Syntheses and applications of 3-bromo-1,2,4,5-tetrazines as novel tools for chemical biology

Enric Ros Simó

Syntheses and applications of 3-bromo-1,2,4,5-tetrazines as novel tools for chemical biology

Enric Ros Simó

Doctoral programme: Química Orgànica

Thesis Directors:

Lluís Ribas de Pouplana and **Antoni Riera Escalé**

Facultat de Química

Departament de Química Orgànica

Universitat de Barcelona

Memòria presentada per l'**Enric Ros Simó** per a optar al grau de DOCTOR en Química Orgànica per la Universitat de Barcelona

Enric Ros Simó

El **Dr. Lluís Ribas de Pouplana**, investigador principal del Laboratori de Traducció Genètica de l'Institut de Recerca Biomèdica (IRB) Barcelona,

i

el **Prof. Antoni Riera Escalé**, catedràtic del departament de Química Inorgànica i Orgànica, Secció de Química Orgànica, de la Facultat de Química de la Universitat de Barcelona,

CERTIFIQUEN: que la present tesi doctoral titulada "Syntheses and applications of 3-bromo-1,2,4,5-tetrazines as novel tools for chemical biology", presentada per **Enric Ros Simó** per optar al títol de doctor per la Universitat de Barcelona, ha estat realitzada sota la seva co-direcció.

Dr. Lluís Ribas de Pouplana **Prof. Antoni Riera Escalé**

Barcelona, de setembre de 2020

This work was performed from October 2016 until September 2020 with the support from La Caixa-Severo Ochoa Ph.D. Fellowship. Research was funded by grants from the Spanish Ministry of Economy, Industry and Competitiveness (MINECO, CTQ2017-87840-P to Prof. Antoni Riera and BIO2015-64572-R to Dr. Lluís Ribas de Pouplana), and La Caixa Foundation (Young BiomedTec ID 100010434 to Enric Ros). IRB Barcelona is the recipient of institutional funding from the Spanish Ministry of Science, Innovation and Universities through the Centres of Excellence Severo Ochoa award and from the CERCA Program of the Catalan Government.

The experimental work was executed at the Gene Translation Laboratory and the Research Unit on Asymmetric Synthesis from the IRB Barcelona, located at the Parc Científic de Barcelona (PCB).

Acknowledgements

És amb una barreja de felicitat i tristesa que escric aquests agraïments, un sentiment agredolç provocat per tots els gran moments que he passat en aquesta etapa i pel fet d'haver-la de tancar per mirar cap a d'altres horitzons.

En primer lloc, mai hauria estat capaç d'arribar aquí sense els meus pares, l'Anna i el Kim. Els valors de vida, basats en l'esforç i la tenacitat, que ens han ensenyat a mi i a la meva germana, la Clara, han estat les eines més útils per fer front a aquest món. Clareta, ara que et toca a tu transmetre'ls a la Gina i l'Axel, estic convençut que ho faràs igual de bé o millor del que ho van fer els nostres pares.

Tampoc seria aquí sense tots els mentors que he trobat en el camí i que han influït en la meva educació i formació. No hi ha prou línies per agrair a tothom que ha utilitzat un temps de la seva vida per ensenyar-me a ser millor professional o persona. En aquest sentit, una part important d'aquesta tasca l'han desenvolupat els meus directors de tesi, el Lluís i el Toni. Més enllà de la recerca, no oblidaré mai que m'acollissin als seus respectius laboratoris després d'un inici insòlit, en el qual estava perdut, desconcertat i plantejant-me seriosament la possibilitat d'abandonar la meva beca doctoral i la investigació. Després de quatre anys, puc dir amb un agraïment immens que han sabut deixar-me la llibertat suficient per seguir les meves idees i donar-me la tutela necessària per arribar a resultats. Aquest equilibri, tan senzill a priori i tan complicat en realitat, és el que ha aconseguit que acabi el doctorat sent molt millor científic del que mai hauria imaginat.

El fet de poder desenvolupar la meva recerca en dos laboratoris ha propiciat que formi part de dos equips de científics i persones dels quals he après molt. Em sento un privilegiat per haver pogut compartir moments, ja siguin de felicitat o frustració, amb tots ells i elles.

Del laboratori de traducció genètica em quedo amb el suport emocional de la Noelia, la persona fonamental del grup d'investigació (tot i que ella no en sigui conscient); amb la perseverança i la passió científica de l'Adrián, sempre disposat a tenir una discussió profitosa; amb l'amabilitat i la capacitat de treballar del Sundar; i sobretot

amb l'Alba, amb qui vam començar junts aquest llarg camí i amb qui he tingut una ajuda constant i incondicional tots aquests anys, especialment en els moments més complicats. També m'enduc molts records d'aquells que hi van ser al principi, quan encara em pensava que canviaria el món, i es van ocupar d'ensenyar-me que la investigació és quelcom diferent: la Marta, l'Albert, l'Helena i l'Àlbert. I per últim, amb els moments compartits amb aquells que han passat més fugaçment pel laboratori, en especial la Nerea i la Irene. Elles em van ajudar en el meu projecte però, sobretot, em van ensenyar el difícil que és ensenyar.

De URSA em quedo amb la importància de la cohesió i el treball en equip. Vaig començar com un extern al grup però mai me'n vaig sentir. I això va ser en gran part gràcies a tots aquells que hi eren quan vaig arribar, alguns dels quals van marxar abans que jo, com el Dan, l'Anna, o l'Ernest. També l'Amparo, un exemple d'investigadora amb qui vaig tenir la sort de compartir projecte durant uns mesos. El Craig, un gran amic amb qui la casualitat ens va unir a compartir vitrina primer, després pis, i a partir d'ara experiències per terres britàniques. O l'Albert, amb qui les incomptables discussions que hem tingut no han fet més que reforçar la nostra curiosa amistat. Segur que l'aventura americana acabarà sent molt MITica. També un agraïment per tots els estudiants que han passat pel laboratori i han deixat petjada, especialment a la Jacqueline i la Dominique, per haver-me ajudat a treure partit de les nostres meravelloses tetrazines. I, finalment, a tots els que us quedeu, serà molt difícil dir-vos adeu. Trobaré a faltar moltíssim riure amb el Pep parlant de la Sotana, el Barça, twitter o del nostre meravellós equip de futbol amateur; també trobaré a faltar a la Marina, el seu *joie de vivre*, i la seva manera de mirar el món. Vaig passar amb tu les setmanes més frustrants del meu doctorat, i encara agraeixo ara el moment en què vas dir "etanol 96%"; també recordaré la fortalesa i personalitat de la Caro; l'excentricitat del Guillem; i l'habilitat del Sergi per treure somriures, ja sigui amb els seus vídeos, frases cèlebres o merdes vàries, fins i tot en moments tristos.

I, més enllà de la gent amb qui he compartit laboratori, de l'IRB me n'emporto un conjunt de persones que espero que segueixin amb mi durant molt temps. Des d'aquells amb qui ja havia compartit històries des de molt abans, com l'Edu, amb qui espero que encara ens en quedin moltíssimes més. També com la Marta o el Salva, precursors en les meves decisions i, en certa manera, grans exemples a seguir. Passant

per aquells amb qui vaig començar aquest viatge que ens ha canviat tant: el Marcos, l'Adrián, l'Adrià, la Sara, el Joel, la Lorena, la Marina i la Isa. I a tots els que vaig anar coneixent a mesura que avançaven els mesos: la Cris, la Leyre, el Ricardo, el Jürgen, la Sara, la Mireia... A tots ells i elles, la nostra etapa junts no s'acaba aquí.

Tampoc no em puc oblidar de les meves arrels. Com més lluny he viscut, més he valorat els amics de tota la vida de Mollerussa i Lleida, els fars que guien el camí de tornada a casa. Vam passar junts la nostra infantesa i adolescència, probablement els anys més importants de les nostres vides per arribar a ser les persones que som ara, i durant els quals vam estar aprenent contínuament els uns dels altres. En aquest sentit, aquells que han estat incansablement al meu costat des del principi han tingut un paper essencial en la meva vida: de l'Edu n'he après la importància de mirar les coses des de diferents punts de vista per arribar a entendre el conjunt; el Bere m'ha ensenyat a afrontar els problemes de cara, amb valor i coratge, fins i tot en el moment més difícil de tots; de l'Albert m'emporto la importància de la sinceritat, el respecte i el compromís; de l'Angel, el nostre germà petit que ens ha superat en tots en maduresa, m'emporto la lliçó que, amb esforç, pots tenir èxit en allò que et proposis; i de l'Àlex n'he après que la cosa més important de la vida és viure-la al màxim, com va fer ell, ja que és terriblement efímera.

Per últim, part d'aquesta tesi també pertany a la persona que ha estat al meu costat durant absolutament tots els minuts que m'he passat escrivint-la. Sense el seu suport, tendresa i amor tot hauria estat molt més complicat. Per tant, g*razie per essere il mio centro di gravità permanente, Medea*.

I senzillament, gràcies a tothom que s'ha interessat per la meva feina, que m'ha ajudat en els moments difícils, amb qui he celebrat els èxits o que, en algun moment o altre, m'ha acompanyat en aquest camí que és la vida.

"Perfection belongs to narrated events, not to those we live".

Primo Levi, in *The Periodic Table* (*Il Sistema periodico*)

"Genius, it is said, is the ability to see the obvious before anyone else".

Nick Lane, in *Life Ascending: The Ten Great Inventions of Evolution*

This doctoral thesis is dedicated to the memory of Àlex Galitó Cortassa,

in remembrance of his vitality, spontaneity and joy.

List of abbreviations

AARS	Aminoacyl-tRNA synthetase
AcOH	Acetic acid
Amp	Ampicillin
APCI	Atmospheric pressure chemical ionization
Ara	Arabinose
AU	Absorbance unit
AUC	Area under the curve
BCN	Bicyclo[6.1.0]nonyne
Boc	*tert*-butyloxycarbonyl
CAN	Acetonitrile
CC_{50}	Cytotoxic concentration 50
CHO	Chinese hamster ovary
CLIPTAC	Click-formed proteolysis targeting chimera
Conv	Conversion
CuAAC	Copper(I)-catalyzed azide-alkyne cycloaddition
DCM	Dichloromethane
DIPEA	N,N-Diisopropylethylamine
DMF	Dimethyl formamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
Dox	Doxorubicin
DSC	Differential scanning calorimetry
DTT	1,4-Dithiotheitol
EDG	Electron-donating group
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
eq	Equivalent
ESI	Electrospray ionization
EtOAc	Ethyl acetate
EtOH	Ethanol
EWG	Electron-withdrawing group
FA	Formic acid
FACS	Fluorescence-activated cell sorting
Fig	Figure
FRET	Förster resonance energy transfer
GCE	Genetic code expansion
GFP	Green Fluorescent Protein
HOMO	Highest occupied molecular orbital
HRMS	High-resolution mass spectrometry
IAA	Indoleacetic acid
iEDDA	Inverse electron-demand Diels-Alder
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Infrared
KDa	Kilodalton
LB	Lysogeny broth
LC-MS	Liquid chromatography-Mass spectrometry
LUCA	Last Universal Common Ancestor
LUMO	Lowest unoccupied molecular orbital
MAPK	Mitogen-activated protein kinase
MeOH	Methanol
MOM	Methoxymethyl ether
Mp	Melting point
mRNA	Messenger RNA
Mw	Molecular weight
MWCO	Molecular weight cut-off
ncAA	Non-canonical amino acid

nEDDA	Normal electron-demand Diels-Alder
NHS	N-Hydroxysuccinimide
NMR	Nuclear magnetic resonance
Nu	Nucleophile
OCT	Octreotide
OD	Optical density
OTS	Orthogonal translation system
PBS	Phosphate buffered saline
PDB	Protein Data Bank
PEG	Polyethylene glycol
pelB	Pectatelyase B
PET	Photoinduced electron transfer
Ph	Phenyl
Ph.D.	Doctor of Philosophy
PNP	*Para*-nitrophenyl carbamate
PROTAC	Proteolysis targeting chimera
Pyl	Pyrrolysine
PylRS	Pyrrolysyl-tRNA synthetase
RF	Release factor
RNA	Ribonucleic acid
RNAse A	Ribonuclease A
SeC	Selenocysteine
S_N(ANRORC)	Addition of the nucleophile, ring opening, ring closure substitution
S_NAr	Nucleophilic aromatic substitution
SPAAC	Strain-promoted azide-alkyne cycloaddition
SPANC	Strain-promoted alkyne-nitrone cycloaddition
SPPS	Solid-phase peptide synthesis
SSTR2	Somatostatin receptor type 2
T	Temperature
TBAF	Tetrabutylammonium fluoride
TBET	Through-bond energy transfer
TCO	*Trans*-cyclooctene
Tet	Tetracycline
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
tRNA	Transfer RNA
TyrRS	Tyrosyl-tRNA synthetase
UPLC	Ultra-performance liquid chromatography
UPS	Ubiquitin-proteasome system
XIC	Extracted ion chromatogram

List of publications

The experimental results of this thesis are presented through eight chapters (from Chapter 2 to Chapter 9). Some of the results have already been published in the following scientific publications during the doctoral period.

I. Synthesis and application of 3-bromo-1,2,4,5-tetrazine for protein labeling to trigger click-to-release biorthogonal reactions

E. Ros, M. Bellido, X. Verdaguer, L. Ribas de Pouplana and A. Riera.

Bioconjug. Chem., 2020, **31**, 933–938. With featured supplementary cover (next page).

II. Synthesis of 3-alkyl-6-methyl-1,2,4,5-tetrazines via a Sonogashira-type cross-coupling reaction

E. Ros, A. Prades, D. Forson, J. Smyth, X. Verdaguer, L. Ribas de Pouplana and A. Riera.

Chem. Commun., 2020, ahead of print, DOI: 10.1039/D0CC03482G

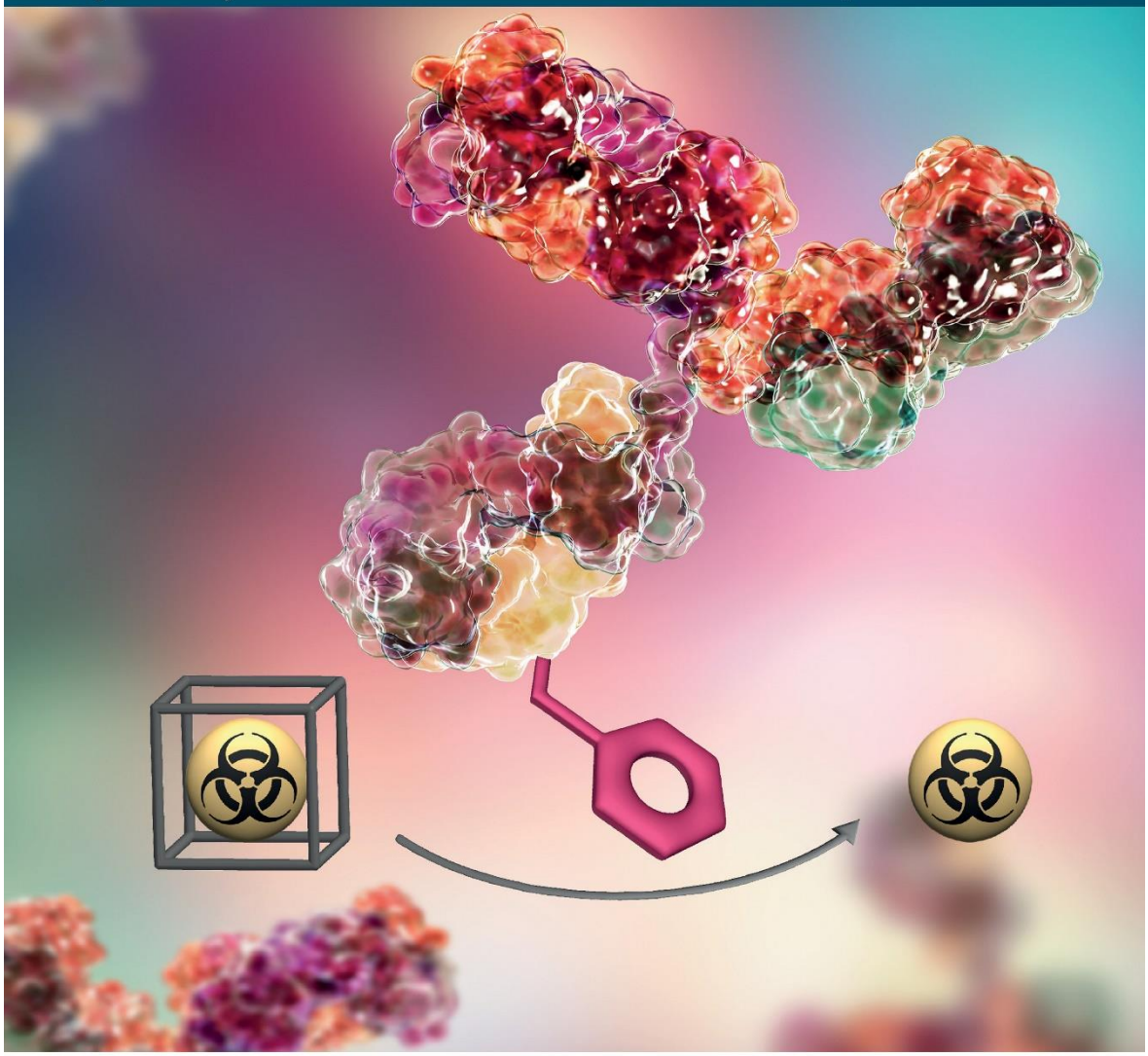
III. Learning from nature to expand the genetic code

E. Ros, A. Torres and L. Ribas de Pouplana.

Trends Biotechnol., 2020, ahead of print, DOI: 10.1016/j.tibtech.2020.08.003

BC Bioconjugate Chemistry

pubs.acs.org/bc March 2020 Volume 31, Number 3



ACS Publications
Most Trusted. Most Cited. Most Read.

www.acs.org

Chapter 1

General introduction and objectives

1.1. Performing chemical reactions *in vivo*:

Life as we know it is based on the coordinated assembly of a myriad of chemical reactions that constitute all metabolic pathways and interactions necessary for organisms to exist. These chemical reactions happen uniquely due to the physicochemical interactions taking place between the different molecules that conform living entities. Hence, one can imagine biological environments as an extremely complex combination of biomolecules, ranging from proteins to lipids, passing by nucleic acids and carbohydrates, all of them floating in an aqueous media flooded with metal ions, cofactor molecules, and probably many other interactors we are still clueless of their existence.

From this perspective, the ability to perform specific and selective chemical reactions under these conditions should be, apparently, of great difficulty. Not only shall the participating chemical groups not cross-react with the functionalities of the different biomolecules, but they also need to be stable and react rapidly to avoid being metabolized. Moreover, the reaction must happen in water at physiological temperature (typically 37ºC) and, importantly, neither the participating chemical groups nor the products or by-products can be toxic. These are the critical characteristics to consider a chemical reaction truly *bioorthogonal*,[1] a concept defined by Prof. Carolyn Bertozzi in 2003, when the term was used for the first time.[2] Obviously, the majority of known chemical reactions do not fulfill these attributes and, by extension, do not take place under biocompatible conditions. Nevertheless, there are a few successful examples of reactions used either in cell culture or directly *in vivo*, each one presenting advantages and drawbacks.

The emergence of bioorthogonal reactions cannot be understood without the previous advent of the so-called *click chemistry*, a term coined by Prof. Karl Barry Sharpless in 2001, the same year in which he was awarded the Nobel Prize in chemistry for his work on catalytic enantioselective oxidations. Click chemistry defines those reactions that are high-yielding, wide in scope, stereospecific and generate harmless by-products. Their reaction conditions are simple and non-sensitive to oxygen or water, and the final product is effortlessly isolated.[3] Of

particular interest is the logical requirement that click reactions must present a large thermodynamic driving force, usually greater than 20 kcal·mol^{-1}.[3]

Examples of such reactions comprehend certain cycloadditions of unsaturated species, for instance the 1,3-dipolar cycloaddition (or Huisgen cycloaddition due to its discoverer)[4] taking place between azides and alkynes with copper catalysis, commonly known as Copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC). The CuAAC, apart from being the premier example of a click chemistry reaction according to Sharpless,[5] has probably become the most widely used click chemistry reaction. Other click reactions include nucleophilic ring-opening reactions of strained heterocyclic electrophiles, oxidation reactions such as epoxidations or aziridinations, certain Michael additions and non-aldolic carbonyl chemistry.[5]

Sharpless inspiration to describe click chemistry reactions came from the observation of nature's preference towards creating carbon-heteroatom bonds over carbon-carbon bonds.[3] In a fascinating coincidence, it was these same click chemistry reactions what prompted the appearance of bioorthogonal reactions shortly after. In essence, click reactions can become bioorthogonal when they are unreactive towards the functional groups found in nature, present very fast kinetics and do not require or generate toxic entities. For example, the abovementioned CuAAC is not an ideal bioorthogonal reaction because of the Cu(I)-mediated generation of reactive oxygen species and subsequent cytotoxic oxidative stress.[6] In response to this limitation, non-catalyzed 1,3-dipolar cycloadditions were developed taking advantage of the ring-strain of the alkyne reactant, resulting in the bioorthogonal strain-promoted azide-alkyne cycloaddition (SPAAC).[7]

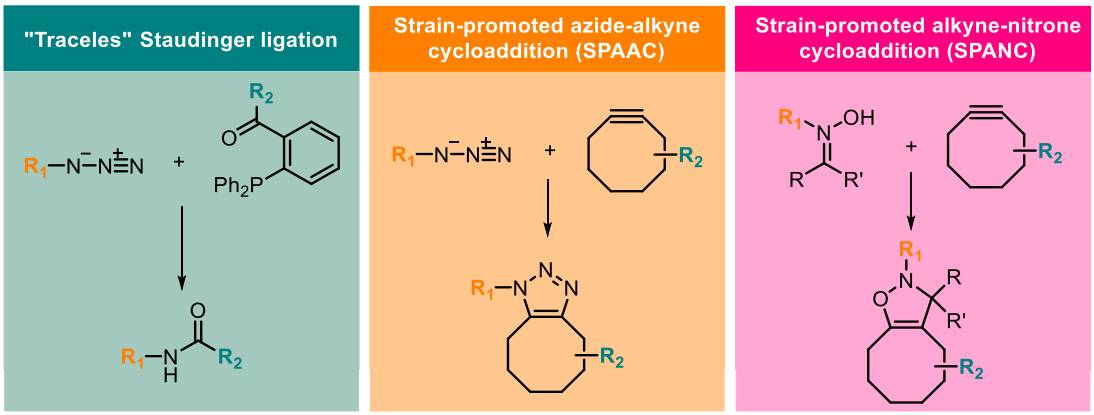


Figure 1-1: Toolbox of bioorthogonal reactions. Outlined, some of the most commonly used bioorthogonal reactions.

The toolbox of truly bioorthogonal reactions is, as expected, relatively limited (**Fig. 1-1**). The first ever bioorthogonal reaction employed in mammalian cell culture, used to label mucin-type O-linked membrane glycoproteins, was the Staudinger ligation between certain phosphines and azides.[2] Despite being a biocompatible reaction, specially its traceless version,[8,9] the oxidation of the phosphine *in vivo* and the sluggish kinetics limit its applicability. On the other hand, similarly to the previously cited SPAAC, the strain-promoted alkyne-nitrone cycloaddition (SPANC) is faster in terms of kinetic rate constant but suffers from limited *in vivo* stability of the nitrone counterpart.[10]

1.2. The inverse electron-demand Diels-Alder cycloaddition:

Probably, the most ideal bioorthogonal ligation described to date is the inverse electron-demand Diels-Alder (iEDDA) cycloaddition. This reaction is highly specific, since none of the participating functional groups is found in natural biomolecules; displays a considerable kinetic rate enhancement in polar solvents, including water, which is positive for chemical biology applications; does not require catalysis of any kind; and the only released by-product, molecular nitrogen (N_2), is harmless for living organisms.[10]

Just like the normal electron-demand Diels-Alder (nEDDA), the iEDDA is also a pericyclic [4+2]-cycloaddition between a diene (a 4π-electron system) and a dienophile (a 2π-electron system, typically an alkene or alkyne). The difference between both reaction arises from the electronic effects of the substituents of the diene and dienophile: whereas the nEDDA is characterized by dienophiles bearing electron-withdrawing groups (EWGs) and electron-rich dienes, the iEDDA undergoes between electron-poor dienes and electron-rich dienophiles (**Fig. 1-2, a**).[11]

For bioorthogonal purposes, the reactivity of the dienophiles participating in the iEDDA can be increased either through the substitution with electron-donating groups (EDGs) or by introducing ring strain into the alkene/alkyne. Over the years, many designer molecules have been used as dienophiles,[12] in all cases trying to find the right compromise between reactivity and stability *in vivo*. The most widely used alkenes and alkynes are, in descendent order of reactivity, *trans*-cyclooctenes (TCOs),

bicyclo[6.1.0]nonynes (BCNs), cyclopropenes, norbornene derivatives and styrenes (**Fig. 1-2, b**).[13]

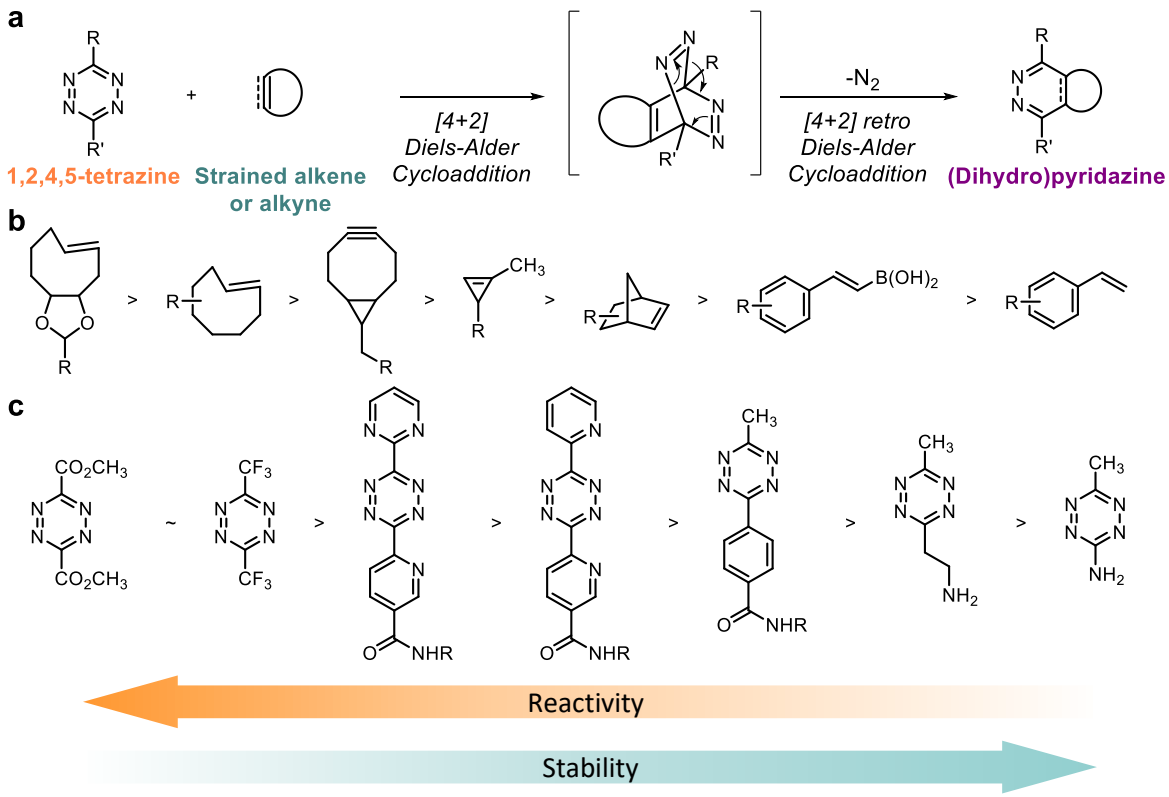


Figure 1-2: Bioorthogonal iEDDA cycloaddition. a) The reaction between a tetrazine and a strained alkene or alkyne is a two-step reaction, the first one being a [4+2] cycloaddition to form a bicyclic bridged intermediate, followed by its break down into the resulting (dihydro)pyridazine through the release of molecular nitrogen. **b)** Mainly used strained alkenes or alkynes acting as dienophiles for bioorthogonal iEDDA cycloadditions, in descendent order of reactivity. **c)** Overview of different tetrazines, whose reactivity can be tuned using different combinations of substituents in position 3 and 6 depending on the given application. In general, EWGs as substituents enhance the electron-poor character of the diene, thus increasing reactivity but lowering stability.

In contrast to the wide variety of dienophiles that can participate in iEDDA reactions, the number of different functional groups that may act as the diene counterpart is much more limited. Except from a few examples showing that 1,2,4-triazines also react with TCO or BCN derivatives under biocompatible conditions,[14–16] nearly all reports employing the iEDDA cycloaddition as a bioorthogonal reaction rely on the use of 1,2,4,5-tetrazines as dienes (**Fig. 1-2, c**). The initial reports on tetrazine synthesis date back from the 19th century, precisely from 1893, year in which Adolf Pinner reported the first synthesis for the tetrazine core. Since then, many convergent and complementary synthetic approaches have been described, most of them sharing

the preparation of a dihydrotetrazine that is subsequently oxidized to the final tetrazine.

Despite numerous efforts to afford tetrazinyl-compounds in quantitative yields, some substrates are still synthetically very challenging.[12,13] Consequently, the first aim of this thesis was the design and development of a useful synthetic route to attain a suitable tetrazine reagent that enabled the late-stage functionalization of a diverse set of small molecules and biomolecules. A detailed description of the different synthetic routes previously described to achieve tetrazines, together with the rationale and optimization of our own synthesis of certain brominated tetrazines, will be discussed in **Chapter 2**.

The participation of tetrazines in iEDDA reactions was initially described in the late 1950s, when their use as suitable precursors of the pyridazine core was demonstrated through the reaction with several unsaturated compounds, like styrene. Since then, and for the rest of the century, the utilization of tetrazines in iEDDA reactions was limited to natural product synthesis.[11] The fact that the reaction does not require a catalyst, together with its high levels of regioselectivity and atom economy, turned tetrazines into excellent precursors of nitrogenated heterocycles.

It was not until 2008 that the tetrazine-based iEDDA cycloaddition was employed for bioorthogonal purposes. Fox and co-workers described the reaction between tetrazines and TCO, which displayed a kinetic rate constant never previously observed for any bioorthogonal reaction, broad range of functional group tolerance and stability in water, cell media an cell lysate.[17] In the same year, Hildebrand's group reported the same observations in the reaction between a monosubstituted tetrazine and a norbornene derivative.[18] Somewhat surprising is the fact that the two applications shown in these two initial papers, bioconjugation and live cell imaging, have been the two main fields in which the tetrazine ligation has been mostly applied.

Since these two pioneering examples, the iEDDA cycloaddition has been by far the most utilized biorthogonal reaction and, by extension, multiple methods to generate useful molecules functionalized with either strained dienophiles or tetrazines have been developed.[12] However, new methods to easily synthesize a higher number of

probes or bioactive compounds are needed to expand the use of this reaction outside the commonly used reagents.

Consequently, the second aim of the thesis was to employ the brominated tetrazine reagents detailed in **Chapter 2** to achieve late-stage functionalization of different small molecules with interesting properties for biorthogonal purposes. The synthesis of a set of compounds achieved through nucleophilic aromatic substitution (S_NAr), with a special emphasis on fluorescent ether-bridged tetrazines, will be described in **Chapter 3**, and the optimization of a cross-coupling reaction of the same synthon to incorporate tetrazines through C-C bond formation, focusing on potential bioactive molecules, will be detailed in **Chapter 4**.

1.3. Decorating biomolecules with tetrazine functionalities:

The appearance of bioorthogonal reactions opened the door for countless possibilities to solve one of the most difficult conundrums in molecular biology research: how can we study the function of a single biomolecule in its native context? Traditionally, this challenge has been overcome to a certain extent using reporter tags fused to the biopolymer of interest to track its subcellular localization, potential interactors and other general properties. As a compelling recognition of the impact of this strategy, Shimomura, Chalfie, and Tsien were awarded the Nobel Prize in chemistry in 2008 for the discovery and development of the green fluorescent protein (GFP). However, these approaches may perturb the biomolecule of study, potentially leading to a deficient or completely altered function and skewed observations.

The eruption of bioorthogonal reactions supposed a substantial advance in modern chemical biology. The potential to label biomolecules with functional groups to gain spatiotemporal control of their reactions, while minimally perturbing their structure or physicochemical properties, enabled several important discoveries. Nevertheless, achieving effective bioconjugation with reactive groups suitable for bioorthogonal reactions remains a challenge.

The most interesting biomolecules for bioconjugation processes are proteins, nucleic acids[19] and polysaccharides.[20] Proteins are, however, the most abundant

biomolecules in mammalian cells, comprising ca. 60% of a cell's dry mass,[21] and play a central role in virtually all metabolic processes of living matter. The study, manipulation and expansion of protein functions is not only of vital importance in research, but also in the therapeutic field, where the use of protein-based drugs (predominantly monoclonal antibodies) has been steadily increasing in the past decade.[22] Therefore, the present doctoral thesis is mainly focused on protein bioconjugation, although the use of the methods developed here will be applied in the future for the bioconjugation of other biomolecules.

Therefore, the fourth aim of the thesis was to find a simple, useful methodology to decorate proteins with tetrazinyl functionalities, with a special focus on their subsequent application. Consequently, a detailed explanation and characterization of a chemoselective lysine-labeling method, followed by a proof-of-concept of the potential utilization of the labeled proteins in a therapeutically relevant context, will be provided in **Chapter 5** and **Chapter 6**.

1.4. Genetic encoding of tetrazine-containing amino acids:

Even though targeting some of the naturally occurring amino acids in a protein sequence is undoubtedly a practical way to achieve bioconjugation, it is also true that changing proteins at the level of their core sequence to introduce desired chemical functionalities is a much more elegant approach.

Proteins are biosynthesized in the ribosome of all living creatures in a process called translation, a fundamental step in the central dogma of molecular biology.[23] In a startling invariability, all known species follow the same set of rules that govern the translation of a 4-letter nucleotide alphabet into the 20-letter alphabet of proteins, namely the genetic code. Consequently, the genetic code is considered universal: from the Last Universal Common Ancestor (or LUCA)[24] to *Homo sapiens,* all forms of life, irrespectively of their apparent simplicity, almost invariably utilize the same combinations of nucleotide triplets (or codons) to encode for the building blocks of proteins: the proteinogenic or "natural" amino acids. Since there are 4 nucleotides (A, U, G and C) organized in groups of 3, there are a total of $4^3 = 64$ possible combinations:

61 of them encode for each one of the 20 natural amino acids, while the remaining 3 result in translation termination or "stop" codons (**Fig. 1-3, a**).

Arising from the simple observation that there are more codons than amino acids, one can easily deduce that the genetic code is redundant and, by extension, degenerate because multiple codons may be translated as the same amino acid. The reasons why the genetic code did not evolve into accepting a broader range of amino acid entities, thereby increasing the chemical repertoire available to generate proteins, is still unclear despite numerous research efforts dedicated to answering this question since Francis Crick referred to it as the "frozen accident" in 1968.[25,26] What seems undeniable is that, accidentally or not, the genetic code is evolved to minimize the deleterious effects of mutations, cleverly reducing their impact by using similar codons to encode amino acids with similar chemical properties.[27] Indeed, the universal genetic code is more robust than, at least, a million randomly generated genetic codes.[28]

There are, however, notable examples of species that make use of the denominated rare amino acids selenocysteine (SeC) and pyrrolysine (Pyl), commonly referred to as the 21st and 22nd proteinogenic amino acids. In both cases, one of the 3 existing stop codons is recoded into encoding for one of these amino acids: the opal (UGA) stop codon in the case of SeC,[29] and the amber (UAG) stop codon in the case of Pyl.[30] Of note, this recoding is accomplished through the presence of a tightly regulated supplementary translation machinery, such as special aminoacyl-tRNA synthetases (AARSs) and other auxiliary factors, that permits their incorporation during translation.[31] Thus, SeC and Pyl are examples of naturally occurring cases of an *expanded* genetic code, in which the rules have been altered in order to encode for more than 20 amino acids.

Maybe inspired by the nature's trick to encode SeC, Schultz's laboratory reported in 2001 an artificially expanded genetic code in the prokaryotic organism *Escherichia coli.*[32] The followed approach was based on the recoding of the amber stop codon through the expression of the tyrosyl-tRNA synthetase (TyrRS) of the thermophilic methanogenic archaea *Methanocaldococcus janaschii,* together with its cognate tRNA with a modified anticodon sequence ($tRNA^{Tyr}_{CUA}$) (**Fig. 1-3, b**).[32] Importantly, the

*Mj*TyrRS/tRNA$^{Tyr}_{CUA}$ pair is orthogonal in *E. coli* (a detailed explanation of the reasons for orthogonality will be displayed in **Chapter 7**).

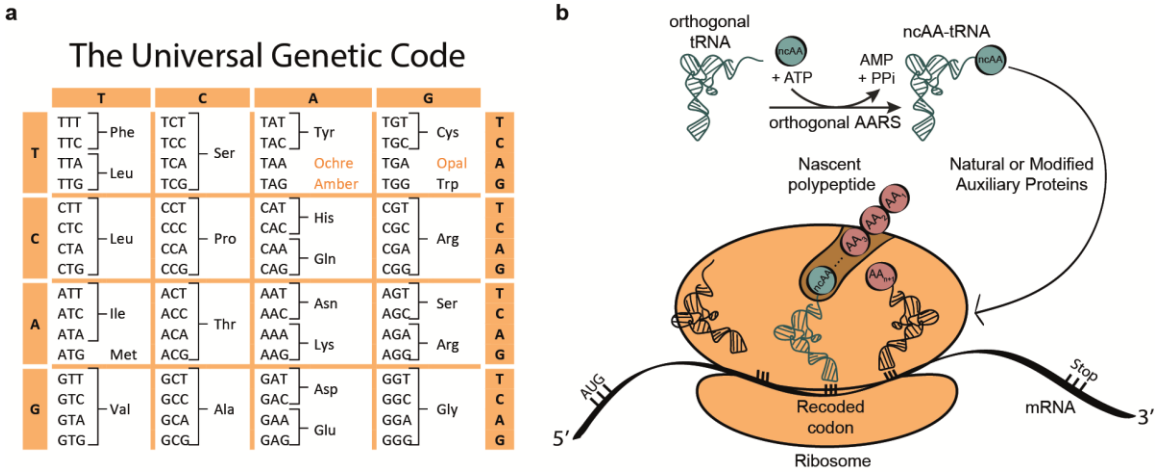


Figure 1-3: The universal genetic code and its expansion. a) Table of the set of rules to translate each codon unambiguously into a natural amino acid or a stop codon (commonly referred to as ochre, opal and amber). **b)** Representation of the behavior of an expanded genetic code. Initially, an orthogonal AARS aminoacylates its cognate tRNA with the desired ncAA, and the ncAA-tRNA complex translocates into the ribosome to participate in the translation of a nascent protein through the recoding of a given codon.

Remarkably, only one year later, in 2002, the presence of Pyl in the proteins of some methanogenic archaea and bacteria was discovered,[33,34] and with it the striking observation of multiple parallelisms with the system developed by Schultz's group: Pyl also used a recoded amber stop codon, and a new AARS/tRNA pair (pyrrolysyl-tRNA Synthetase, PylRS), orthogonal to the rest of the AARS/tRNA machinery of the organism. Unsurprisingly, the PylRS/tRNA pair was ultimately also utilized as an alternative to the *Mj*TyRS/tRNA pair to expand the genetic code of prokaryotes and, furthermore, to date is the most reliable system for genetic code expansion in eukaryotic cells.[30] Alternatively, it also represents an example in which, unwillingly and unexpectedly, research mimicked with great fidelity natural evolution, creating a remarkable coincidence.

The ability to perform genetic code expansion (GCE) of a given host organism with an unnatural amino acid, or non-canonical amino acid (ncAA) in this specific case, bearing a desired chemical functionality is obviously highly attractive. Not only allows the expression of proteins with expanded functionalities, but they are already biosynthesized in a particular cellular context, enabling the study of their function in

quasi-endogenous conditions. Moreover, the ncAAs are incorporated at a specific position in the protein sequence, enabling a level of fine-tuning hardly accessible for any other protein labeling methods.[35]

We rationalized it would be highly appealing to increase the toolbox of available ncAAs bearing tetrazinyl functionalities, in order to perform site-selective protein labeling with groups suitable for bioorthogonal reactions. Precedents showed it was possible to perform GCE in prokaryotes with the *Mj*TyrRS system, albeit only with a limited number of unsymmetrically substituted tetrazine structures in the incorporated ncAAs.[36,37] However, to date there are still no reports of successful ncAA incorporation with tetrazinyl functionalities in mammalian cells.

From this perspective, the fifth aim of this thesis was to perform GCE with the novel tetrazine-containing ncAAs developed in the previous chapters. Initially, a detailed explanation of the strategies followed to expand the genetic code will be provided in **Chapter 7**. Next, the *Mj*TyrRS/tRNA$^{Tyr}_{CUA}$ pair to expand the genetic code of *E. coli* with a ncAA bearing a tetrazine moiety will be provided in **Chapter 8**. Finally, the progress made towards the development of a tetrazine-containing ncAA to expand the genetic code of mammalian cells through the PylRS/tRNA pair will be presented in **Chapter 9**.

As it has been highlighted throughout the introduction, the beginning of the XXI century brought with it the appearance of concepts like *click chemistry* (Sharpless, 2001) and *bioorthogonal reactions* (Bertozzi, 2003), which were a landmark in the **chemical biology** field. Contemporarily to these discoveries, the *expansion of the universal genetic code* (Schultz, 2001) of living organisms started a whole new era in what is now considered **synthetic biology**, a fast-developing field that is completely changing the paradigm of how we understand life and living matter. Despite being two separate areas of knowledge, they share the common objective of modulating and customizing the limitless tools offered by chemistry and biology to our advantage. Consequently, plenty of excellent research has been performed combining these fields, and this present thesis is just another attempt to bring these two disciplines together.

1.5. Objectives:

In summary, the objectives of the present doctoral thesis were:

Objective 1. Design and development of a useful synthetic route to a 1,2,4,5-tetrazine reagent that enabled the late-stage functionalization of a diverse set of biomolecules (**Chapter 2**).

Objective 2. To study the nucleophilic aromatic substitution (S_NAr) reactions of the previously developed halogenated tetrazine reagents in late-stage functionalization of small molecules with interesting properties for biorthogonal purposes (**Chapter 3**).

Objective 3, To study the cross-coupling reaction of the previously developed halogenated tetrazine reagents for the synthesis of (dialkyl)tetrazines (**Chapter 4**).

Objective 4. To find a simple, useful methodology to decorate proteins with tetrazinyl functionalities, developing a chemoselective lysine-labeling method, and applying it in a therapeutically relevant context (**Chapters 5** and **6**).

Objective 5. To perform genetic code expansion (GCE) with the novel tetrazine-containing ncAAs developed in the previous chapters (**Chapters 7**, **8** and **9**).

1.6. Bibliography:

1 E. M. Sletten and C. R. Bertozzi. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chem., Int. Ed.* 2009, **48**, 6974–6998.

2 H. C. Hang, C. Yu, D. L. Kato and C. R. Bertozzi. A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *Proc. Natl. Acad. Sci.* 2003, **100**, 14846–14851.

3 H. C. Kolb, M. G. Finn and K. B. Sharpless. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem., Int. Ed.* 2001, **40**, 2004–2021.

4 R. Huisgen. Centenary Lecture - 1,3-Dipolar Cycloadditions. *Proc. Chem. Soc.* 1961, 357–369.

5 H. C. Kolb and K. B. Sharpless. The growing impact of click chemistry on drug discovery. *Drug Discov. Today* 2003, **8**, 1128–1137.

6 L. M. Gaetke and C. K. Chow. Copper toxicity, oxidative stress, and antioxidant nutrients. *Toxicology* 2003, **189**, 147–163.

7 J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli and C. R. Bertozzi. Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci.* 2007, **104**, 16793–16797.

8 E. Saxon, J. I. Armstrong and C. R. Bertozzi. A 'traceless' Staudinger ligation for the chemoselective synthesis of amide bonds. *Org. Lett.* 2000, **2**, 2141–2143.

9 B. L. Nilsson, L. L. Kiessling and R. T. Raines. Staudinger ligation: A peptide from a thioester and azide. *Org. Lett.* 2000, **2**, 1939–1941.

10 B. L. Oliveira, Z. Guo and G. J. L. Bernardes. Inverse electron demand Diels–Alder reactions in chemical biology. *Chem. Soc. Rev.* 2017, **46**, 4895–4950.

11 R. A. A. Foster and M. C. Willis. Tandem inverse-electron-demand hetero-/retro-Diels-Alder reactions for aromatic nitrogen heterocycle synthesis. *Chem. Soc. Rev.* 2013, **42**, 63–76.

12 H. Wu and N. K. Devaraj. Advances in Tetrazine Bioorthogonal Chemistry Driven by the Synthesis of Novel Tetrazines and Dienophiles. *Acc. Chem. Res.* 2018, **51**, 1249–1259.

13 S. Mayer and K. Lang. Tetrazines in Inverse-Electron-Demand Diels–Alder Cycloadditions and Their Use in Biology. *Synthesis* 2016, **49**, 830–848.

14 D. N. Kamber, Y. Liang, R. J. Blizzard, F. Liu, R. A. Mehl, K. N. Houk and J. A. Prescher. 1,2,4-Triazines Are Versatile Bioorthogonal Reagents. *J. Am. Chem. Soc.* 2015, **137**, 8388–8391.

15 A. Vázquez, R. Dzijak, M. Dračínský, R. Rampmaier, S. J. Siegl and M. Vrabel. Mechanism-Based Fluorogenic trans-Cyclooctene–Tetrazine Cycloaddition. *Angew. Chem., Int. Ed.* 2017, **56**, 1334–1337.

16 K. A. Horner, N. M. Valette and M. E. Webb. Strain-Promoted Reaction of 1,2,4-Triazines with Bicyclononynes. *Chem. Eur. J.* 2015, **21**, 14376–14381.

17 M. L. Blackman, M. Royzen and J. M. Fox. Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels–Alder Reactivity. *J. Am. Chem. Soc.* 2008, **130**, 13518–13519.

18 N. K. Devaraj, R. Weissleder and S. A. Hilderbrand. Tetrazine-based cycloadditions: Application to pretargeted live cell imaging. *Bioconjug. Chem.* 2008, **19**, 2297–2299.

19 J. Šečkutė, J. Yang and N. K. Devaraj. Rapid oligonucleotide-templated fluorogenic tetrazine ligations. *Nucleic Acids Res.* 2013, **41**, e148–e148.

20 N. K. Devaraj, G. M. Thurber, E. J. Keliher, B. Marinelli and R. Weissleder. Reactive polymer enables efficient in vivo bioorthogonal chemistry. *Proc. Natl. Acad. Sci.* 2012, **109**, 4762–4767.

21 A. Uzman. Molecular biology of the cell (4th ed.): Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. *Biochem. Mol. Biol. Educ.* 2003, **31**, 212–214.

22 L. Urquhart. Top drugs and companies by sales in 2018. *Nat. Rev. Drug Discov.* 2019, **18**, 2019.

23 F. Crick. Central Dogma of Molecular Biology. *Nature* 1970, **227**, 561–563.

24 D. L. Theobald. A formal test of the theory of universal common ancestry. *Nature* 2010, **465**, 219–222.

25 F. H. C. Crick. The origin of the genetic code. *J. Mol. Biol.* 1968, **38**, 367–379.

26 L. R. de Pouplana, A. G. Torres and À. Rafels-Ybern. What froze the genetic code? *Life* 2017, **7**, 1–6.

27 C. R. Woese. On the evolution of the genetic code. *Proc. Natl. Acad. Sci.* 1965, **54**, 1546–1552.

28 S. J. Freeland and L. D. Hurst. The genetic code is one in a million. *J. Mol. Evol.* 1998, **47**, 238–248.

29 T. C. Stadtman. Selenocysteine. *Annu. Rev. Biochem.* 1996, **65**, 83–100.

30 W. Wan, J. M. Tharp and W. R. Liu. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* 2014, **1844**, 1059–1070.

31 J. Yuan, P. O'Donoghue, A. Ambrogelly, S. Gundllapalli, R. L. Sherrer, S. Palioura, M. Simonović and D. Söll. Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems. *FEBS Lett.* 2010, **584**, 342–349.

32 L. Wang, A. Brock, B. Herberich and P. G. Schultz. Expanding the genetic code of Escherichia coli. *Science* 2001, **292**, 498–500.

33 B. Hao, W. Gong, T. K. Ferguson, C. M. James, J. A. Krzycki and M. K. Chan. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* 2002, **296**, 1462–1466.

34 G. Srinivasan, C. M. James and J. A. Krzycki. Pyrrolysine encoded by UAG in archaea: Charging of a UAG-decoding specialized tRNA. *Science* 2002, **296**, 1459–1462.

35 C. C. Liu and P. G. Schultz. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* 2010, **79**, 413.

36 J. L. Seitchik, J. C. Peeler, M. T. Taylor, M. L. Blackman, T. W. Rhoads, R. B. Cooley, C. Refakis, J. M. Fox and R. A. Mehl. Genetically Encoded Tetrazine Amino Acid Directs Rapid Site-Specific in Vivo Bioorthogonal Ligation with trans-Cyclooctenes. *J. Am. Chem. Soc.* 2012, **134**, 2898–2901.

37 R. J. Blizzard, D. R. Backus, W. Brown, C. G. Bazewicz, Y. Li and R. A. Mehl. Ideal Bioorthogonal Reactions Using A Site-Specifically Encoded Tetrazine Amino Acid. *J. Am. Chem. Soc.* 2015, **137**, 10044–10047.

Chapter 2

Syntheses of 3-bromo-1,2,4,5-tetrazines

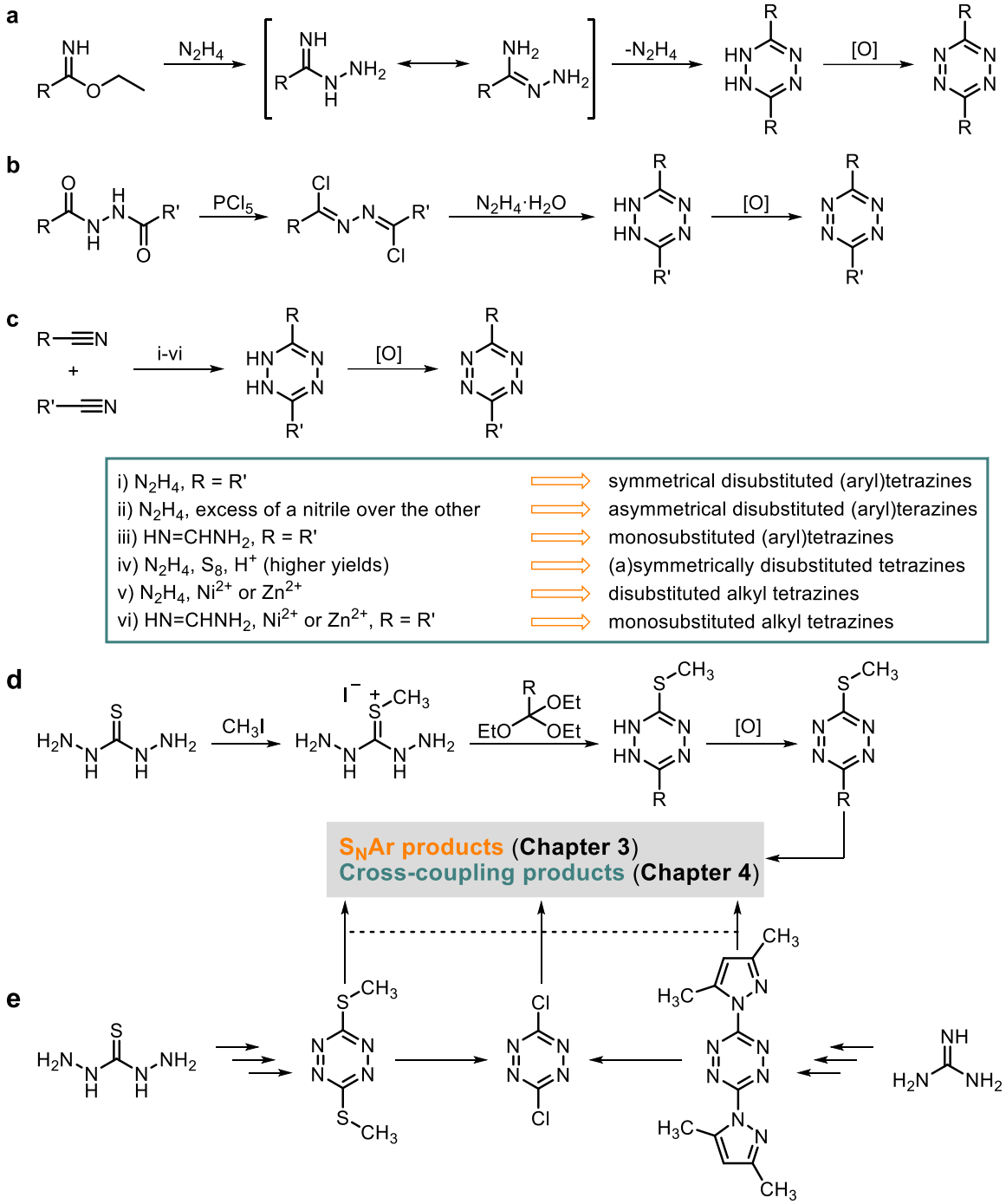
2.1. Overview of the synthetic approaches to 1,2,4,5-tetrazines

2.1.1. Synthesis of the 1,2,4,5-tetrazine core:

The first report on 1,2,4,5-tetrazines was published by Adolf Pinner back in 1893.[1] He described the condensation of imidoesters with hydrazine to the corresponding amidrazone intermediate, which in turn reacts with excess hydrazine leading to the formation of dihydrotetrazine, that is finally oxidized to the final tetrazine (**Scheme 2-1, a**). Interestingly, some of the common features and limitations encountered in tetrazine synthesis were already observed in this initial report: firstly, the generation of a dihydrotetrazine that requires further oxidation with appropriate reagents, which will be discussed in detail on **section 2.2.** Secondly, the formation of amidrazone intermediates, which is believed to happen in most synthetic methods for tetrazine formation employing hydrazine. And thirdly, the challenge to afford asymmetrically substituted tetrazines on position 3 and 6.

Some years later, in 1906,[2] an alternative tetrazine synthesis was developed by R. Stollé that involved the treatment of acylhydrazides with phosphorous pentachloride to achieve the formation of 1,2-dichloromethylene hydrazines, which were subsequently condensed with hydrazine and oxidized to afford the final tetrazine product (**Scheme 2-1, b**). By tuning the substituents in the initial acylhydrazide, different symmetrical and unsymmetrical tetrazines can be achieved in moderate yields.[3]

Except in certain occasions where the desired substrate requires these original procedures, typically tetrazines are synthesized through the method developed by Hofmann and Ehrhart in 1912.[4] It involves the condensation of two nitriles in the presence of hydrazine, leading to the formation of dihydrotetrazine that is subsequently oxidized (**Scheme 2-1, c**). In this synthesis, the use of anhydrous hydrazine normally increases the yields, but nowadays, its commercialization is restricted in Europe due to safety concerns.[5] Therefore, either hydrazine hydrate ($N_2H_4 \cdot H_2O$) or solutions in THF or ethanol are the ones normally employed.[6]



Scheme 2-1: Main synthetic routes to generate 1,2,4,5-tetrazines. a) Initial Pinner synthesis. **b)** Stollé's synthetic route. **c)** Hofmann and Ehrhart nitrile condensation with hydrazine, with all the developed variations. **d)** Asymmetrical 3-(methylthio)tetrazines generated from thiocarbohydrazide. **e)** Symmetrical electrophilic tetrazines synthesized from thiocarbohydrazide or guanidine. The utilization of tetrazines generated in routes **d** and **e** will be discussed in detail in **Chapter 3** and **Chapter 4**.

The simplest application of this route is to prepare symmetrical and asymmetrical (aryl)tetrazines, which can be achieved using a single starting nitrile (**Scheme 2-1, c-i**) or adding two different nitriles, one of them in excess (**Scheme 2-1, c-ii**). In the

latter case, the formation of a statistical mixture of products, which can be difficult to purify, is an unavoidable limitation. If the target product is a monosubstituted tetrazine, formamidine acetate can be employed with a suitable nitrile (**Scheme 2-1, c-iii**). Over the years, several modifications on this synthesis have been reported aiming at improving the existing methods. For instance, a significant progress was the addition of elemental sulfur to facilitate the tetrazine formation, thereby increasing the yields (**Scheme 2-1, c-iv**).[7] Despite these methods are normally effective for the synthesis of aryl tetrazines, the generation of tetrazines with aliphatic substituents is considerably harder due to the inactivated alkyl nitriles required as precursors. In that regard, Devaraj's group reported the use of certain Lewis acids, specifically Ni^{2+} or Zn^{2+} salts, effectively activated the initial nitriles and promoted the nucleophilic addition of hydrazine. This significant contribution enhanced greatly the synthetic yields for 3,6-(dialkyl)tetrazines (**Scheme 2-1, c-v**)[8] and monosubstituted (alkyl)tetrazines through reaction with formamidine acetate (**Scheme 2-1, c-vi**).[8,9]

These abovementioned procedures achieve the tetrazine formation in the last step, with the substituents in position 3 and/or 6 already in the starting molecules. Although this is a useful strategy because it normally does not involve further synthesis in the formed products, some sensitive functional groups may become an inconvenient under the reaction conditions required for these condensations. Also, the use of hydrazine, which is at the same time a good nucleophile and a reductant, may bring substrate incompatibilities.

To overcome these limitations, the presence of substituents acting as good leaving groups might be a potential strategy to achieve late-stage tetrazine functionalization. For instance, a well-known synthesis for these class of asymmetric tetrazines starts with thiocarbohydrazide, which is initially methylated to the corresponding iodide salt followed by ring closure with an orthoester (acting as a carboxylic acid synthetic equivalent), yielding the typical dihydrotetrazine that is finally oxidized (**Scheme 2-1, d**).[10] A typical drawback that was encountered in the use of thiocarbohydrazide as the tetrazine ring precursor was the formation of alternative 5-membered nitrogenated heterocylces as undesired by-products. The methylthio group prevented their formation through the deactivativation of the internal latent nitrogens.[11,12] Furthemore, it provided a handle for further derivatization. The

resulting 6-aryl or 6-alkyl-3-(methylthio)tetrazines are suitable precursors to undergo both SNAr reactions[10] or metal-catalyzed cross-couplings.[13]

Tetrazines can also be symmetrically substituted with good leaving groups. Starting from thiocarbohydrazide, 3,6-bis(methylthio)tetrazine was synthesized in a similar strategy as in the previous case,[14] and likewise 3,6-bis(3,5-dimethylpyrazol-1-yl)tetrazine was afforded starting from guanidine (**Scheme 2-1, e**).[15] Both products can participate in SNAr reactions, and can also be subsequently chlorinated to the corresponding 3,6-dichlorotetrazine,[15] which apart from being a suitable precursor in nucleophilic substitution reactions, it can also be employed in metal-catalyzed cross-couplings.[16,17] A detailed description of the role of these reagents in late-stage functionalization through SNAr reactions and C-C bond formation reactions will be displayed in **Chapter 3** and **Chapter 4**, respectively.

2.1.2. Oxidation of dihydrotetrazines:

Every route for 1,2,4,5-tetrazine synthesis requires a final oxidation step from the reduced dihydrotetrazine intermediate. Except from a limited number of cases where this oxidation can be exerted without the requirement of any reagent, normally this process is accomplished through the addition of different oxidizing agents. Over the years, several methods have been reported, although the most commonly used is the addition sodium nitrite under acidic conditions.[18,19] The generation of toxic nitrous gases and the instability of certain functional groups at low pH are the main drawbacks for this methodology.

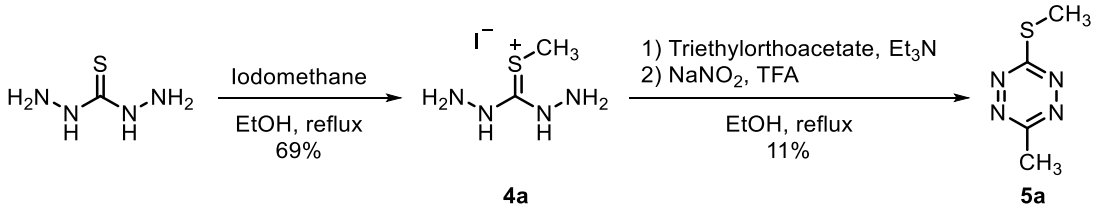
Other standard oxidizing agents have also been explored for this purpose. For instance, metal oxides, like MnO_2[20] or CrO_3,[21] afford the final tetrazines with moderate yields; m-CPBA[22] and benzoyl peroxide[23] are also potential agents; DDQ[24] can oxidize certain substrates more efficiently than $NaNO_2$, but its by-products can be challenging to purify; hypervalent iodine species, such as (diacetoxyiodo)benzene,[23] can also be exploited as oxidants under mild conditions (an application for this reagent will be discussed in **Chapter 4**). Alternatively, modern methodologies for dihydrotetrazine oxidation have also been explored. For instance, methylene blue irradiated at 660 nm wavelength or the enzyme horseradish peroxidase[25] can efficiently achieve the same conversion, although their implementation is less widespread for obvious reasons.

2.2. Synthesis of 3-bromotetrazines

2.2.1. Initial attempts on 1,2,4,5-tetrazine synthesis and functionalization:

Our first goal was to synthesize a tetrazine reagent that allowed the simple functionalization of a wide variety of small molecules and biomolecules. Ultimately, we accomplished our aim through the synthesis of 3-bromo-6-methyl-1,2,4,5-tetrazine (**1**), 3-bromo-1,2,4,5-tetrazine (**2**) and 3-bromo-6-phenyl-1,2,4,5-tetrazine (**3**). However, we initially pursued the synthesis of 6-methyl-3-(methylthio)-1,2,4,5-tetrazine because it was reported as an excellent precursor for S_NAr reactions to afford asymmetrically substituted tetrazines.

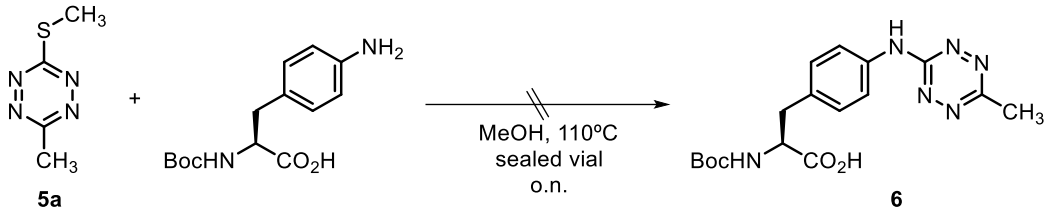
We started the synthesis of 3-methyl-6-(methylthio)-1,2,4,5-tetrazine (**5a**) following the reported route (**Scheme 2-2**).[10] The synthesis started with the commercially available thiocarbohydrazide, which was methylated to afford iodide salt **4a** as colorless crystals. Compound **4a** was then reacted with the correspondent orthoester (in this case, triethylorthoacetate) in the presence of triethylamine, followed by the *in situ* oxidation of the formed dihydrotetrazine with the usual sodium nitrite in acidic media. Under these conditions, (methylthio)tetrazine **5a** was afforded in low yield. Apart from the generation of toxic nitrous gases, the main drawback of this synthesis was the formation of the volatile tetrazine **5a**, which was difficult to purify, handle and store.



Scheme 2-2: Synthesis of 3-methyl-6-(methylthio)-1,2,4,5-tetrazine (5a).

Despite the limitations in the synthesis of **5a**, we decided to employ it in the generation of the previously published unnatural amino acid N^{α}-Boc-N^{4}-(6-methyl-1,2,4,5-tetrazin-3-yl)-4-amino-*L*-Phenylalanine (**6**) of our interest for genetic code expansion (**Scheme 2-3**).[26] Despite employing the same reported experimental procedure and conditions, we were only able to observe traces of the target product

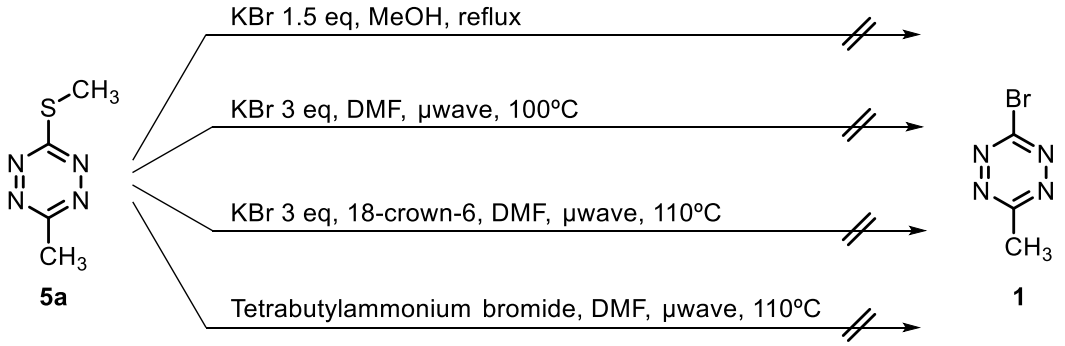
by NMR. Furthermore, the need to perform the reaction in a sealed vial limited the amount of amino acid that could be synthesized in the future.



Scheme 2-3: Failed reaction to achieve the desired phenylalanine derivative 6.

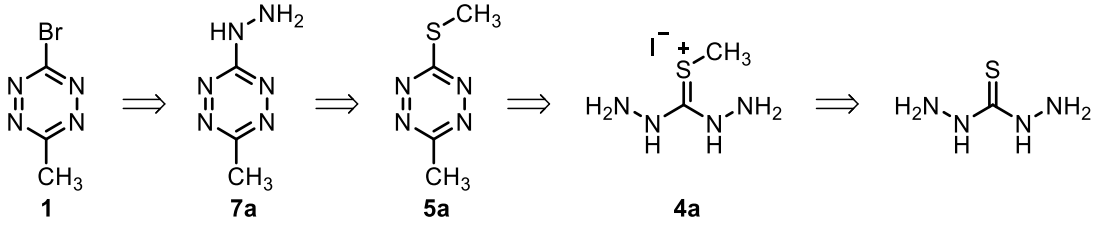
2.2.2. 3-Bromo-1,2,4,5-tetrazines as useful reagents in substitution reactions:

Considering the inefficient synthesis and the limited set of nucleophiles suitable to undergo substitution reactions with **5a**, we decided to explore other asymmetric tetrazines with higher electrophilicity. We envisaged that the use of a halogenated tetrazine would be ideal, since the corresponding halogen could be a better leaving group than the methylthio substituent. We initially focused on 3-bromo-6-methyl-1,2,4,5-tetrazine (**1**) as our target compound, and we attempted the direct substitution in compound **5a** with different nucleophilic bromide species. We tested different equivalents of potassium bromide under different solvent and temperature conditions, alone and together with 18-crown-6 to increase its solubility. Also aiming at solubilizing the potassium salt, tetrabutylammonium bromide was used, but no formation of the desired bromotetrazine **1** was observed in any of the test reactions (**Scheme 2-4**).



Scheme 2-4: Different failed attempts for the generation of bromotetrazine 1 directly from (methylthio)tetrazine 5a.

Hence, after a literature review, we found that the only report† of bromotetrazine **1** was dating back from 1978,[27] when a direct bromination from 3-hydrazinyl-6-methyl-1,2,4,5-tetrazine (**7a**) was achieved. We believed intermediate **5a** could be converted to **7a** simply by reaction with hydrazine, given its good nucleophilicity, releasing methanethiol. Similar hydrazinolysis-type reactions on tetrazines had been reported in the past.[28,29] Therefore, we designed the retrosynthetic **scheme 2-5** aiming at finding a simple, straightforward route to bromotetrazine **1**.



Scheme 2-5: Retrosynthesis for 3-bromo-6-methyl-1,2,4,5-tetrazine (1).

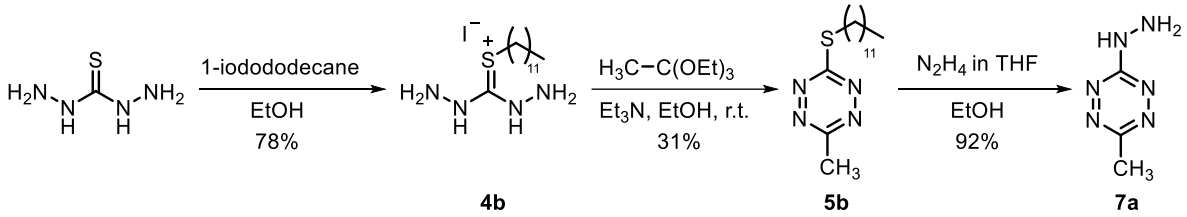
2.2.3. Improved synthesis for reagent **1**:

Still, the synthesis of (methylthio)tetrazine **5a** remained inefficient due to several factors: low overall yield, partially due to the high volatility of the final product; strong oxidation conditions needed ($NaNO_2$, TFA, reflux), which form toxic nitrous gases; and the generation of the malodorous methanethiol in the hydrazinolysis reaction. For these reasons, we envisaged that the alkylation with a longer chain in the initial reaction could solve the volatility of the (alkylthio)tetrazine intermediate by increasing its molecular weight, plus it would prevent the formation of volatile organosulfur by-products, overall generating a more user-friendly synthesis suitable for scale-up.

Therefore, we decided to alkylate thiocarbohydrazide with 1-iodododecane, as the substitution product would result in the non-volatile organosulfur dodecanethiol. Salt **4b** was afforded uneventfully, and it was then subjected to the same cyclisation procedure with triethylorthoacetate. Strikingly, (dodecylthio)tetrazine **5b** was afforded without adding any oxidant, and the addition of $NaNO_2$/TFA only resulted in a lower yield. The reasons behind this different behavior in comparison with (methylthio)tetrazine **5a** are still unclear, and could be a subject for investigation

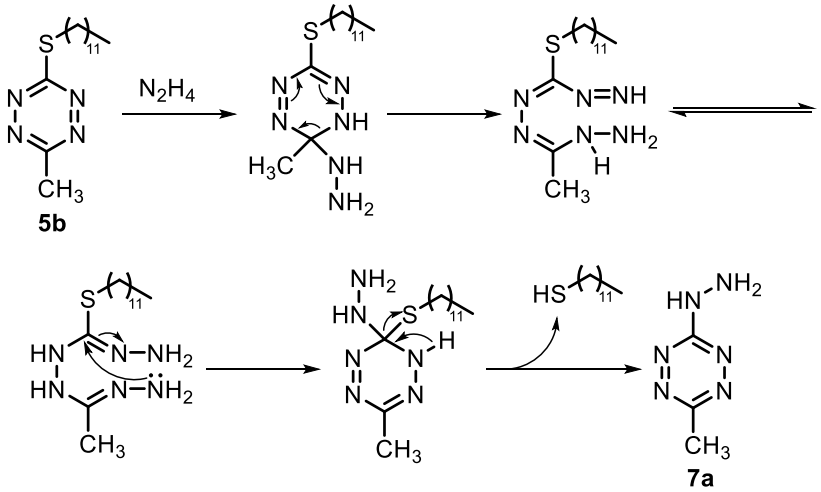
†A few months later, an article with a similar synthetic route to attain bromotetrazine **1** was published.[31]

beyond this Ph.D. thesis. However, we were pleased to observe that conversion from **4b** to **7a** was afforded with considerably improved yields (**Scheme 2-6**).



Scheme 2-6: Synthesis of 3-hydrazinyl-6-methyl-1,2,4,5-tetrazine (7a) through a (dodecylthio)tetrazine intermediate.

Despite the extra steric hindrance exerted by the bulkier dodecylthio substituent, reaction of **5b** to (hydrazino)tetrazine **7a** was afforded with an excellent yield (**Scheme 2-6**). Interestingly, a previously reported mechanistic study on the hydrazinolysis of 6-methyl-3-amino-1,2,4,5-tetrazine[27] revealed that the reaction does not proceed through a direct substitution on position 3. Instead, an unforeseen stepwise ring-opening and ring-closing mechanism, which was referred to as S_N(ANRORC), takes place.[30] Our speculation is that this reaction undergoes a similar rearrangement (**Scheme 2-7**).

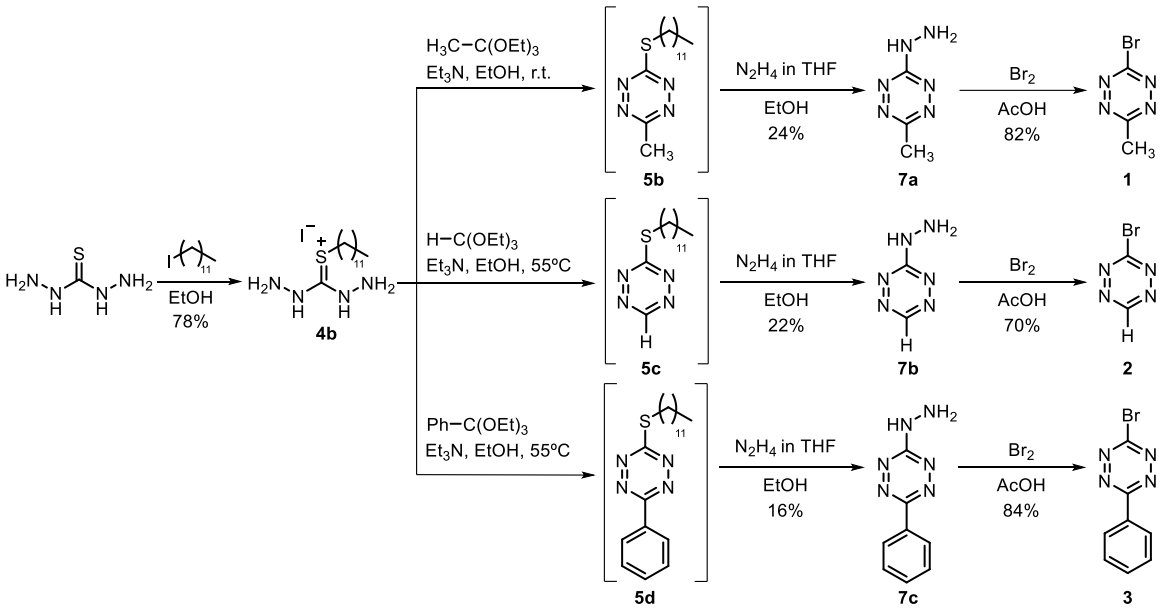


Scheme 2-7: Proposed mechanism for the hydrazinolysis from intermediate 5b to 7a through a S_N(ANRORC)[27,30] mechanism.

2.2.4. Optimized synthetic route for the differently substituted 3-bromo-1,2,4,5-tetrazines:

Consequently, following the retrosynthesis described in **scheme 2-5** with the dodecylthio substituent instead of the methylthio one, we were able to synthesize

bromotetrazine **1** with a practical, straightforward route (**Scheme 2-8, top**). After some amendments aiming at a more practical synthesis, we decided not to purify intermediate **5b**, but instead submit the crude to reaction with hydrazine in THF. The overall yield for this synthesis is of 15%, much larger than the highest reported to date (2%).[27,31] Moreover, the methodology is scalable, as it does not require metal catalysis or any oxidant. Employing an analogous route with different orthoesters we were able to synthesize the two alternative 3-bromotetrazines: using triethylorthoformate, compound **2** was synthesized through intermediates **5c** and **7b** (**Scheme 2-8, middle**); with triethylorthobenzoate, **3** was formed through **5d** and **7c** (**Scheme 2-8, bottom**). In both cases, comparable yields were obtained. Importantly, compound **2** had not been previously described in the literature.†† The use of bromotetrazine reagents **1**, **2** and **3** in small molecule functionalization trough S_NAr reactions and Pd-mediated cross-coupling reactions will be discussed in **Chapter 3** and **Chapter 4**. Likewise, their applicability for protein labeling will be highlighted in **Chapter 5**.



Scheme 2-8: Syntheses of monosubstituted 3-bromotetrazine (2) and the corresponding analogues with methyl (1) and phenyl (3) substituents.

††Our synthesis of 3-bromo-1,2,4,5-tetrazine (**2**) was published simultaneously with another article[32] that also described its synthesis through a similar route.

2.3. Summary:

We have developed an alternative method for the synthesis of 3-bromo-1,2,4,5-tetrazines that is free from oxidants and metal catalysis, does not generate volatile intermediates, and where all reactions proceed at room temperature or 55 ºC. Altogether, this has enabled the synthesis of the previously unreported simplest bromotetrazine, namely 3-bromo-1,2,4,5-tetrazine (**2**), and the production of 3-bromo-6-methyl-1,2,4,5-tetrazine (**1**) and 3-bromo-6-phenyl-1,2,4,5-tetrazine (**3**) in higher overall yields.

2.4. Bibliography:

1 A. Pinner. Über sauerstoffreie Pyrimidine. *Berichte der Dtsch. Chem. Gesellschaft* 1893, **26**, 2122–2126.

2 R. Stollé. Über die Überführung von Hydrazinabkömmlingen in heterocyklische Verbindungen. *J. für Prakt. Chemie* 1906, **73**, 277–287.

3 D. S. Liu, A. Tangpeerachaikul, R. Selvaraj, M. T. Taylor, J. M. Fox and A. Y. Ting. Diels–Alder Cycloaddition for Fluorophore Targeting to Specific Proteins inside Living Cells. *J. Am. Chem. Soc.* 2012, **134**, 792–795.

4 K. A. Hofmann and O. Ehrhart. Einwirkung von Hydrazin auf Dicyandiamid. *Berichte der Dtsch. Chem. Gesellschaft* 1912, **45**, 2731–2740.

5 H. Wu and N. K. Devaraj. Advances in Tetrazine Bioorthogonal Chemistry Driven by the Synthesis of Novel Tetrazines and Dienophiles. *Acc. Chem. Res.* 2018, **51**, 1249–1259.

6 B. L. Oliveira, Z. Guo and G. J. L. Bernardes. Inverse electron demand Diels–Alder reactions in chemical biology. *Chem. Soc. Rev.* 2017, **46**, 4895–4950.

7 P. Audebert, S. Sadki, F. Miomandre, G. Clavier, M. Claude Vernières, M. Saoud and P. Hapiot. Synthesis of new substituted tetrazines: electrochemical and spectroscopic properties. *New J. Chem.* 2004, **28**, 387–392.

8 J. Yang, M. R. Karver, W. Li, S. Sahu and N. K. Devaraj. Metal-Catalyzed One-Pot Synthesis of Tetrazines Directly from Aliphatic Nitriles and Hydrazine. *Angew. Chem., Int. Ed.* 2012, **51**, 5222–5225.

9 D. L. Alge, D. F. Donohue and K. S. Anseth. Facile and efficient Lewis acid catalyzed synthesis of an asymmetric tetrazine useful for bio-orthogonal click chemistry applications. *Tetrahedron Lett.* 2013, **54**, 5639–5641.

10 S. C. Fields, M. H. Parker and W. R. Erickson. A Simple Route to Unsymmetrically Substituted 1,2,4,5-Tetrazines. *J. Org. Chem.* 1994, **59**, 8284–8287.

11 L. M. Werbel, D. J. Mcnamara, N. L. Colbry, J. L. Johnson, M. J. Degnan and B. Whitney. Synthesis and antimalarial effects of N,N-dialkyl-6-(substituted phenyl)-1,2,4,5-tetrazin-3-amines. *J. Heterocycl. Chem.* 1979, **16**, 881–894.

12 R. Esmail and F. Kurzer. Heterocyclic compounds from urea derivatives. Part XXII. Thiobenzoylated carbonohydrazides and their cyclisation. *J. Chem. Soc. Perkin Trans.* 1975, **18**, 1781.

13 N. Leconte, A. Keromnes-Wuillaume, F. Suzenet and G. Guillaumet. Efficient palladium-catalyzed synthesis of unsymmetrical (het)aryltetrazines. *Synlett* 2007, 204–210.

14 D. L. Boger and S. M. Sakya. Inverse Electron Demand Diels-Alder Reactions of 3,6-bis(methylthio)-1,2,4,5-Tetrazine: 1,2-Diazine Introduction and Direct Implementation of a Divergent 1,2,4,5-Tetrazine → 1,2-Diazine → Benzene (Indoline/Indole) Diels-Alder Strategy. *J. Org. Chem.* 1988, **53**, 1415–1423.

15 S. G. Tolshchina, G. L. Rusinov and V. N. Charushin. 1,2,4,5-Tetrazines and Azolo[1,2,4,5]tetrazines: Synthesis and Reactions with Nucleophiles. *Chem. Heterocycl. Compd.* 2013, **49**, 66–91.

16 Z. Novák and A. Kotschy. First Cross-Coupling Reactions on Tetrazines. *Org. Lett.* 2003, **5**, 3495–3497.

17 F. Pop, J. Ding, L. M. L. Daku, A. Hauser and N. Avarvari. Tetrathiafulvalene-s-tetrazine: Versatile platform for donor-acceptor systems and multifunctional ligands. *RSC Adv.* 2013, **3**, 3218–3221.

18 R. A. Carboni and R. V. Lindsey Jr. Reactions of Tetrazines with Unsaturated Compounds. A New Synthesis of Pyridazines. *J. Am. Chem. Soc.* 1959, **81**, 4342–4346.

19 C.-H. Lin, E. Lieber and J. P. Horwitz. The Synthesis of sym-Diaminotetrazine 1a,1b. *J. Am. Chem. Soc.* 1954, **76**, 427–430.

20 D. Nhu, S. Duffy, V. M. Avery, A. Hughes and J. B. Baell. Antimalarial 3-arylamino-6-benzylamino-1,2,4,5-tetrazines. *Bioorg. Med. Chem. Lett.* 2010, **20**, 4496–4498.

21 J. Sauer, G. R. Pabst, U. Holland, H.-S. Kim and S. Loebbecke. 3,6-Bis(2H-tetrazol-5-yl)-1,2,4,5-tetrazine: A Versatile Bifunctional Building Block for the Synthesis of Linear Oligoheterocycles. *European J. Org. Chem.* 2001, 697–706.

22 S. Pican, V. Lapinte, J.-F. Pilard, E. Pasquinet, L. Beller, L. Fontaine and D. Poullain. Synthesis of 3,6-Divinyl-1,2,4,5-Tetrazine, the First Member of the Elusive Vinyltetrazine Family. *Synlett* 2009, **5**, 731–734.

23 R. Selvaraj and J. M. Fox. An efficient and mild oxidant for the synthesis of s-tetrazines. *Tetrahedron Lett.* 2014, **55**, 4795–4797.

24 M. L. Blackman, M. Royzen and J. M. Fox. Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels–Alder Reactivity. *J. Am. Chem. Soc.* 2008, **130**, 13518–13519.

25 H. Zhang, W. S. Trout, S. Liu, G. A. Andrade, D. A. Hudson, S. L. Scinto, K. T. Dicker, Y. Li, N. Lazouski, J. Rosenthal, C. Thorpe, X. Jia and J. M. Fox. Rapid Bioorthogonal Chemistry Turn-on through Enzymatic or Long Wavelength Photocatalytic Activation of Tetrazine Ligation. *J. Am. Chem. Soc.* 2016, **138**, 5978–5983.

26 J. L. Seitchik, J. C. Peeler, M. T. Taylor, M. L. Blackman, T. W. Rhoads, R. B. Cooley, C. Refakis, J. M. Fox and R. A. Mehl. Genetically Encoded Tetrazine Amino Acid Directs Rapid Site-Specific in Vivo Bioorthogonal Ligation with trans-Cyclooctenes. *J. Am. Chem. Soc.* 2012, **134**, 2898–2901.

27 A. Counotte-Potman and H. van der Plas. Degenerate Ring Transformations in Reactions of 1,2,4,5-Tetrazines with Hydrazine. *J. Heterocycl. Chem.* 1978, **15**, 445–448.

28 A. Counotte-Potman, H. van der Plas, B. van Veldhuizen and C. Landheer. Occurrence of the S_N(ANRORC) Mechanism in the Hydrazination of 1,2,4,5-Tetraines. *J. Org. Chem.* 1981, **46**, 5102–5109.

29 A. Counotte-potman and H. van der Plas. A New Synthesis of 6-(Alkyl)amino-3-aryl(alkyl)-1,2,4,5-tetrazines. *J. Heterocycl. Chem.* 1981, **18**, 123–127.

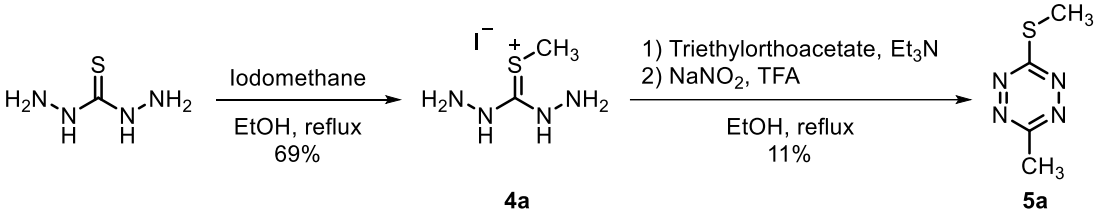
30 H. C. Van Der Plas. The S_N(ANRORC) Mechanism: A New Mechanism for Nucleophilic Substitution. *Acc. Chem. Res.* 1978, **11**, 462–468.

31 A. Wieczorek, P. Werther, J. Euchner and R. Wombacher. Green- to far-red-emitting fluorogenic tetrazine probes – synthetic access and no-wash protein imaging inside living cells. *Chem. Sci.* 2017, **8**, 1506–1510.

32 S. D. Schnell, L. V. Hoff, A. Panchagnula, M. H. H. Wurzenberger, T. M. Klapötke, S. Sieber, A. Linden and K. Gademann. 3-Bromotetrazine: labelling of macromolecules via monosubstituted bifunctional s -tetrazines. *Chem. Sci.* 2020, **11**, 3042–3047.

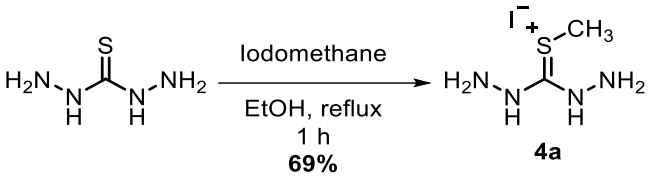
2.5. Experimental section

2.5.1. Synthesis of 3-methyl-6-(methylthio)-1,2,4,5-tetrazine (**5a**):



The synthesis was carried out following the reported procedure.[10]

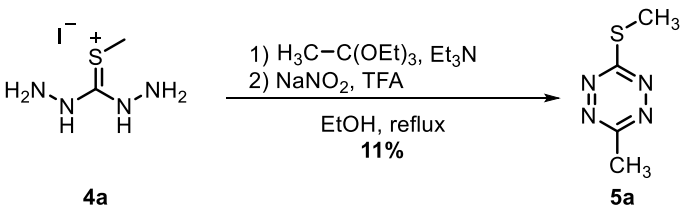
Methyl-thiocarbohydrazide iodide salt (4a):



Thiocarbohydrazide (5.00 g, 47.1 mmol, 1 eq) was suspended in EtOH (200 mL, 0.23 M), and the suspension was heated under reflux. Iodomethane (3.40 mL, 51.8 mmol, 1.1 eq) was dissolved in EtOH (200 mL) and added dropwise over 10 min. The reaction mixture was heated under reflux for 1 h, after which the resulting yellow solution was directly filtered over a C-type crucible filter. The reaction mixture was allowed to cool at room temperature overnight, thereby forming fine white crystals that were separated through filtration and dried under vacuum to obtain compound **4a**, 8.02 g, (69% yield), as colourless crystals.

^{1}H NMR (400 MHz, DMSO-d$_6$) δ 2.38 (s, 3H) ppm

3-methyl-6-(methylthio)-1,2,4,5-tetrazine (5a):

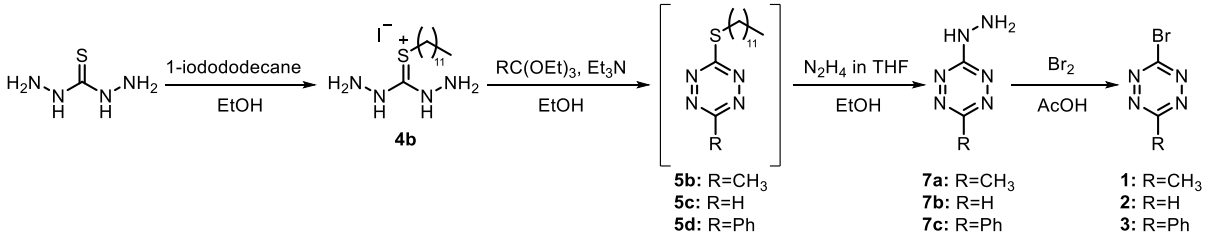


Salt **4a** (7.48 g, 30.1 mmol, 1 eq) was suspended in EtOH (200 mL, 0.15 M) at room temperature, and triethyl orthoacetate (Merck; 6.20 mL, 33.1 mmol, 1.1 eq) was added at room temperature, turning the solution clear yellow. After 5 min,

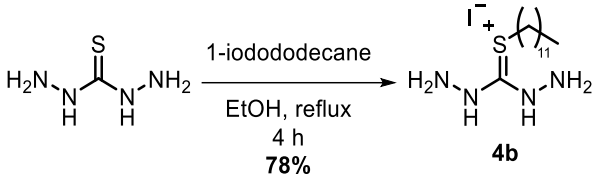
triethylamine (4.15 mL, 30.1 mmol, 1 eq) was added and the reaction mixture was stirred at reflux temperature for 30 minutes, turning the reaction mixture orange. $NaNO_2$ (4.26 g, 60.2 mmol, 2 eq) was then added, immediately followed by the dropwise addition of TFA (2.31 mL, 30.1 mmol, 1 eq), which lead to a heavy steam formation and the colour changed to deep red. After stirring for 30 minutes at reflux temperature, the reaction mixture was cooled down, hexane (200 mL) and water (150 mL) were added, and the desired compound was extracted with diethyl ether (x3). The combined organic phases were dried over $MgSO_4$ and concentrated under vacuum (product extremely volatile, mild temperature and vacuum required to improve yields). The resulting red oil was filtered through a silica pad using hexane/EtOAc (9:1) to remove polar impurities, yielding desired compound **5a**, 452 mg (11% yield), as a red oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ 2.98 (s, 3H), 2.73 (s, 3H) ppm

2.5.2. Synthesis of 3-bromotetrazines **1**, **2** and **3**:



Dodecyl-thiocarbohydrazide iodide salt (4b):



Thiocarbohydrazide (7.50 g, 69.2 mmol, 1 eq) was suspended in EtOH (350 mL, 0.2 M), and the suspension was heated under reflux. 1-iodododecane (19.15 mL, 77.6 mmol, 1.1 eq) was dissolved in EtOH (20 mL) and added dropwise over 15 min. The reaction mixture was heated under reflux for 4 h.

After cooling down the reaction mixture, the solvent was evaporated under reduced pressure, the resulting residue suspended in fresh EtOH (≈ 100 mL) and the suspension filtered to remove the unreacted thiocarbohydrazide (3.70 g, NMR consistent with starting material). The filtrate solution was evaporated, the resulting

residue suspended in hexane (≈ 100 mL), filtered again and the desired product **4b** was recovered as a white solid, which was dried in the Hvac line to afford 11.21 g (51% conversion, 78% yield).

^{1}H NMR (400 MHz, DMSO-d$_6$) δ 2.90 (t, *J* = 7.4 Hz, 2H), 1.62–1.50 (m, 2H), 1.42 – 1.17 (m, 20H), 0.85 (t, *J* = 6.9 Hz, 3H) ppm

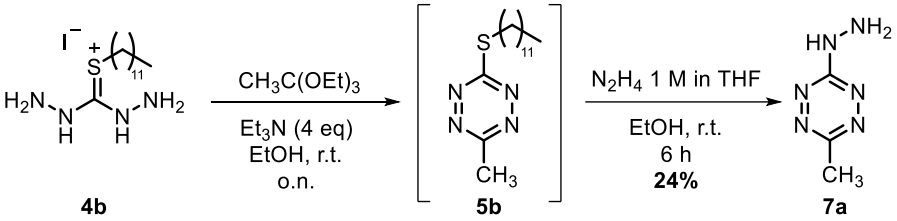
^{13}C NMR (101 MHz, DMSO-d$_6$) δ 31.7, 29.5–29.4, 29.3, 29.1, 28.9, 28.7, 28.5, 22.5, 14.4 ppm

HRMS (ESI$^+$) calculated for $C_{13}H_{31}N_4S$ $[M+H]^+$: 275.22639 / found 275.22628

Melting point (DSC): 92.9 ºC

IR (KBr) υ_{max}: 3308, 3166, 3085, 2914, 2848, 1658, 958 cm^{-1}

3-hydrazinyl-6-methyl-1,2,4,5-tetrazine (7a):



Salt **4b** (11.21 g, 27.8 mmol, 1 eq) was suspended in EtOH (200 mL, 0.14 M) at room temperature, and triethyl orthoacetate (Merck; 10.55 mL, 55.6 mmol, 2 eq) was added, turning the solution clear yellow. After 10 min, triethylamine (15.5 mL, 111.2 mmol, 4 eq) was added, immediately turning the solution light red and eventually magenta. The reaction mixture was stirred vigorously overnight at room temperature, and TLC (50:50 Hexane/EtOAc) showed the desired product (pink, front) and a by-product (yellow, baseline). The solvent was evaporated under reduced pressure, the resulting crude diluted in water and extracted with diethyl ether (x3). The combined organics layers were dried over $MgSO_4$, concentrated under reduced pressure to 20 mL. The resulting crude was purified by column chromatography (50:50 Hexane/Et_2O) to afford 4.20 g of **5b** as an impure pink oil.

To a stirred solution of **5b** (4.20 g) in EtOH (90 mL) was added hydrazine 1 M in THF (14.2 mL, 14.20 mmol, 1 eq considering **5b** as pure). The reaction mixture was stirred at room temperature and TLC (100% EtOAc) analysis showed completion after 6 h. The solvent was evaporated under reduced pressure and the resulting residue was diluted in EtOAc and dry-loaded to an 80 g silica column using a 0-100% Hexane/EtOAc gradient to obtain **7a**, 838 mg (24% from salt **4b**), as a red solid.

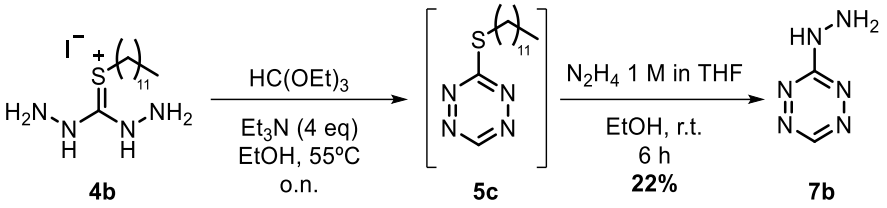
^{1}H NMR (400 MHz, DMSO-d_6) δ 9.28 (s, 1H), 4.50 (s, 2H), 2.70 (s, 3H) ppm

^{13}C NMR (101 MHz, DMSO-d_6) δ 163.9, 161.2, 20.0 ppm

HRMS (ESI$^+$) calculated for $C_3H_6N_6$ $[M+H]^+$ 127.07267 / found 127.07272

IR U_{max}: 3268, 3199, 2924, 1540, 1496, 909 cm^{-1}

3-hydrazinyl-1,2,4,5-tetrazine (7b):



An analogous methodology was followed for the synthesis of **7a.** Salt **4b** (10.40 g, 25.84 mmol, 1 eq) was suspended in EtOH (175 mL, 0.15 M) and the reaction mixture was heated at 55ºC. Triethyl orthoformate 98% (4.4 mL, 25.84 mmol, 1 eq) was added over 5 minutes, turning de reaction mixture a clear yellow solution. After 5 mins, triethylamine (14.4 mL, 103.36 mmol, 4 eq) was added, turning the reaction light red after 5 mins and eventually magenta. The reaction mixture was cooled to room temperature and stirred overnight. Then, the solvent was evaporated under reduced pressure, the resulting crude diluted in water and extracted with EtOAc (x3). The combined organics layers were dried over $MgSO_4$ and concentrated under reduced pressure. The resulting crude was filtered through silica (using EtOAc as the mobile phase) to afford, after evaporation of the filtrate, 2.93 g of **5c** as an impure pink solid.

The intermediate **5c** was then dissolved in EtOH (50 mL) and hydrazine 1M in THF (10.4 mL, 10.4 mmol, 1 eq considering **5c** as a pure intermediate) was added. The reaction mixture was stirred overnight at room temperature, and TLC (100% EtOAc) showed completion. The reaction mixture was dried under reduced pressure, and the obtained residue was adsorbed into silica to be purified by flash chromatography (0-100% EtOAc in hexane). After evaporation of the relevant fractions, **7b**, 604 mg (22% from **4b**), was obtained as a pure red solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ 9.76 (s, 1H), 6.95 (s, 1H), 4.11 (s, 1H) ppm

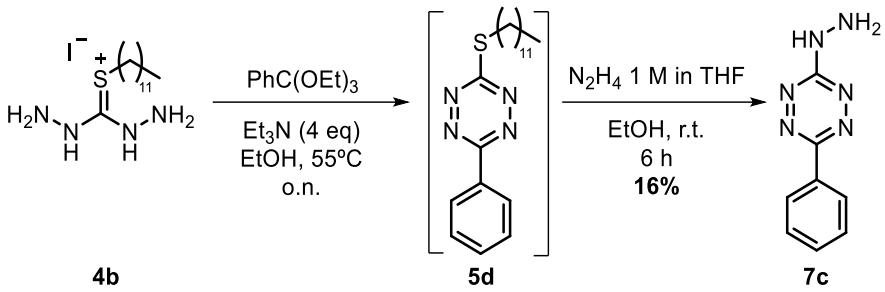
^{13}C NMR (101 MHz, CD_3OD) δ 166.22, 154.39 ppm

HRMS (ESI$^+$) calculated for $C_2H_4N_6$ $[M+H]^+$: 113.05702 / found 113.05707

Melting point: 89.0 ºC

IR (KBr) U_{max}: 3212, 2925, 1641, 1549, 1118, 957, 844 cm^{-1}

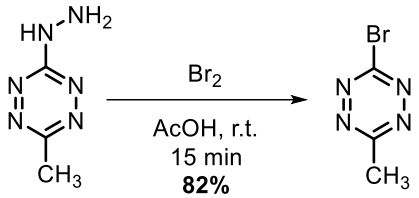
3-hydrazinyl-1,2,4,5-tetrazine (7c):



4b **5d** **7c**

Intermediate **7c** was synthesized with the same methodology as **7b**.

3-bromo-6-methyl-1,2,4,5-tetrazine (1):



7a **1**

To a stirred solution of **7a** (480 mg, 3.81 mmol, 1 eq) in 2 mL glacial acetic acid was added a solution of bromine (0.2 mL) in glacial acetic acid (4 mL) dropwise. After 15 mins, TLC (90:10 Hexane/EtOAc) showed completion. The reaction mixture was diluted in water, extracted with DCM (x3) and the combined organic layers were washed with water (x3). The organic phase was dried over $MgSO_4$ and concentrated under reduced pressure to obtain **1**, 544 mg (82% yield), as a red crystalline solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ 3.05 (s, 3H) ppm.

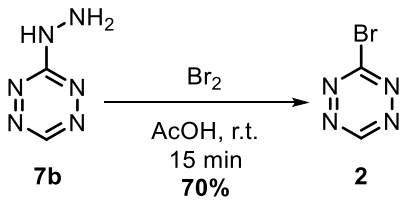
^{13}C NMR (101 MHz, $CDCl_3$) δ 168.1, 161.2, 20.6 ppm.

Melting point: 86.0 ºC.

APCI calculated for $C_3H_4BrN_4$ $[M+H]^+$: 174.9614 and 176.9593 / found: 174.9608 and 176.9588

IR (KBr) U_{max}: 2920, 2846, 1305, 1140, 865 cm^{-1}

Synthesis of 3-bromo-1,2,4,5-tetrazine (2):



7b **2**

An analogous methodology was followed for the synthesis of **2**. **7b** (602 mg, 5.37 mmol, 1 eq) was dissolved in 3 mL glacial acetic acid, and a solution of Br_2 (280 µL) in glacial acetic acid (5 mL) was added dropwise. After stirring at room temperature for 15 minutes, TLC (1:1 Hexane/EtOAc) showed completion. The reaction mixture was diluted with water and extracted with DCM (x3). The combined organics were washed with water (x3), dried over $MgSO_4$ and the solvent evaporated under reduced pressure to afford **2**, 608 mg (70% yield), as a red crystalline solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ 10.34 (s, 1H) ppm

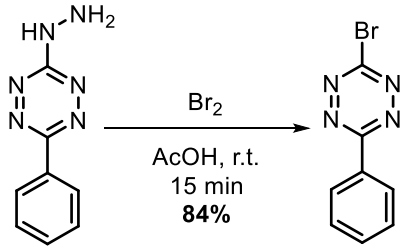
^{13}C NMR (101 MHz, $CDCl_3$) δ 164.16, 157.98 ppm

Melting point: 76.1 ºC

APCI calculated for $C_2H_4N_4Br$ (*in situ* reduction of the Tz observed): 162.9614 and 164.9593 (isotopic profile) / found: 162.9611 and 194.9589

IR (KBr) U_{max}: 3081, 2922, 2850, 2365, 2238, 1848, 1736, 1686, 1485, 1388, 1370, 1225, 1197, 1167, 1126, 1091, 1033, 882 cm^{-1}

Synthesis of 3-bromo-6-phenyl-1,2,4,5-tetrazine (3):



An analogous methodology was followed for the synthesis of **3**. **7c** (203 mg, 1.08 mmol, 1 eq) was dissolved in 0.6 mL glacial acetic acid, and a solution of Br_2 (56 µL) in glacial acetic acid (1 mL) was added dropwise. After stirring at room temperature for 15 minutes, TLC (1:1 Hexane/EtOAc) showed completion. The reaction mixture was diluted with water and extracted with DCM (x3). The combined organics were washed with water (x3), dried over $MgSO_4$ and the solvent evaporated under reduced pressure to afford **3**, 220 mg (84% yield), as a red crystalline solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.60 – 8.53 (m, 2H), 7.71 – 7.56 (m, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 164.74, 161.11, 133.62, 130.61, 129.64, 128.49 ppm

Chapter 3

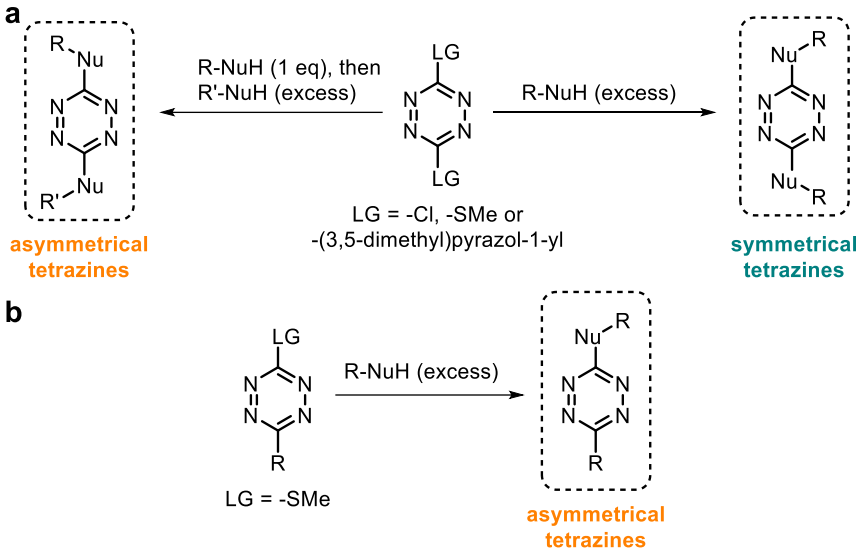
Small molecule functionalization with 3-bromo-1,2,4,5-tetrazines through nucleophilic aromatic substitutions

Synthesis, fluorescence and use as probes

3.1. Tetrazine functionalization through S_NAr reactions

3.1.1. Background:

As it has been highlighted in **Chapter 2**, tetrazines can be functionalized through S_NAr reactions starting from adequate precursors. For this purposes, either 3,6-dichloro-1,2,4,5-tetrazine, 3,6-bis(metylthio)-1,2,4,5-tetrazine or 3,6-bis(3,5-dimethylpyrazol-1-yl)-1,2,4,5-tetrazine have been typically used as electrophilic tetrazines to be substituted by different nucleophilic functional groups (**Scheme 3-1, a**). Either symmetrical or asymmetrical tetrazines can be synthesized employing these reagents, by controlling the number of equivalents of the nucleophiles. This is a useful approach to attain tetrazines bearing EDGs in position 3 and/or 6, which have been employed, for instance, in peptide stapling,[1] to generate fluorescent probes,[2] or to develop novel explosive materials.[3]



Scheme 3-1: Tetrazine functionalization through S_NAr reactions. a) With symmetrical electrophilic tetrazine precurors. **b)** With asymmetrical electrophilic tetrazine precursors.

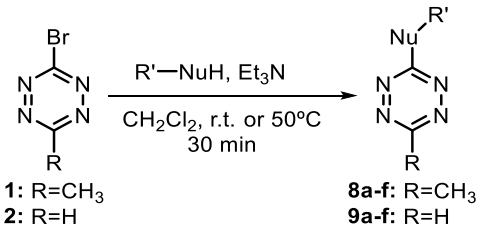
Alternatively, a simpler method to produce asymmetrically substituted tetrazines via a single S_NAr reaction relies on a different class of electrophilic precursors, such as 3-alkyl- or 3-aryl-6-(methylthio)-1,2,4,5-tetrazines,[4] which upon reaction with a given nucleophile generate the substitution product and release the unpleasant methanethiol as a gas (**Scheme 3-1, b**). In particular, 3-methyl-6-(methylthio)-1,2,4,5-tetrazine has been used to generate probes for antibody bioconjugation,[5] for

bioorthogonal labelling of membrane proteins[6], and in the synthesis of a ncAA for GCE.[7] However, the volatility of the compound is a drawback in terms of its synthesis and storage. Moreover, while the substitution with good nucleophiles such as primary or secondary amines normally works well for (methylthio)tetrazines, the substitution with worse nucleophiles is less effective (i.e.: anilines) or completely inefficient (i.e.: alcohols and phenols).

3.1.2. Results:

With the two halogenated tetrazines **1** and **2** in hand, we decided to investigate their potential for being precursors for tetrazine functionalization through S_NAr reactions. We anticipated that the strong electrophilicity of these tetrazines, together with the presence of bromine as a better leaving group than methanethiol, would facilitate the substitution with both strong and weak nucleophiles. Therefore, we tested the reactivity of **1** and **2** against different aliphatic and aromatic amines, alcohols and thiols (**Table 3-1**).

Table 3-1: S_NAr reaction of 1 or 2 with different functionalities acting as nucleophiles. Reactions were performed with 1.2 eq of nucleophile (R'-NuH) in DCM.



	-R	R'-	-Nu-	Yield (%)
8a	-CH_3	BocHN	-NH-	76
9a	-H			69
8b	-CH_3	MeO	-NH-	70
9b	-H			50
8c	-CH_3	BocHN	-O-	79
9c	-H			60
8d	-CH_3	Cl	-O-	92
9d	-H			56
8e	-CH_3	H_3C (11)	-S-	89
9e	-H			67
8f	-CH_3	F	-S-	88
9f	-H			51

As expected, aliphatic amines and anilines reacted at room temperature to afford the desired products **8a-b** and **9a-b** with moderate to good yields. Likewise, aliphatic alcohols and phenols also afforded ethers **8c-d** and **9c-d** with similar yields, but the temperature of the reaction had to be increased to 50 ºC; finally, aliphatic thiols and thiophenols yielded thioethers **8e-f** and **9e-f**, also at room temperature.

3.2. Fluorescence spectroscopy of heteroatom-substituted tetrazines:

Encouraged by the efficient synthesis of these asymmetric tetrazines substituted with different heteroatoms, we decided to exploit their properties aiming at developing new tetrazine-based tools with bioorthogonal applications. To that end, we focused on existing reports referring to the fluorescent behavior of tetrazines bearing oxygen substitutions.[2,8,9] Aided by Dr. Laura Rodriguez' laboratory at the Inorganic section of the Department of Inorganic and Organic Chemistry of the Universitat de Barcelona, we performed a fluorescent spectroscopic characterization of the previously described tetrazinyl compounds.

Due to substrate availability, we focused on compounds **8a-f** derived from 3-bromo-6-methyl-1,2,4,5-tetrazine **1**. The absorption and emission spectra for each compound were acquired (**Fig. 3-1**). From them, the absorption (λ_{Abs}) and emission (λ_{Em}) maxima, and the difference between these two values (Stokes shift), were measured. Also, the molar extinction coefficient (ε) and the fluorescence quantum yield (Φ_F), defined as the ratio of photons absorbed to photons emitted through fluorescence, were determined (**Table 3-2**). Our results were in concordance with previous reports, and only ether-bridged tetrazines showed a moderate fluorescent behavior (Φ_F > 12% for **8c** and **8d**), with absorption wavelengths in the green region (~ 530-540 nm) and emissions at 580-590 nm.

Table 3-2: Fluorescence spectroscopy of compounds 8a-f. *Quantum yield determination performed in DCM using Rhodamine B in ethanol as reference.

	-Nu-	λ_{abs} (nm)	ε ($M^{-1}cm^{-1}$)	λ_{Em} (nm)	Stokes shift (nm)	Φ_F* (%)
8a	-NH-	531	416	587	56	0.3
8b		533	303	572	39	0.6
8c	-O-	536	696	581	45	12.2
8d		539	688	588	49	12.4
8e	-S-	537	550	586	49	1.6
8f		537	554	593	56	1.0

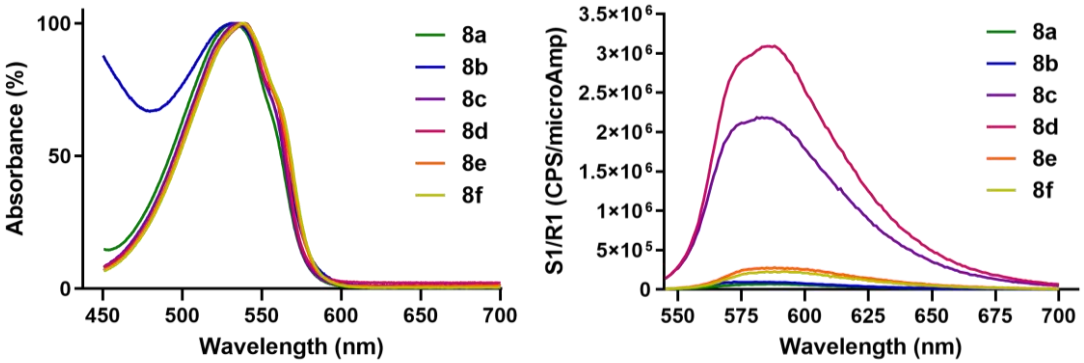
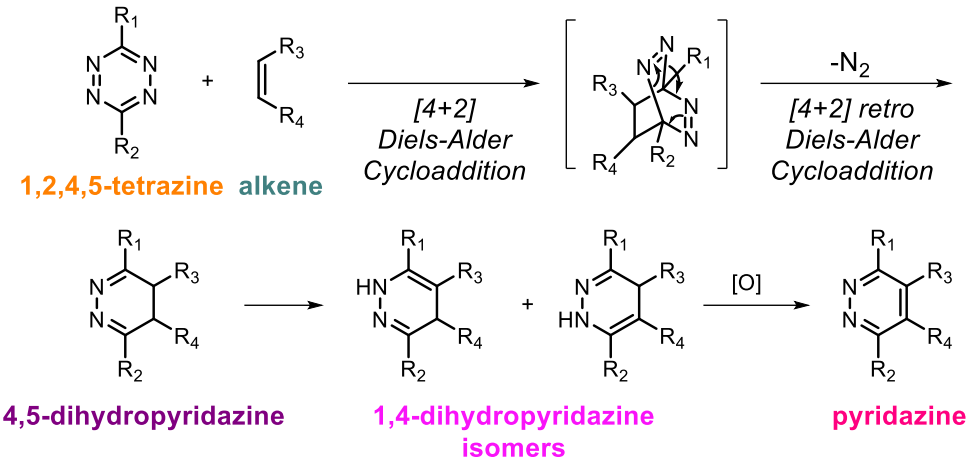


Figure 3-1: Absorption (left) and emission (right) spectra of compounds **8a-f**.

3.3. Kinetics in the iEDDA cycloaddition

3.3.1. Background:

The main reactivity of 1,2,4,5-tetrazines is in iEDDA cycloadditions with certain alkenes or alkynes (**Scheme 3-2**). These two functionalities react through an initial [4+2] cycloaddition, leading to a bicyclic intermediate that undergoes a retro Diels-Alder with the concomitant release of a molecule of elemental nitrogen (N_2). If the initial cycloaddend is an alkyne, the product after the retro Diels-Alder is a fully oxidized pyridazine. However, if the initial reactant is an alkene, the obtained product is a 4,5-dihydropiridazine, which isomerizes to both possible 1,4-dihydropyridazines, finally spontaneously oxidizing to the final pyridazine core (the importance of these reduced intermediates will be highlighted in **Chapter 6**). Importantly, the iEDDA cycloaddition is highly stereo-, regio- and chemoselective.[10]



Scheme 3-2: iEDDA cycloaddition mechanism between a 1,2,4,5-tetrazine and a suitable alkene.

The iEDDA bioorthogonal reaction has been applied as an extremely useful chemical biology tool both for research and therapeutic purposes. For instance, many applications in the imaging field have been developed, aiming mostly at developing fluorophores that do not require washing steps during sample preparation (more details will be provided in **Section 3.4.1.**).[11] Furthermore, the use of this reaction to develop new drug delivery approaches[12] or to improve the pharmacokinetic profile of certain drugs[13] is a continuously advancing field. All these applications take advantage of one of the most important properties of the iEDDA cycloaddition: its extremely fast kinetics compared to other bioorthogonal reactions.

Our aim was to apply our functionalized tetrazines, in particular the fluorescent ones, in bioorthogonal reactions. To this end, the kinetic parameters of these compounds in the iEDDA cycloaddition is the most critical parameter to consider. As it has been pointed out in **Chapter 1**, the suitability of a bioorthogonal reaction is directly proportional to its reaction kinetics: if the reactants are too slow at the required low biological concentrations, the reaction will not be compatible with biological time-frames.

The kinetic parameter to compare different reactions is the rate constant, which in the case for 2nd order reactions (the type that will be detailed herein, including the iEDDA cycloaddition) can be defined as the 2nd order rate constant (k_2). Some bioorthogonal reactions are highly specific in biological environments but display sluggish kinetics, thus becoming their main limitation to be used *in vivo*. Staudinger ligations, for instance, present k_2 values of 10^{-4}-10^{-3} $M^{-1}s^{-1}$, requiring several hours to

react at significant conversion efficiencies at physiological concentrations (**Fig. 3-2**).[14] iEDDA cycloadditions, on the other hand, have the advantage of being the fastest bioorthogonal reactions described to date, normally exhibiting k_2 values between 1-10^5 $M^{-1}s^{-1}$.[14]

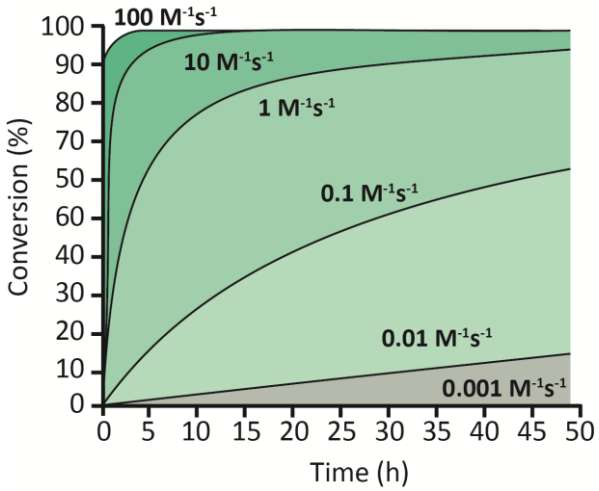


Figure 3-2: Simulation of 2nd order reactions between reactants at 100 µM concentration. Adapted from Oliveira *et al.* (reference 10).

The kinetics of Diels-Alder-type reactions are determined by several factors, but the most critical are: a) the electronic characteristics of the reacting groups, b) the strain of the dienophile, c) the steric hindrance between the cycloaddends, and d) the solvent in which the reaction takes place.

Firstly, the predominant factor is the energy gap between the frontier highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) of the diene and the dienophile (**Fig. 3-3**).[10,15] When the energy difference between the two frontier molecular orbitals is smaller, the subsequent reaction is faster. In the case of the nEDDA, the presence of EDGs in the diene and EWGs in the dienophile decrease the energy gap, and are therefore favorable for the reaction to take place. Conversely, in the case of the iEDDA, the opposite substituent effect can be observed: EWGs in the diene (tetrazine) and EDGs in the dienophile (alkene or alkyne) enhance the reaction rate of the cycloaddition. From this principle, several reports have exemplified how the reactivity of tetrazines can be fine-tuned depending on the electronic effects exerted by the substituents in positions 3 and 6.[16,17]

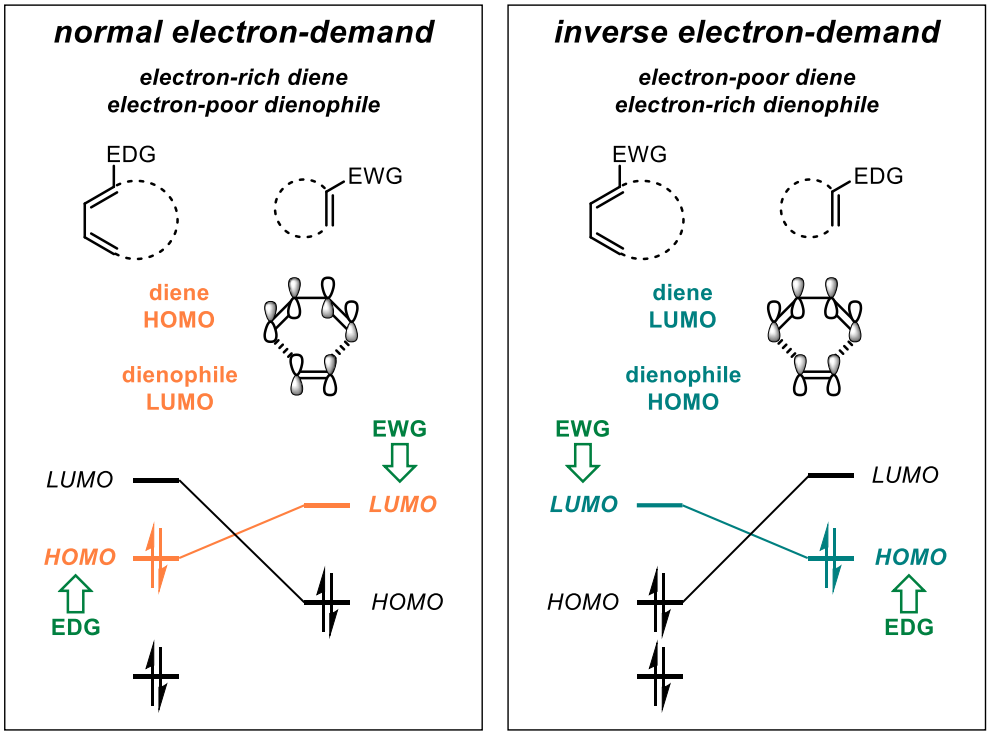


Figure 3-3: Frontier molecular orbital theory for Diels-Alder reactions. Kinetics of the nEDDA and iEDDA cycloadditions depend on the energy gap between the frontier molecular orbitals. Adapted from Mayer and Lang (reference 11).

Secondly, the ring tension on the cyclic alkenes or alkynes acting as dienophiles is also of utmost importance for the kinetics of the iEDDA cycloaddition.[18] Indeed, strained dienophiles are more reactive because of their pre-distorted conformations, thus requiring less distortion energies to react and proceed to the transition state.[19,20] A typical example for this effect is cyclooctene: the *trans*- isomer (*trans*-cyclooctene, TCO) adopts a 'crown-like' conformation, much more reactive than *cis*-cyclooctene, which is found in a 'half-chair' conformation. Alternatively, ring strain also results in an electronic effect, since it raises the HOMO of the dienophile just like EDGs do.

Another factor influencing the reaction rate of tetrazines is the steric hindrance. Bulky substituents, like *tert*-butyl groups, slow down the iEDDA reaction due to unfavorable steric repulsion with the dienophile compared, for instance, to methyl groups.[21]

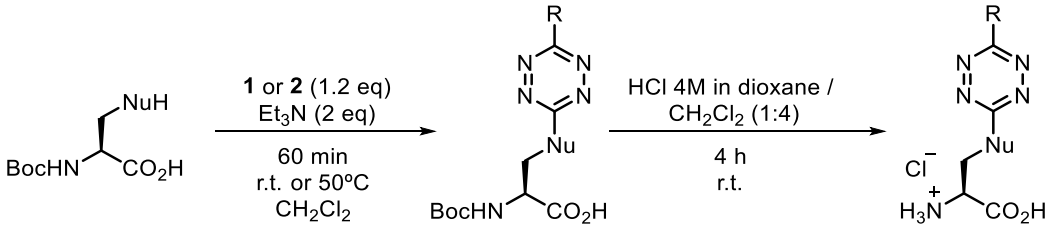
Finally, the solvent is also of major importance when considering reaction rates. Protic solvents, and specially water, accelerate the cycloaddition partially due to a facilitated hydrophobic interaction of the diene and dienophile, but also due to a smaller energy gap between HOMO and LUMO through the induction of polarity by hydrogen-bonding with the tetrazine ring.[22,23]

3.3.2. Results:

Due to the importance of reaction rate when considering bioorthogonal applications, we initially decided to determine the k_2 of the previously obtained **8a-f** and **9a-f**. Our first consideration was to perform these experiments in water, as it is the universal solvent of biological systems and, as it has been previously mentioned, accelerates the iEDDA cycloaddition. However, since model compounds **8a-f** and **9a-f** were predicted to be poorly soluble in aqueous media, we decided to synthesize a set of unnatural amino acids, which tend to be highly water-soluble due to their zwitterionic character. Moreover, we anticipated that having an available set of tetrazinyl-amino acids in hand would be useful for future *in vitro* determinations (see **Chapter 6**) or to test their incorporation in proteins *in vivo* through GCE (see **Chapters 7, 8** and **9**)

We therefore derivatized natural amino acids bearing nucleophilic groups in their side chains with bromo-tetrazines **1** or **2** (**Table 3-3**).

Table 3-3: Unnatural amino acid synthesis via S_NAr reaction of 1 and 2. *Combined yield for both steps.



N-Boc-(10a-e): R = CH_3
N-Boc-(11a-e): R = H

10a-e: R = CH_3
11a-e: R = H

	Derived from	Nu	-R	Yield* (%)
10a	Lys	-C_3H_6-NH-	-CH_3	81
11a			-H	70
10b	*p*-aminoPhe	-C_6H_4-NH-	-CH_3	80
11b			-H	42
10c	Ser	-O-	-CH_3	20
11c			-H	25
10d	Tyr	-C_6H_4-O-	-CH_3	72
11d			-H	23
10e	Cys	-S-	-CH_3	58
11e			-H	29

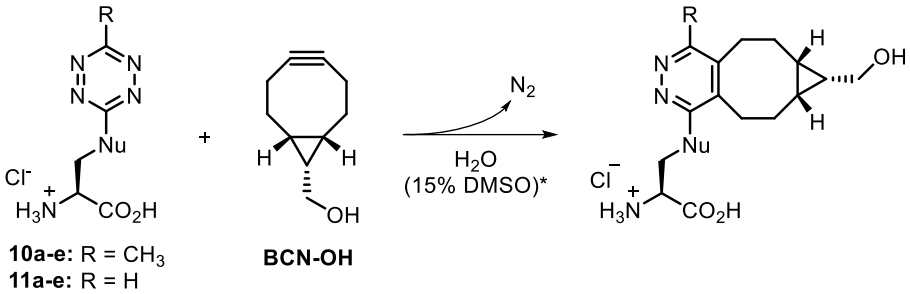
Since the alpha nitrogen (N^{α}) in the amino acid could also be a potential substitution site, we started from N^{α}-Boc protected amino acids for the S_NAr reaction, followed by a Boc-deprotection step with HCl to afford the hydrochloride salts as the final products.

Under these conditions, we synthesized lysine-derivatives **10a** and **11a**, *para*-amino-phenylalanine-derivatives **10b** and **11b**, serine-derivatives **10c** and **11c**, tyrosine-derivatives **10d** and **11d**, and cysteine-derivatives **10e** and **11e**, which were used as model compounds of tetrazine-substituted aliphatic and aromatic amines and alcohols, and aliphatic thiols (since no commercially available thiophenol-containing amino acid could be found).

With the tetrazinyl-amino acids **10a-e** and **11a-e** in hand, we explored different techniques to determine experimentally their kinetic rate constant. For fast iEDDA cycloadditions, the method of choice is normally stopped-flow spectroscopy, as it provides the most accurate data for reactions that take place within a few seconds. However, since our set of tetrazines were substituted with EDGs, we were expecting them to react slower. Hence, we investigated their kinetic parameters using a UV-Vis spectrometer, exploiting the color loss upon reaction with a suitable dienophile: 1,2,4,5-tetrazines are always seen as red, magenta or pink compounds, but the resulting pyridazines after the iEDDA reaction are normally colorless. Through the decay of the absorption at a specific wavelength, the speed of the reaction can be determined.

As the iEDDA cycloaddition is a 2nd order reaction, its k_2 can be approximated under pseudo-1st order conditions, using known concentrations of the dienophile of choice in excess. By plotting the different observed rate constants (k_{obs}), the slope of the reaction can be extrapolated as the k_2. Following this approach, we determined the k_2 for the majority of synthesized tetrazinyl-amino acids in reaction with **BCN-OH** (**Table 3-4**). **BCN-OH** was used as the model dienophile because it is known as one of the fastest strained alkene/alkyne in iEDDA cycloadditions. Since we envisaged that our tetrazines substituted with EDGs would be relatively slow (compared to other tetrazines bearing EWGs), a highly reactive dienophile was necessary.

Table 3-4: 2nd order rate constant (k_2) determination of tetrazinyl-amino acids 10a-e and 11a-e in reaction with BCN-OH. *15% DMSO was added to the reaction to solubilize **BCN-OH**. n/d: reaction was too fast to be accurately determined in the used experimental conditions.



	Nu	R	λ_{abs} (nm)	k_2 ($M^{-1}s^{-1}$)
10a	-C_3H_6-NH-	-CH_3	503	0.04
11a		-H	505	1.33
10b	-C_6H_4-NH-	-CH_3	508	0.53
11b		-H	510	n/d
10c	-O-	-CH_3	527	0.38
11c		-H	518	n/d
10d	-C_6H_4-O-	-CH_3	512	10.23
11d		-H	510	n/d
10e	-S-	-CH_3	515	0.41
11e		-H	513	1.77

Interestingly, *O*-substituted tetrazines **10d** and **11d** reacted faster than their nitrogen analogues, **10a** and **11a** respectively. Aromatic chains (**10b** and **10d**) improve the kinetics of the cycloaddition when compared to the aliphatic ones (**10a** and **10c**). Also, due to the steric hindrance generated by the methyl substituent compared to the monosubstituted tetrazines, products from **1** react slower than products of **2** (**10a** compared to **11a**; **10e** compared to **11e**). Importantly, the k_2 for ethers observed for the tyrosine derivative (**10d**, 10.23 $M^{-1}s^{-1}$) is in the range of some of the most commonly used bioorthogonal reactions.

3.4. Ether-bridged tetrazines as fluorescent probes for bioorthogonal labeling

3.4.1. Background:

The use of tetrazines as bioorthogonal fluorescent probes for research purposes, most usually in imaging experiments, has been an intense topic for investigation in the last 15 years.[11] Typically, tetrazines are used to quench the emission of the fluorophores to which they are coupled to. This effect can happen either by photoinduced electron transfer (PET), Förster resonance energy transfer (FRET) or through-bond energy transfer (TBET).[24] Indeed, several tetrazinyl-conjugates with a range of fluorophores such as BODIPY,[25,26] coumarin[27], xanthene[28], phenoxazine[29], siliconrhodamine[30] or cyanine[31] derivatives have been developed. Upon iEDDA cycloaddition with a given dienophile, the fluorophore is released from the quench exerted by the tetrazine and can emit, thus acting as a *fluorogenic* reaction or a *turn-on* fluorescent system (**Fig. 3-4**).[32,33] This approach is advantageous for *in vivo* imaging experiments, because there is no requirement for washing steps of the excess fluorophores.[33]

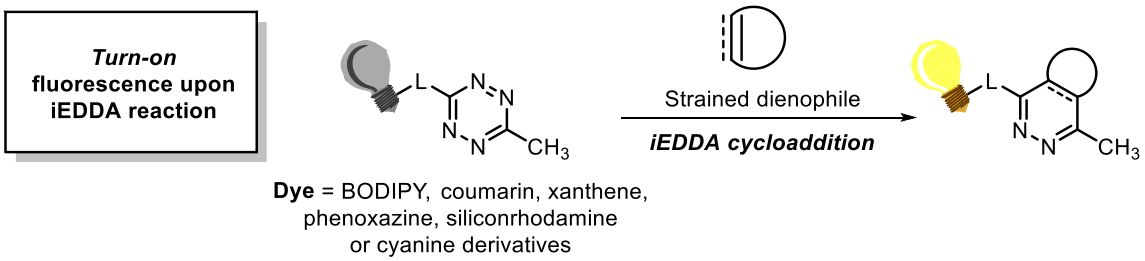


Figure 3-4: Fluorogenic reaction of tetrazinyl conjugates with a variety of fluorophores (shaped as a lightbulb) through different linkers (L). The fluorescence of BODIPY, coumarin and xanthene derivatives is quenched by the tetrazine core, but released by reaction with a strained dienophile.

In contrast, intrinsically fluorescent tetrazines are far less common. Being initially reported in the 1960s,[34] fluorescent tetrazines are mostly achieved by the reaction of alcohols with 3,6-dichloro-1,2,4,5-tetrazine[35]. In this fashion, both symmetrical and unsymmetrical ether-bridged tetrazines have been reported and their physicochemical properties characterised.[8,9,36] However, there are no precedents on the use of fluorescent tertazines to obtain a

biological read-out. Furthermore, their reactivity as bioorthogonal reagents had never been investigated.

3.4.2. Results:

When we considered the bioorthogonal applicability of the emissive tetrazines synthesized from the reaction between alcohols or phenols and **1**, there were different factors that prompted us to use the tyrosine derivative **10d** as a suitable candidate molecule. Firstly, its efficient synthesis over 2 reactions, starting from **1** (72% yield). Also, its excellent aqueous solubility and the convenience that its synthetic intermediate, **N-Boc-10d**, could be used directly in solid-phase peptide synthesis (SPPS) to couple it to a peptide of choice. Finally, it presented the highest k_2 out of all the tested tetrazinyl-amino acids (10.23 $M^{-1}s^{-1}$) and it maintained the fluorescent nature of ether-bridged tetrazines (λ_{Abs} = 511 nm and λ_{Em} = 580 nm, Φ_F = 10.4%). Interestingly, this fluorescence disappeared upon iEDDA reaction due to the destruction of the tetrazine core and formation of the resulting pyridazine. Therefore, we envisioned a system in which, instead of a fluorescence *turn-on*, a *turn-off* effect was observed upon iEDDA cycloaddition (**Fig. 3-5**).

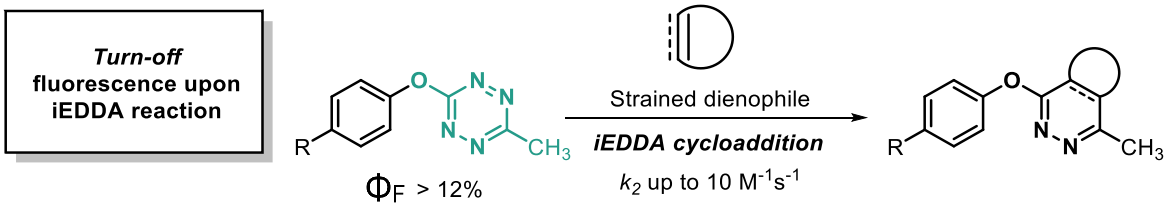


Figure 3-5: Fluorescence *turn-off* upon iEDDA cycloaddition. Intrinsically fluorescent ether-bridged tetrazines derived from a phenol lose their fluorescent through reaction with a strained dienophile.

In order to determine whether **10d** could be a suitable compound to use *in vivo*, we determined its stability in PBS pH 7.4 and FBS (**Fig. 3-6**). In this regard, more than 50% of the sample remained in solution after 48h of incubation at 37 ºC in both cases, thereby indicating that this derivative shows sufficient stability to be used in *in vivo* settings, where a trade-off between reactivity and stability is needed.[37]

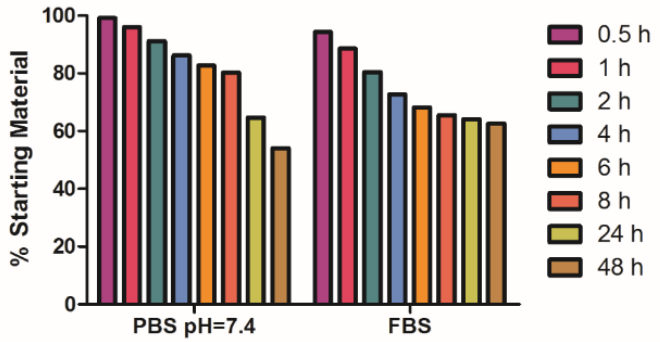


Figure 3-6: Stability of amino acid 10d. Experiment carried out in PBS pH 7.4 and FBS, in both cases at 37 ºC.

Next, in order to showcase the potential of these intrinsically fluorescent tetrazines in a biologically relevant setting, we synthesized an octreotide (**OCT**)[38] derivative by coupling **N-Boc-10d** to its *N*-terminal residue (**OCT-10d**) using SPPS (**Fig. 3-7, a;** see **Experimental Section 3.6.1.3.** for technical details). Octreotide is a marketed drug with affinity for Somatostatin receptors type 2 (SSTR2) used in several endocrine disorders.[39] As expected, the spectroscopic characteristics of **OCT-10d** (λ_{abs}=513 nm and λ_{em}=579 nm) matched those observed for **10d** alone, and the fluorescence was equally lost by reaction with **BCN-PEG**, a water-soluble version of **BCN-OH** (**Fig. 3-7, b**).

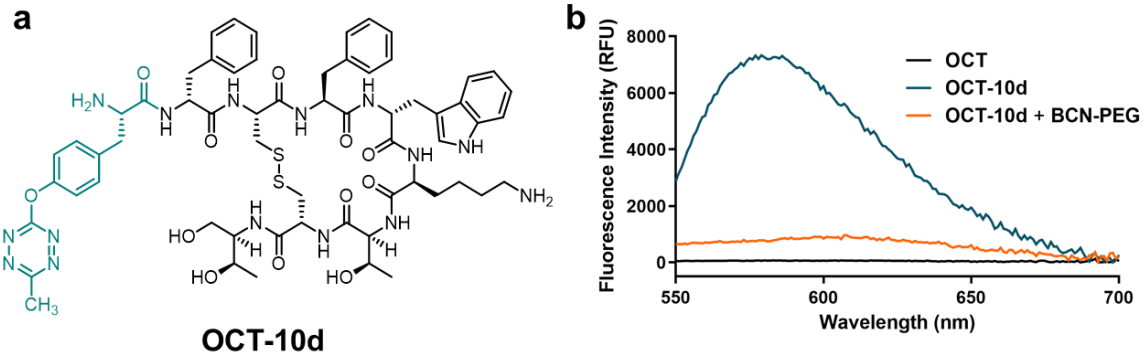


Figure 3-7: OCT-10d spectroscopic characteristics. a) Structure of **OCT-10d** (parental **OCT** shown in black, **10d** in blue). **b)** Emission spectra of **OCT**, **OCT-10d** and **OCT-10d** in which **BCN-PEG** in excess was added.

As a cellular model to test the functionality of **OCT-10d**, we used Chinese Hamster Ovary (CHO)-K1 cells with overexpression of SSTR2 receptors. We initially attempted to exemplify the dual fluorescent and bioorthogonal character of **OCT-10d** in imaging experiments, but we were unable to obtain clear microscopy images were **OCT-10d** was clearly detectable. This could be due to an inefficient internalisation of the peptide by the cells, or due to a weak emission signal that could not be distinguished

from the auto-fluorescent cellular background. For this reason, we used flow cytometry, as we envisaged it would be more sensitive to mild fluorescent signals.

Therefore, after an overnight incubation of CHO-K1 cells with **OCT-10d** (**Fig. 3-8, a**), we were able to detect an increase in the total relative fluorescence in comparison with the control cells treated with **OCT** (**Fig. 3-8, b**).

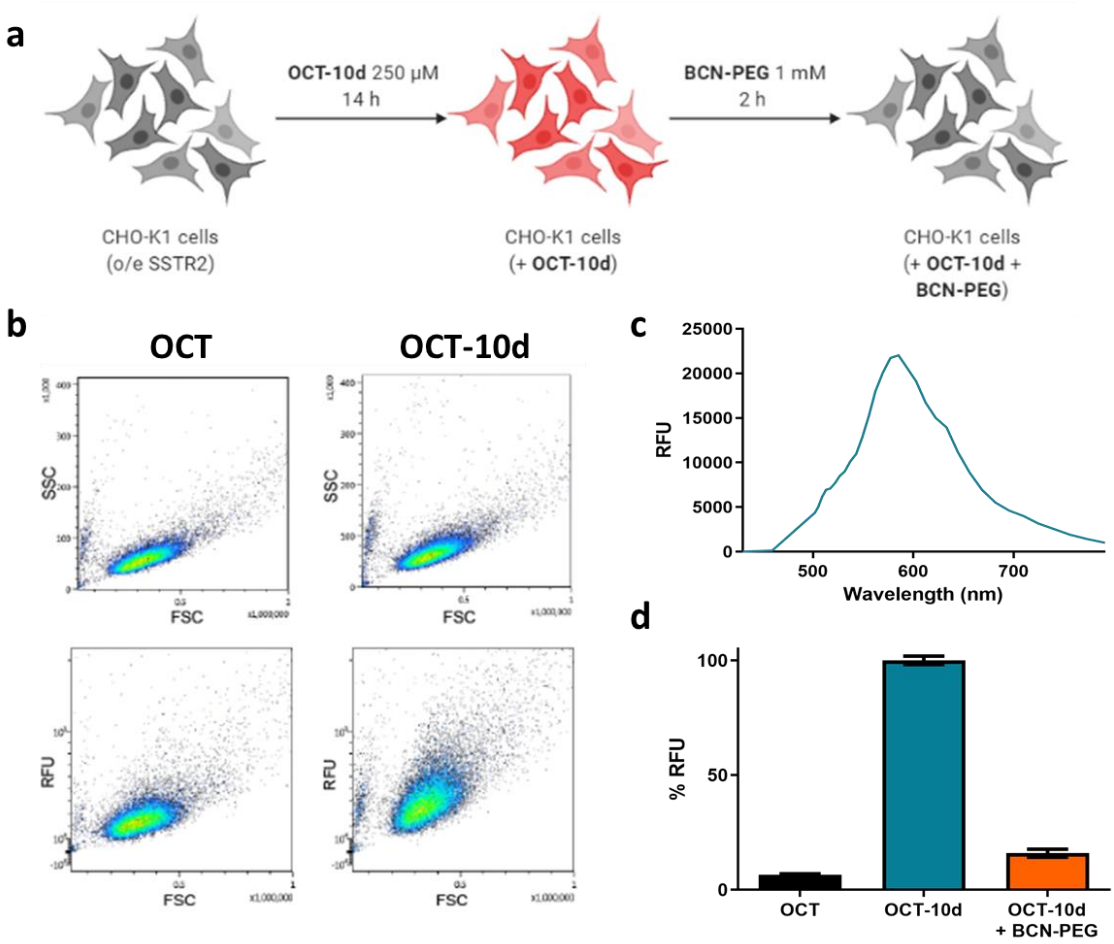


Figure 3-8. Fluorescence *turn-off* upon iEDDA bioorthogonal reaction of OCT-10d in CHO-K1 cell culture. a) Experimental procedure for the fluorescence detection of **OCT-10d**, and its subsequent loss upon bioorthogonal reaction with **BCN-PEG**. **b)** Flow cytometry density plots of CHO-K1 cells treated with **OCT** and **OCT-10d. c)** Mean emission spectra of cells treated with **OCT-10d**. **d)** Quantification of the emission signal of cells treated with **OCT**, **OCT-10d** and **OCT-10d** + **BCN-PEG** at the maximum emission wavelength (585.3 nm). SSC = Side Scatter, FSC = Forward Scatter, RFU = Relative Fluorescence Units.

We determined the maximum of the averaged emission spectra of the **OCT-10d** treated cells, which matched the expected (λ_{Em}=585.3 nm, **Fig. 3-8, c**), and quantified the relative fluorescence of the control peptide (**OCT**) and the intrinsically fluorescent peptide (**OCT-10d**). The iEDDA reaction was then performed with the water-soluble

dienophile **BCN-PEG**, which led to a fluorescence loss (**Fig. 3-8, d**), demonstrating that the iEDDA reaction was indeed taking place bioorthogonally in living cells.

In summary, we have reported that the S_NAr of 3-bromo-tetrazines **1** and **2** with aromatic or aliphatic amines, alcohols or thiols proceeds rapidly under mild conditions and with good yields, thereby providing a facile and efficient method for late-stage functionalization of small molecules through S_NAr reactions.

The corresponding oxygen-substituted tetrazines derived from alcohols or phenols showed a remarkable fluorescent behaviour that is not present in the nitrogen or sulphur analogues. Such fluorescence is lost after the corresponding iEDDA reaction.

We have studied the kinetics of the iEDDA reaction in water using a range of tetrazinyl-containing unnatural amino acids. Our results reveal that the tyrosine-derived compound (**10d**), containing a tetrazine-ether, showed a k_2 in a suitable range for rapid bioorthogonal labelling. In cellular experiments, we used amino acid **10d** as a fluorescent probe that loses its emission signal after the iEDDA reaction. We incorporated **10d** into a model peptide (octreotide) and used it in cell flow cytometry, where we could *turn-off* its fluorescence upon reaction with a strained dienophile. In conclusion, we have developed a straightforward strategy to achieve dual fluorescent and bioorthogonal probes derived from the reaction bromo-tetrazines with alcohols that can be used to synthesize useful chemical biology tools.

3.5. Summary:

3-bromotetrazines **1** and **2** are excellent electrophilic reagents to undergo S_NAr reactions with aliphatic and aromatic heteroatoms acting as nucleophiles under mild conditions (compounds **8a-f** and **9a-f**). A set of tetrazine-containing amino acids (**10a-e** and **11a-e**) have been synthesized and used to unravel the kinetics of the subsequent iEDDA cycloaddition with a model strained cyclooctyne, revealing a substituent effect in the speed of the reaction, with tetrazine ethers being the most reactive.

Tetrazine ethers also exhibit a fluorescent behavior compared to tetrazines with other substituents. We have exploited this feature to incorporate the tyrosine derivative **10d** into a model peptide by solid-phase peptide synthesis and used it in

cell culture as a dual fluorescent and clickable peptide that loses its emissive properties upon iEDDA cycloaddition, showing for the first time their applicability *in vivo.*

3.6. Bibliography:

1 S. P. Brown and A. B. Smith. Peptide/protein stapling and unstapling: Introduction of s-tetrazine, photochemical release, and regeneration of the peptide/protein. *J. Am. Chem. Soc.* 2015, **137**, 4034–4037.

2 E. Jullien-Macchi, V. Alain-Rizzo, C. Allain, C. Dumas-Verdes and P. Audebert. S-Tetrazines functionalized with phenols: Synthesis and physico-chemical properties. *RSC Adv.* 2014, **4**, 34127–34133.

3 D. E. Chavez and M. A. Hiskey. 1,2,4,5-tetrazine based energetic materials. *J. Energ. Mater.* 1999, **17**, 357–377.

4 S. C. Fields, M. H. Parker and W. R. Erickson. A Simple Route to Unsymmetrically Substituted 1,2,4,5-Tetrazines. *J. Org. Chem.* 1994, **59**, 8284–8287.

5 A. Maggi, E. Ruivo, J. Fissers, C. Vangestel, S. Chatterjee, J. Joossens, F. Sobott, S. Staelens, S. Stroobants, P. Van Der Veken, L. Wyffels and K. Augustyns. Development of a novel antibody–tetrazine conjugate for bioorthogonal pretargeting. *Org. Biomol. Chem.* 2016, **14**, 7544–7551.

6 S. E. Pidgeon and M. M. Pires. Metabolic remodeling of bacterial surfaces via tetrazine ligations. *Chem. Commun.* 2015, **51**, 10330–10333.

7 J. L. Seitchik, J. C. Peeler, M. T. Taylor, M. L. Blackman, T. W. Rhoads, R. B. Cooley, C. Refakis, J. M. Fox and R. A. Mehl. Genetically Encoded Tetrazine Amino Acid Directs Rapid Site-Specific in Vivo Bioorthogonal Ligation with trans-Cyclooctenes. *J. Am. Chem. Soc.* 2012, **134**, 2898–2901.

8 Q. Zhou, P. Audebert, G. Clavier, F. Miomandre and J. Tang. New unsymmetrical alkyl-s-tetrazines: original syntheses, fluorescence and electrochemical behaviour. *RSC Adv.* 2014, **4**, 7193.

9 Z. Qing, P. Audebert, G. Clavier, F. Miomandre, J. Tang, T. T. Vu and R. Méallet-Renault. Tetrazines with hindered or electron withdrawing substituents: Synthesis, electrochemical and fluorescence properties. *J. Electroanal. Chem.* 2009, **632**, 39–44.

10 S. Mayer and K. Lang. Tetrazines in Inverse-Electron-Demand Diels–Alder Cycloadditions and Their Use in Biology. *Synthesis* 2016, **49**, 830–848.

11 G. B. Cserép, A. Herner and P. Kele. Bioorthogonal fluorescent labels: A review on combined forces. *Methods Appl. Fluoresc.* , DOI:10.1088/2050-6120/3/4/042001.

12 R. M. Versteegen, R. Rossin, W. Ten Hoeve, H. M. Janssen and M. S. Robillard. Click to release: Instantaneous doxorubicin elimination upon tetrazine ligation. *Angew. Chem., Int. Ed.* 2013, **52**, 14112–14116.

13 H. Lebraud, D. J. Wright, C. N. Johnson and T. D. Heightman. Protein degradation by in-cell self-assembly of proteolysis targeting chimeras. *ACS Cent. Sci.* 2016, **2**, 927–934.

14 B. L. Oliveira, Z. Guo and G. J. L. Bernardes. Inverse electron demand Diels–Alder reactions in chemical biology. *Chem. Soc. Rev.* 2017, **46**, 4895–4950.

15 R. A. A. Foster and M. C. Willis. Tandem inverse-electron-demand hetero-/retro-Diels-Alder reactions for aromatic nitrogen heterocycle synthesis. *Chem. Soc. Rev.* 2013, **42**, 63–76.

16 A. Hamasaki, R. Ducray and D. L. Boger. Two novel 1,2,4,5-tetrazines that participate in inverse electron demand Diels-Alder reactions with an unexpected regioselectivity. *J. Org. Chem.* 2006, **71**, 185–193.

17 D. L. Boger, R. P. Schaum and R. M. Garbaccio. Regioselective Inverse Electron Demand Diels-Alder Reactions of N-Acyl 6-Amino-3-(methylthio)-1,2,4,5-tetrazines. *J. Org. Chem.* 1998, **63**, 6329–6337.

18 F. Thalhammer, U. Wallfahrer and J. Sauer. Reaktivität einfacher offenkettiger und cyclischer dienophile bei Diels-Alder-reaktionen mit inversem elektronenbedarf. *Tetrahedron Lett.* 1990, **31**, 6851–6854.

19 F. Liu, R. S. Paton, S. Kim, Y. Liang and K. N. Houk. Diels–Alder Reactivities of Strained and Unstrained Cycloalkenes with Normal and Inverse-Electron-Demand Dienes: Activation Barriers and Distortion/Interaction Analysis. *J. Am. Chem. Soc.* 2013, **135**, 15642–15649.

20 F. Liu, Y. Liang and K. N. Houk. Theoretical Elucidation of the Origins of Substituent and Strain Effects on the Rates of Diels–Alder Reactions of 1,2,4,5-Tetrazines. *J. Am. Chem. Soc.* 2014, **136**, 11483–11493.

21 J. Yang, Y. Liang, J. Šečkutė, K. N. Houk and N. K. Devaraj. Synthesis and Reactivity Comparisons of 1-Methyl-3-Substituted Cyclopropene Mini-tags for Tetrazine Bioorthogonal Reactions. *Chem. Eur. J.* 2014, **20**, 3365–3375.

22 D. Wang, W. Chen, Y. Zheng, C. Dai, K. Wang, B. Ke and B. Wang. 3,6-Substituted-1,2,4,5-tetrazines: Tuning reaction rates for staged labeling applications. *Org. Biomol. Chem.* 2014, **12**, 3950–3955.

23 J. W. Wijnen, S. Zavarise, J. B. F. N. Engberts and M. Charton. Substituent effects on an inverse electron demand hetero Diels-Alder reaction in aqueous solution and organic solvents: Cycloaddition of substituted styrenes to di(2-pyridyl)-1,2,4,5-tetrazine. *J. Org. Chem.* 1996, **61**, 2001–2005.

24 B. L. Oliveira, Z. Guo and G. J. L. Bernardes. Inverse electron demand Diels–Alder reactions in chemical biology. *Chem. Soc. Rev.* 2017, **46**, 4895–4950.

25 N. K. Devaraj, S. Hilderbrand, R. Upadhyay, R. Mazitschek and R. Weissleder. Bioorthogonal turn-on probes for imaging small molecules inside living cells. *Angew. Chem., Int. Ed.* 2010, **49**, 2869–2872.

26 A. Wieczorek, T. Buckup and R. Wombacher. Rigid tetrazine fluorophore conjugates with fluorogenic properties in the inverse electron demand Diels-Alder reaction. *Org. Biomol. Chem.* 2014, **12**, 4177–4185.

27 L. G. Meimetis, J. C. T. Carlson, R. J. Giedt, R. H. Kohler and R. Weissleder. Ultrafluorogenic coumarin-tetrazine probes for real-time biological imaging. *Angew. Chem., Int. Ed.* 2014, **53**, 7531–7534.

28 H. Wu, J. Yang, J. Šečkute and N. K. Devaraj. In situ synthesis of alkenyl tetrazines for highly fluorogenic bioorthogonal live-cell imaging probes. *Angew. Chem., Int. Ed.* 2014, **53**, 5805–5809.

29 G. Knorr, E. Kozma, A. Herner, E. A. Lemke and P. Kele. New Red-Emitting Tetrazine-Phenoxazine Fluorogenic Labels for Live-Cell Intracellular Bioorthogonal Labeling Schemes. *Chem. Eur. J.* 2016, **22**, 8972–8979.

30 E. Kozma, G. Estrada Girona, G. Paci, E. A. Lemke and P. Kele. Bioorthogonal double-fluorogenic siliconrhodamine probes for intracellular super-resolution microscopy. *Chem. Commun.* 2017, **53**, 6696–6699.

31 G. Knorr, E. Kozma, J. M. Schaart, K. Németh, G. Török and P. Kele. Bioorthogonally Applicable Fluorogenic Cyanine-Tetrazines for No-Wash Super-Resolution Imaging. *Bioconjug. Chem.* 2018, **29**, 1312–1318.

32 A. Vázquez, R. Dzijak, M. Dračínský, R. Rampmaier, S. J. Siegl and M. Vrabel. Mechanism-Based Fluorogenic trans-Cyclooctene–Tetrazine Cycloaddition. *Angew. Chem., Int. Ed.* 2017, **56**, 1334–1337.

33 A. Wieczorek, P. Werther, J. Euchner and R. Wombacher. Green- to far-red-emitting fluorogenic tetrazine probes – synthetic access and no-wash protein imaging inside living cells. *Chem. Sci.* 2017, **8**, 1506–1510.

34 M. A. El-Sayed. Spin-orbit coupling and the radiationless processes in nitrogen heterocyclics. *J. Chem. Phys.* 1963, **38**, 2834–2838.

35 P. Audebert, F. Miomandre, G. Clavier, M. C. Verniéres, S. Badré and R. Méallet-Renault. Synthesis and properties of new tetrazines substituted by heteroatoms: Towards the world's smallest organic fluorophores. *Chem. Eur. J.* 2005, **11**, 5667–5673.

36 E. Jullien-Macchi, V. Alain-Rizzo, C. Allain, C. Dumas-Verdes and P. Audebert. S-Tetrazines functionalized with phenols: Synthesis and physico-chemical properties. *RSC Adv.* 2014, **4**, 34127–34133.

37 A. C. Knall and C. Slugovc. Inverse electron demand Diels-Alder (iEDDA)-initiated conjugation: A (high) potential click chemistry scheme. *Chem. Soc. Rev.* 2013, **42**, 5131–5142.

38 S. W. J. Lamberts, A.-J. van der Lely, W. W. de Herder and L. J. Hofland. Octreotide. *N. Engl. J. Med.* 1996, **334**, 246–254.

39 C. Bruns, F. Raulf, D. Hoyer, J. Schloos, H. Lübbert and G. Weckbecker. Binding properties of somatostatin receptor subtypes. *Metabolism.* 1996, **45**, 17–20.

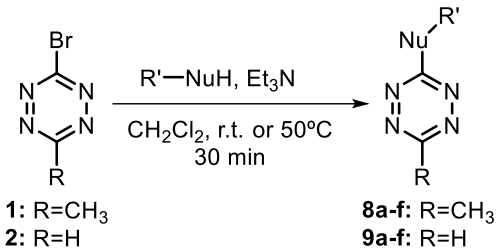
40 E. Kaiser, R. L. Colescott, C. D. Bossinger and P. I. Cook. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal. Biochem.* 1970, **34**, 595–598.

41 G. A. Crosby and J. N. Demas. Measurement of photoluminescence quantum yields. Review. *J. Phys. Chem.* 1971, **75**, 991–1024.

3.7. Experimental section

3.7.1. Organic synthesis

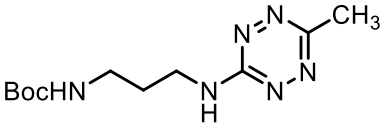
3.7.1.1. General procedure for the S_NAr reaction of **1** or **2** with nucleophiles:



The corresponding R'-NuH (0.21 mmol, 1.2 eq) was dissolved in 0.5 mL DCM, and **1** or **2** (0.17 mmol, 1.0 eq) was dissolved in 0.5 mL of DCM and added dropwise to the previous solution. Et_3N (0.17 mmol, 1.0 eq) was then added, and the reaction mixture was stirred for 30 min. In the cases were completion was not confirmed by TLC, the reaction mixture was heated at 50 ºC until completion.

The reaction mixtures were concentrated under reduced pressure, diluted in water and extracted with DCM (x3). The combined organic layers were dried over $MgSO_4$ and evaporated under vacuum, followed by flash column chromatography purification in silica (12g) in hexane with increasing concentrations of EtOAc.

N^1-Boc-N^3-(6-methyl-1,2,4,5-tetrazin-3-yl)-1,3-diaminopropane (8a)



Red solid (35.0 mg, 76%). **Mp =** 90.2 ºC

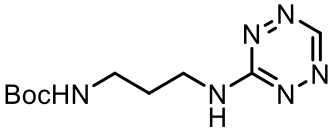
^{1}H NMR (400 MHz, $CDCl_3$) δ 6.41 (s, 1H), 4.82 (s, 1H), 3.64 (q, *J* = 6.4 Hz, 2H), 3.26 (dd, *J* = 12.5, 6.3 Hz, 2H), 2.82 (s, 3H), 1.82 (m, 2H), 1.45 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 161.9, 161.68, 156.7, 79.7, 38.1, 37.6, 29.9, 28.5, 20.2 ppm

HRMS (ESI^+) calculated for $C_{11}H_{21}O_2N_6$ $[M+H]^+$: 269.17205 / found 269.17196

IR (KBr) U_{max}: 3307, 2972, 2929, 2276, 1688, 1557, 1523, 1362, 1253, 1166, cm^{-1}

N^1-Boc-N^3-(1,2,4,5-tetrazin-3-yl)-1,3-diaminopropane (9a)



Red solid (32.7 mg, 69%). **Mp =** 94.5 ºC

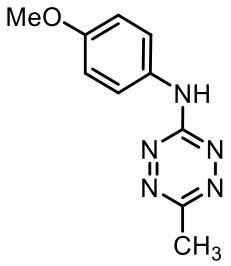
^{1}H NMR (400 MHz, $CDCl_3$) δ 9.60 (s, 1H), 6.69 (s, 1H), 4.78 (s, 1H), 3.67 (q, *J* = 6.4 Hz, 2H), 3.28 (q, *J* = 6.3 Hz, 2H), 1.83 (p, *J* = 6.3 Hz, 2H), 1.45 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 163.16, 156.91, 153.35, 79.87, 37.93, 37.44, 29.80, 28.52 ppm

HRMS (ESI$^+$) calculated for $C_{10}H_{18}N_6O_2$ $[M+H]^+$: 255.1564 / found 255.1563

IR (KBr) U_{max}: 3338, 3216, 3120, 3078, 3033, 2968, 2928. 2875, 1690, 1577, 1534, 1462, 1445, 1380, 1364, 1329, 1269, 1164, 1134, 1083, 1046 cm^{-1}

N-(4-methoxyphenyl)-6-methyl-1,2,4,5-tetrazin-3-amine (8b)



Red solid (26.0 mg, 70%). **Mp =** 114.6 ºC

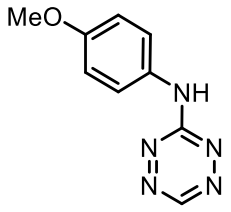
^{1}H NMR (400 MHz, DMSO-d_6) δ 7.65-7.53 (m, 2H), 7.01–6.92 (m, 2H), 3.75 (s, 3H), 2.75 (s, 3H) ppm

^{13}C NMR (101 MHz, DMSO-d_6) δ 161.2, 160.7, 155.4, 131.3, 121.5, 114.1, 55.2, 19.7 ppm

HRMS (ESI$^+$) calculated for $C_{10}H_{12}N_5O$ $[M+H]^+$: 218.1036 / found 218.1044

IR (KBr) U_{max}: 3262, 3199, 3097, 2929, 2826, 1608, 1561, 1506, 1417, 1240, 1028 cm^{-1}

N-(4-methoxyphenyl)-1,2,4,5-tetrazin-3-amine (9b)



Red-brown solid (18.9 mg, 50%). **Mp =** 124.7 ºC

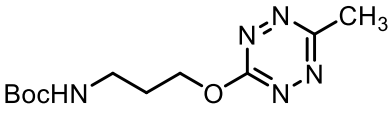
^{1}H NMR (400 MHz, DMSO-d_6) δ 10.62 (s, 1H), 9.87 (s, 1H), 7.62 – 7.56 (m, 2H), 7.00 – 6.92 (m, 2H), 3.74 (s, 3H) ppm

^{13}C NMR (101 MHz, DMSO-d_6) δ 162.39, 156.18, 153.86, 131.23, 122.58, 114.52, 55.70 ppm

HRMS (ESI$^+$) calculated for $C_9H_{10}N_5O$ $[M+H]^+$: 204.08799 / found 204.08803

IR (KBr) U_{max}: 3418 (broad band), 3261, 3201, 3144, 3096, 3020, 2930, 2828, 2202, 2165, 2051, 1609, 1561, 1508, 1419, 1296, 1267, 1242, 1174, 1068, 1028, 939 cm^{-1}

N-Boc-O-(6-methyl-1,2,4,5-tetrazin-3-yl)-3-aminopropanol (8c)



Pink oil (36.5 mg, 79%).

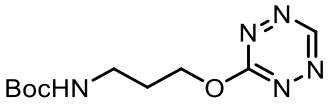
^{1}H NMR (400 MHz, $CDCl_3$) δ 4.68 (t, *J* = 6.2 Hz, 2H), 3.37 (q, *J* = 6.4 Hz, 2H), 2.99 (s, 3H), 2.12 (p, *J* = 6.4 Hz, 2H), 1.43 (d, *J* = 4.9 Hz, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 166.8, 165.5, 156.1, 67.4, 37.5, 32.0, 29.4, 28.5, 20.3 ppm

HRMS (ESI$^+$) calculated for $C_{11}H_{20}O_3N_5$ $[M+H]^+$: 270.15607 / found 270.15602

IR (KBr) U_{max}: 3360, 2960, 2920, 2851, 1697, 1519, 1480, 1427, 1366, 1170, 909 cm^{-1}

N-Boc-O-(1,2,4,5-tetrazin-3-yl)-3-aminopropanol (9c)



Pink oil (28.6 mg, 60%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 10.06 (s, 1H), 4.74 (t, *J* = 6.2 Hz, 2H), 3.38 (q, *J* = 6.5 Hz, 2H), 2.15 (p, *J* = 6.4 Hz, 2H), 1.43 (s, 9H) ppm

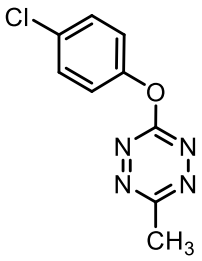
^{13}C NMR (101 MHz, $CDCl_3$) δ 167.86, 155.97, 77.19, 67.58, 37.28, 29.16, 28.35 ppm*

* C of the tetrazine ring (attached to O) could not be detected.

HRMS (ESI$^+$) calculated for $C_{10}H_{17}N_5O_3Na$ $[M+Na]^+$: 278.12236 / found 278.12232

IR (KBr) U_{max}: 3357, 2979, 2930, 2850, 1682, 1518, 1478, 1455, 1367, 1338, 1277, 1254, 1238, 1159, 1131, 1012, 969, 942 cm^{-1}

3-(4-chlorophenoxy)-6-methyl-1,2,4,5-tetrazine (8d)



Pink solid (34.4 mg, 92%). **Mp =** 90.2 ºC

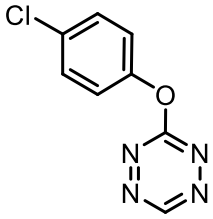
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.47-7.41 (m, 2H), 7.25–7.19 (m, 2H), 3.04 (s, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 167.6, 166.6, 150.4, 132.0, 130.2, 122.4, 20.3 ppm

HRMS (ESI$^+$) calculated for $C_9H_8ClN_4O$ $[M+H]^+$: 223.0381 / found 223.0383

IR (KBr) U_{max}: 3404, 3095, 2918, 1491, 1402, 1361, 1200, 909 cm^{-1}

3-(4-chlorophenoxy)-1,2,4,5-tetrazine (9d)



Pink solid (21.8 mg, 56%). **Mp =** 71.3 ºC

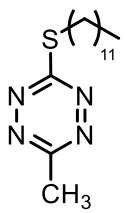
^{1}H NMR (400 MHz, $CDCl_3$) δ 10.18 (s, 1H), 7.52 – 7.40 (m, 2H), 7.30 – 7.17 (m, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 168.89, 157.03, 150.16, 132.53, 130.43, 122.58 ppm

HRMS (ESI$^+$) calculated for $C_8H_6ClN_4O$ $[M+H]^+$: 209.0225 / found 209.0228

IR (KBr) U_{max}: 3095, 3058, 2959, 2920, 2847, 2709, 2427, 2323, 2250, 2176, 2049, 1910, 1485, 1439, 1402, 1354, 1207, 1162, 1116, 1088, 1056, 1013 cm^{-1}

3-methyl-6-(dodecylthio)- 1,2,4,5-tetrazine (8e)



Starting from 40 mg of **1**. Pink oil (44.3 mg, 89%).

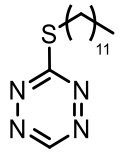
^{1}H NMR (400 MHz, $CDCl_3$) δ 3.35–3.25 (m, 2H), 2.96 (s, 3H), 1.86–1.73 (m, 2H), 1.55–1.40 (m, 2H), 1.26 (s, 18H), 0.87 (t, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 175.3, 165.0, 31.9, 30.4, 29.6, 29.6, 29.5, 29.4, 29.3, 29.1, 28.8, 28.7, 22.7, 20.6, 14.1 ppm

HRMS (ESI$^+$) calculated for $C_{15}H_{29}N_4S$ $[M+H]^+$: 297.2017 / found 297.2112

IR (KBr) U_{max}: 3407, 2924, 2853, 1466, 1384, 1317, 1162, 1067, 880 cm^{-1}

3-(dodecylthio)-1,2,4,5-tetrazine (9e)



Pink oil (35.1 mg, 67%).

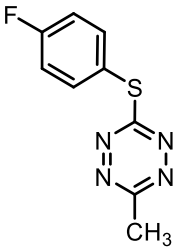
^{1}H NMR (400 MHz, $CDCl_3$) δ 9.95 (s, 1H), 3.36 – 3.27 (m, 2H), 1.80 (dt, J = 15.0, 7.4 Hz, 2H), 1.48 (dt, J = 15.1, 7.1 Hz, 2H), 1.38 – 1.19 (m, 16H), 0.87 (t, J = 6.9 Hz, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 178.94, 155.77, 32.05, 30.66, 29.76, 29.75, 29.69, 29.59, 29.47, 29.22, 28.94, 28.72, 22.82, 14.25 ppm

HRMS (ESI$^+$) calculated for for $C_{14}H_{27}N_4S$ $[M+H]^+$: 283.1951 / found 283.1948

IR (KBr) U_{max}: 2924, 2852, 2341, 1458, 1398, 1217, 1033 cm^{-1}

3-[(4-fluorophenyl)thio]-6-methyl-1,2,4,5-tetrazine (8f)



Pink solid (33.5 mg, 88%). **Mp =** 82.1 ºC

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.69–7.60 (m, 2H), 7.22–7.13 (m, 2H), 2.95 (s, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 175.6, 165.7, 165.4, 162.9, 137.9, 120.9, 120.9, 117.3, 117.0, 20.6 ppm

HRMS (ESI$^+$) calculated for $C_9H_8FN_4S$ $[M+H]^+$: 223.0448 / found 223.0452

IR (KBr) U_{max}: 3093, 2922, 1587, 1490, 1380, 1311, 1219, 1157, 872, 838 cm^{-1}

3-[(4-fluorophenyl)thio]-1,2,4,5-tetrazine (9f)

Pink oil (19.7 mg, 51%).

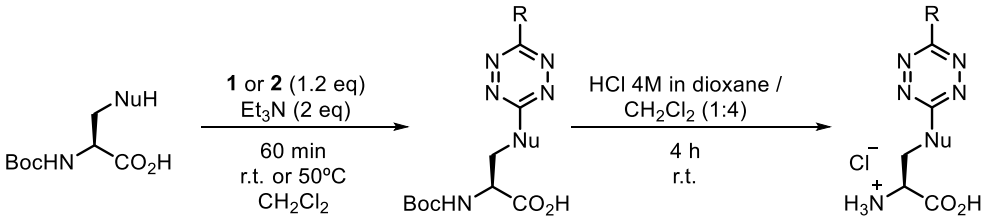
^{1}H NMR (400 MHz, $CDCl_3$) δ 9.99 (s, 1H), 7.69 – 7.63 (m, 2H), 7.23 – 7.17 (m, 2H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.66, 163.15, 156.32, 138.08 (d, C-F coupling, J = 8.8 Hz), 120.45, 117.49 (d, C-F coupling, J = 22.2 Hz) ppm

HRMS (ESI$^+$) calculated for $C_8H_5FN_4S$ $[M+H]^+$: 209.0292 / found 209.0296

IR (KBr) U_{max}: 3092, 2953, 2923, 2851, 2360, 2341, 1589, 1490, 1396, 1216, 1157, 1091 cm^{-1}

3.7.1.2. General procedure for preparing amino acid derivatives with **1** and **2**:



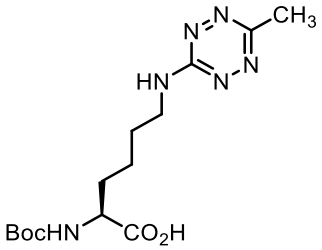
The corresponding N^{α}-Boc-protected amino acid bearing a nucleophilic side chain (0.12 mmol, 1 eq) was dissolved in 0.5 mL anh. DCM in a sealed vial with N_2 atmosphere, and **1** or **2** (0.14 mmol, 1.2 eq) dissolved in 0.5 mL of DCM were added dropwise. Then, Et_3N (0.24 mmol, 2 eq) was added to the previous solution. The reaction mixture was stirred for 60 min at room temperature, when completion was confirmed by TLC. In cases when only starting material was observed, the reaction mixture was heated at 50 ºC.

Each reaction mixture was concentrated under reduced pressure, diluted in HCl 0.1 M aqueous solution, and extracted with EtOAc (x3). The combined organic layers were dried over $MgSO_4$ and evaporated under vacuum. Purification by flash column

chromatography was performed in hexane with increasing concentrations of EtOAc + 1% acetic acid.

For the preparation of the unnatural amino acids, the relevant quantities of the Boc-protected amino acids (detailed in each case) were deprotected in 4:1 DCM / HCl 4 M in dioxane at a final concentration of 0.2 M. After 4 hours at room temperature under N_2 atmosphere, the resulting red/pink solid was filtered over a glass filter crucible to obtain the desired product.

N^{α}-Boc-N^{ε}-(6-methyl-1,2,4,5-tetrazin-3-yl)-*L*-Lysine (N-Boc-10a)



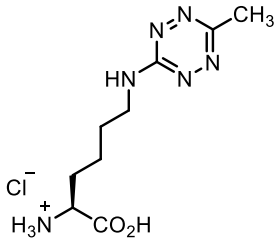
Red gum (43.2 mg, 89%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 6.75 (d, *J* = 33.1 Hz, 1H), 5.38 (d, *J*= 8.1 Hz, 1H), 4.36 (d, *J* = 6.3 Hz, 1H), 3.67–3.53 (m, 2H), 2.81 (s, 3H), 1.87 (dd, *J* = 14.1, 10.6 Hz, 2H), 1.83–1.68 (m, 2H), 1.56 (d, *J* = 6.5 Hz, 2H), 1.42 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 176.6, 161.5, 155.9, 80.3, 53.4, 41.0, 40.6, 32.2, 31.7, 28.5, 22.6, 20.1 ppm

HRMS (ESI^-) calculated for $C_{14}H_{23}N_6O_4$ $[M-H]^-$: 339.17863 / found 339.17862

N^{ε}-(6-methyl-1,2,4,5-tetrazin-3-yl)-*L*-Lysine hydrochloride (10a)



Red solid (26.9 mg, 91% starting from 36.3 mg of **N-Boc-10a**).

^{1}H NMR (400 MHz, CD_3OD) δ 3.99 (t, *J* = 6.3 Hz, 1H), 3.53 (t, *J* = 7.0 Hz, 2H), 2.74 (s, 3H), 2.06–1.90 (m, 2H), 1.81–1.72 (m, 2H), 1.57 (tdd, *J* = 13.8, 10.8, 6.6 Hz, 2H) ppm

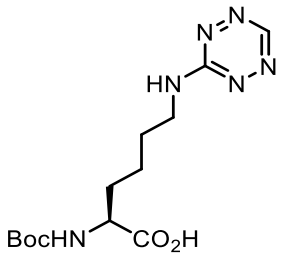
^{13}C NMR (101 MHz, CD_3OD) δ 171.7, 163.1, 161.9, 53.8, 41.4, 31.2, 29.4, 23.3, 19.8 ppm

HRMS (ESI⁺) calculated for $C_9H_{16}N_6O_2$ $[M+H]^+$: 241.14075 / found 241.14097

$[\alpha]_D$: +13.9 (*c* 0.51, CH_3OH)

IR U_{max}: 2875, 2540, 1980, 1735, 1658, 1421, 1213, 1190, 1022, 781, 650 cm^{-1}

N^{α}-Boc-N^{ε}-(1,2,4,5-tetrazin-3-yl)-*L*-Lysine (N-Boc-11a)



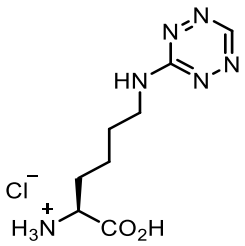
Red gum (34.7 mg, 76%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 9.61 (s, 1H), 7.02 – 6.77 (m, 1H), 5.34 (d, J = 7.8 Hz, 1H), 4.42 – 4.14 (d, J = 5.5 Hz, 1H), 3.62 (d, J = 5.6 Hz, 2H), 1.98 – 1.69 (m, 4H), 1.57 (m, 2H), 1.43 (s, 9H) ppm.

^{13}C NMR (101 MHz, $CDCl_3$) δ 176.47, 162.84, 155.86, 153.12, 80.44, 53.32, 46.19, 40.92, 32.19, 29.83, 28.43, 22.55 ppm.

HRMS (ESI⁺) calculated for $C_{13}H_{22}N_6O_4$ $[M+H]^+$: 327.17753 / found 327.17806

N^{ε}-(1,2,4,5-tetrazin-3-yl)-*L*-Lysine hydrochloride (11a)



Red solid (24.4 mg, 92% starting from 33.0 mg of **N-Boc-11a**).

^{1}H NMR (400 MHz, CD_3OD) δ 9.57 (s, 1H), 3.91 (t, J = 6.3 Hz, 1H), 3.66 (s, 1H), 3.55 (t, J = 7.0 Hz, 2H), 2.01 – 1.92 (m, 2H), 1.81 – 1.73 (m, 2H), 1.62 – 1.54 (m, 2H) ppm

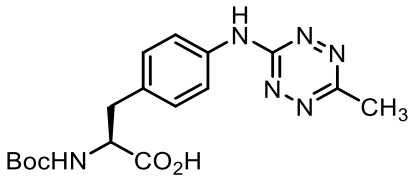
^{13}C NMR (101 MHz, CD_3OD) δ 171.37, 164.60 153.75, 54.26, 41.27, 31.37, 29.40, 23.37 ppm

HRMS (ESI⁺) calculated for $C_8H_{14}N_6O_2$ $[M+H]^+$: 227.12510 / found 227.12509

$[\alpha]_D$: +7.8 (*c* 0.11, CH_3OH)

IR (KBr) U_{max}: 3210, 3063, 3014, 2920, 2850, 2628, 2547, 1737, 1581, 1496, 1458, 1255, 1080 cm^{-1}

N^{α}-Boc-N^4-(6-methyl-1,2,4,5-tetrazin-3-yl)-4-amino-*L*-Phenylalanine[7] (6)

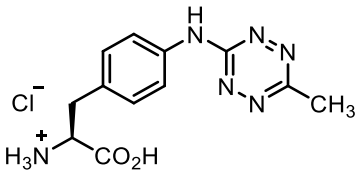


Orange gum (45.2 mg, 85%).

^{1}H NMR (400 MHz, CD_3OD) δ 7.63 (d, *J* = 8.4 Hz, 2H), 7.24 (d, *J* = 8.5 Hz, 2H), 4.39–4.23 (m, 1H), 3.21–2.86 (m, 2H), 2.81 (s, 3H), 1.39 (s, 9H) ppm

^{13}C NMR (101 MHz, CD_3OD) δ 175.4, 163.1, 162.5, 157.8, 138.3, 133.8, 130.8, 120.9, 80.5, 56.4, 38.2, 28.7, 20.0 ppm

N^4-(6-methyl-1,2,4,5-tetrazin-3-yl)-4-amino-*L*-Phenylalanine hydrochloride (10b, Tet-v1.0[7])



Orange solid (29.8 mg, 94% starting from 38.2 mg of **6**).

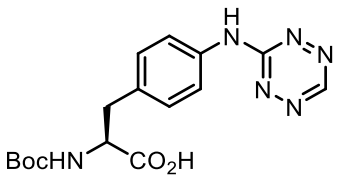
^{1}H NMR (400 MHz, CD_3OD) δ 7.75–7.70 (d, 2H), 7.35–7.30 (d, 2H), 4.27 (dd, *J* = 7.7, 5.4 Hz, 1H), 3.32 (dd, *J* = 14.6, 7.7 Hz, 1H), 3.17 (dd, *J* = 14.6, 7.7 Hz, 1H), 2.82 (s, 3H) ppm

^{13}C NMR (101 MHz, CD_3OD) δ 171.2, 163.3, 162.6, 139.5, 131.1, 130.5, 121.6, 55.2, 36.7, 20.1 ppm

$[\alpha]_D$: +20.6 (*c* 4.15, CH_3OH)

IR υ_{max}: 3259, 3183, 2920, 2854, 2630, 1688, 1558, 1492, 1419, 1368, 1065 cm^{-1}

N^{α}-Boc-N^4-(1,2,4,5-tetrazin-3-yl)-4-amino-*L*-Phenylalanine (N-Boc-11b)



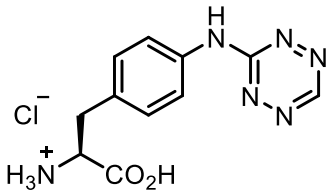
Orange gum (29.7 mg, 58%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 9.72 (s, 1H), 7.64 (d, J = 8.3 Hz, 2H), 7.42 (d, J = 7.9 Hz, 1H), 7.25 (d, J = 8.3 Hz, 2H), 7.17 (d, J = 8.4 Hz, 1H), 4.36 (dd, J = 8.8, 4.9 Hz, 1H), 3.25 – 2.80 (m, 2H), 1.38 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 175.31, 163.84, 157.76, 154.62, 137.83, 134.31, 130.86, 121.35, 80.53, 56.28, 38.13, 28.68 ppm

HRMS (ESI$^+$) calculated for $C_{16}H_{20}N_6O_2$ $[M+H]^+$: 361.16188 / found 361.16189

N^4-(1,2,4,5-tetrazin-3-yl)-4-amino-*L*-Phenylalanine hydrochloride (11b)



Orange solid (19.5 mg, 73% starting from 32.4 mg of **N-Boc-11b**).

^{1}H NMR (400 MHz, CD_3OD) δ 9.77 (s, 1H), 7.77 – 7.72 (d, 2H), 7.36 – 7.32 (d, 2H), 4.27 (dd, J = 7.8, 5.4 Hz, 1H), 3.34 (dd, J = 14.6, 7.8 Hz, 1H), 3.17 (dd, J = 14.6, 7.8 Hz, 1H) ppm

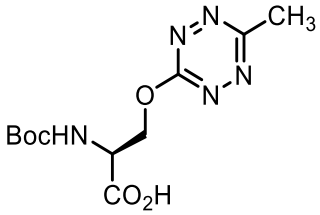
^{13}C NMR (101 MHz, CD_3OD) δ 171.20, 164.03, 154.85, 139.02, 131.14, 131.08, 122.11, 55.15, 36.78 ppm

HRMS (ESI$^-$) calculated for $C_{11}H_{12}N_6O_2$ $[M-H]^-$: 261.10945 / found 261.10932

$[\alpha]_D$: +16.0 (*c* 0.41, CH_3OH)

IR (KBr) U_{max}: 3266, 2963, 2925, 2852, 1688, 1605, 1492, 1366, 1259, 1155 cm^{-1}

N^α-Boc-O-(6-methyl-1,2,4,5-tetrazin-3-yl)-*L*-Serine (N-Boc-10c)



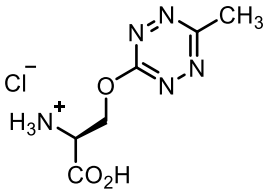
Pink gum (25.6 mg, 60%).

^{1}H NMR (400 MHz, CD_3OD) δ 5.00 (dd, *J* = 10.9, 3.8 Hz, 1H), 4.88 (m, 1H), 4.86 (m, 1H), 2.94 (s, 3H) ppm

^{13}C NMR (101 MHz, CD_3OD) δ 171.3, 166.7, 165.5, 79.5, 68.3, 52.8, 29.3, 27.2, 18.7 ppm

HRMS (ESI⁻) calculated for $C_{11}H_{16}N_5O_5$ $[M-H]^-$: 298.11569 / found 298.11555

O-(6-methyl-1,2,4,5-tetrazin-3-yl)-*L*-Serine hydrochloride (10c)



Pink solid (5.2 mg, 34% starting from 19.4 mg of **N-Boc-10c**).

^{1}H NMR (400 MHz, D_2O) δ 4.76 (d, J = 11.6 Hz, 1H), 4.11 (s, 2H), 2.81 (s, 3H) ppm

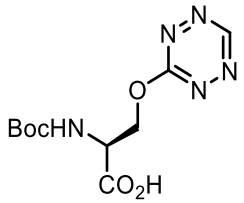
^{13}C NMR (101 MHz, D_2O) δ 174.5, 161.4, 161.1, 61.2, 56.6, 18.8 ppm

$[\alpha]_D$: 6.41 (*c* 0.26, CH_3OH)

HRMS (ESI⁺) calculated for $C_6H_{10}N_5O_3$ $[M+H]^+$: 200.07782 / found 200.07784

IR U_{max}: 2943, 2833, 1655, 1449, 1415, 1015, 567 cm^{-1}

N^{α}-Boc-O-(1,2,4,5-tetrazin-3-yl)-*L*-Serine (N-Boc-11c)



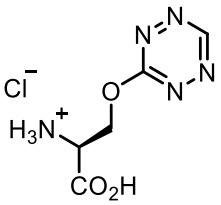
Pink gum (19.0 mg, 47%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 10.10 (s, 1H), 5.64 (d, *J* = 8.3 Hz, 1H), 5.04 (dd, *J* = 33.9, 9.6 Hz, 2H), 4.89 (s, 1H), 1.45 (s, 9H) ppm ("grease" and diethyl ether signals observed)

^{13}C NMR (101 MHz, $CDCl_3$) δ 172.15, 167.80, 156.52, 155.66, 81.11, 69.53, 52.84, 28.41 ppm ("grease" and diethyl ether signals observed)

HRMS (ESI⁺) calculated for $C_{10}H_{16}N_5O_5$ $[M+H]^+$: 286.11460 / found 286.11466

O-(1,2,4,5-tetrazin-3-yl)-*L*-Serine hydrochloride (11c)



Pink solid (12.8 mg, 54% starting from 30.8 mg of **N-Boc-11c**).

^{1}H NMR (400 MHz, CD_3OD) δ 10.26 (s, 1H), 5.16 (ddd, *J* = 16.9, 12.0, 4.0 Hz, 2H), 4.69 (dd, *J* = 4.9, 3.2 Hz, 1H) ppm

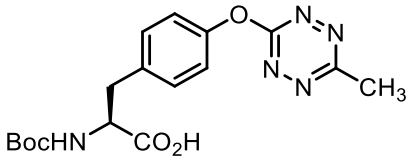
^{13}C NMR (101 MHz, CD_3OD) δ 169.02, 168.64, 158.10, 67.66, 53.37 ppm

$[\alpha]_D$: 6.87 (*c* 1.28, CH_3OH)

HRMS (ESI$^+$) calculated for $C_5H_{10}N_5O_3$ $[M+H]^+$: 188.07782 / found 188.07801

IR U_{max}: 2846 (broad band), 1726, 1561, 1475, 1452, 1342, 1218, 1138, 1056, 1012, 981 cm^{-1}

N$^\alpha$-Boc-O-(6-methyl-1,2,4,5-tetrazin-3-yl)-*L*-Tyrosine (N-Boc-10d)



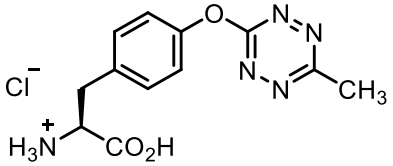
Pink gum (37.8 mg, 84%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.30 (d, *J* = 8.5 Hz, 2H), 7.23–7.19 (d, 2H), 5.02 (d, *J* = 8.0 Hz, 1H), 4.64 (m, 1H), 3.29–3.20 (m, 1H), 3.17–3.05 (m, 1H), 3.03 (s, 3H), 1.43 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 175.6, 167.8, 166.4, 155.5, 151.2, 134.6, 131.2, 121.1, 80.6, 54.3, 37.4, 28.4, 20.4 ppm

HRMS (ESI$^+$) calculated for $C_{17}H_{21}N_5O_5$ $[M+Na]^+$: 398.14349 / found 398.14387

O-(6-methyl-1,2,4,5-tetrazin-3-yl)-*L*-Tyrosine hydrochloride (10d)



Pink solid (27.0 mg, 86% starting from 37.8 mg of **N-Boc-10d**).

^{1}H NMR (400 MHz, CD_3OD) δ 7.48–7.43 (d, 2H), 7.37–7.32 (d, 2H), 4.31 (dd, *J* = 8.1, 5.3 Hz, 1H), 3.41 (dd, *J* = 14.5, 5.3 Hz, 1H), 3.21 (dd, *J* = 14.6, 8.1 Hz, 1H), 2.97 (s, 3H) ppm

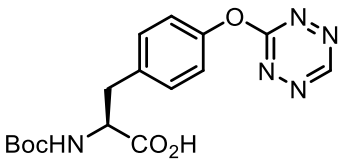
^{13}C NMR (101 MHz, CD_3OD) δ 171.1, 169.3, 167.7, 153.4, 134.1, 132.3, 122.9, 55.2, 36.7, 20.12 ppm

$[\alpha]_D$: -12.6 (*c* 3.90, CH_3OH)

HRMS (ESI$^+$) calculated for $C_{12}H_{13}N_5O_3$ $[M+H]^+$: 276.1091 / found 276.1091

IR U_{max}: 3200, 2912, 2863, 2579, 2447, 1760, 1577, 1502, 1402, 1366, 1251 cm^{-1}

N^{α}-Boc-O-(1,2,4,5-tetrazin-3-yl)-*L*-Tyrosine (N-Boc-11d)



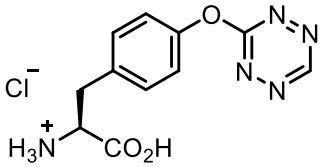
Pink gum (17.4mg, 34%).

^{1}H NMR (400 MHz, CD_3OD) δ 10.22 (s, *J* = 1.3 Hz, 1H), 7.38 (d, *J* = 8.5 Hz, 2H), 7.26 (d, *J* = 8.5 Hz, 2H), 4.39 (dd, *J* = 8.8, 4.9 Hz, 1H), 3.28 – 3.18 (m, 1H), 2.98 (dd, *J* = 14.0, 9.1 Hz, 1H), 1.40 (s, 9H) ppm

^{13}C NMR (101 MHz, CD_3OD) δ 168.99, 156.59, 156.38, 150.85, 135.94, 130.61, 120.53, 79.11, 54.94, 36.75, 27.27 ppm

HRMS (ESI^-) calculated for $C_{16}H_{19}N_5O_5$ $[M-H]^-$: 360.1313 / found 360.1304

O-(1,2,4,5-tetrazin-3-yl)-*L*-Tyrosine hydrochloride (11d)



Pink solid (8.5 mg, 68% starting from 15.2 mg of **N-Boc-11d**).

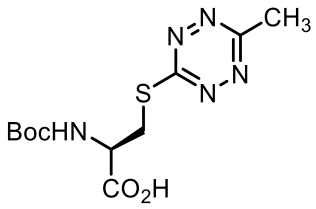
^{1}H NMR (400 MHz, CD_3OD) δ 10.26 (s, 1H), 7.47 (d, *J* = 8.6 Hz, 2H), 7.37 (d, *J* = 8.6 Hz, 2H), 4.32 (dd, *J* = 8.0, 5.3 Hz, 1H), 3.42 (dd, *J* = 14.5, 5.3 Hz, 1H), 3.23 (dd, *J* = 14.6, 8.1 Hz, 1H) ppm

^{13}C NMR (101 MHz, CD_3OD) δ 171.11, 170.40, 158.12, 153.08, 134.34, 132.40, 122.94, 55.12, 36.68 ppm

HRMS (ESI^+) calculated for $C_{11}H_{12}N_5O_3$ $[M+H]^+$: 262.0935 / found 262.0936

IR U_{max}: 3364, 2924 (broad band), 1729, 1595, 1507, 1439, 1358, 1198, 1055, 1018, 937, 896, 820 cm^{-1}

N^{α}-Boc-S-(6-methyl-1,2,4,5-tetrazin-3-yl)-*L*-Cysteine (N-Boc-10e)



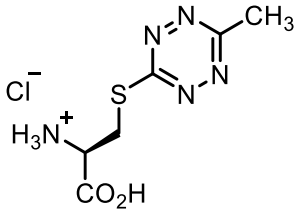
Pink gum (38.4 mg, 73%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 6.81 (s, 1H), 5.46 (s, 1H), 4.72 (m, 2H), 4.01 (dd, J = 14.1, 4.7 Hz, 2H), 3.67 (dd, J = 13.9, 6.9 Hz, 1H), 2.98 (s, 3H), 1.41 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 174.3, 173.6, 165.6, 155.5, 80.9, 52.7, 32.3, 28.2, 20.6 ppm

HRMS (ESI^-) calculated for $C_{11}H_{16}N_5O_4S$ $[M-H]^-$: 314.0928 / found 314.0924

S-(6-methyl-1,2,4,5-tetrazin-3-yl)-*L*-Cysteine hydrochloride (10e)



Pink solid (24.3 mg, 79% starting from 38.4 mg of **N-Boc-10e**).

^{1}H NMR (400 MHz, CD_3OD) δ 4.55 (dd, J = 7.3, 4.4 Hz, 1H), 4.16 (dd, J = 15.2, 4.4 Hz, 1H), 3.78 (dd, J = 15-.2, 7.3 Hz, 1H), 2.97 (s, 3H) ppm

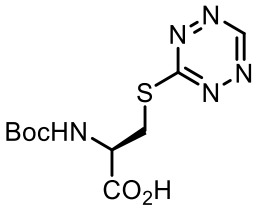
^{13}C NMR (101 MHz, CD_3OD) δ 174.6, 169.7, 167.6, 53.6, 30.7, 20.7 ppm

$[\alpha]_D$: -41.2 (*c* 0.51, CH_3OH)

HRMS (ESI^+) calculated for $C_6H_{10}N_5O_2S$ $[M]^+$: 216.055 / found 216.055

IR U_{max}: 2893, 2629, 1733, 1601, 1573, 1494, 1378, 1313, 1223, 1168, 831, 802 cm^{-1}

N^{α}-Boc-S-(1,2,4,5-tetrazin-3-yl)-*L*-Cysteine (N-Boc-11e)

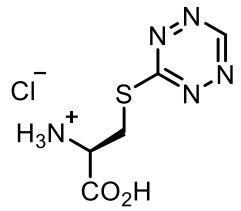


Pink gum (20.2 mg, 47%).

^{1}H NMR (400 MHz, CD_3OD) δ 10.11 (s, 1H), 4.54 (dd, *J* = 8.2, 4.3 Hz, 1H), 4.10 (dd, *J* = 13.9, 4.3 Hz, 1H), 3.52 (dd, *J* = 13.9, 8.5 Hz, 1H), 1.41 (s, 9H) ppm

^{13}C NMR (101 MHz, CD_3OD) δ 179.27, 157.82, 157.34, 80.70, 54.44, 33.68, 28.68 ppm

S-(1,2,4,5-tetrazin-3-yl)-*L*-Cysteine hydrochloride (11e)



Pink solid (34.8 mg, 62% starting from 70.8 mg of **N-Boc-11e**).

^{1}H NMR (400 MHz, CD_3OD) δ 10.24 (s, 1H), 4.61 – 4.53 (m, 1H), 4.20 (dd, *J* = 15.2, 4.4 Hz, 1H), 3.81 (dt, *J* = 12.6, 6.3 Hz, 1H) ppm

^{13}C NMR (101 MHz, CD_3OD) δ 177.94, 169.64, 157.93, 53.51, 30.71 ppm

$[\alpha]_D$: -75.4 (*c* 3.48, CH_3OH)

HRMS (ESI$^+$) calculated for $C_5H_{10}O_2N_5S$ $[M+H]^+$ (*in situ* reduction of the tetrazine observed): 204.05497 / found 204.05512

IR U_{max}: 3205, 2898, 2830, 2713, 2649, 2574, 2081, 1936, 1728, 1631, 1564, 1496, 1403, 1218, 1196, 1094, 1011, 953 cm^{-1}

3.7.1.3. Preparation of **OCT-10d**:

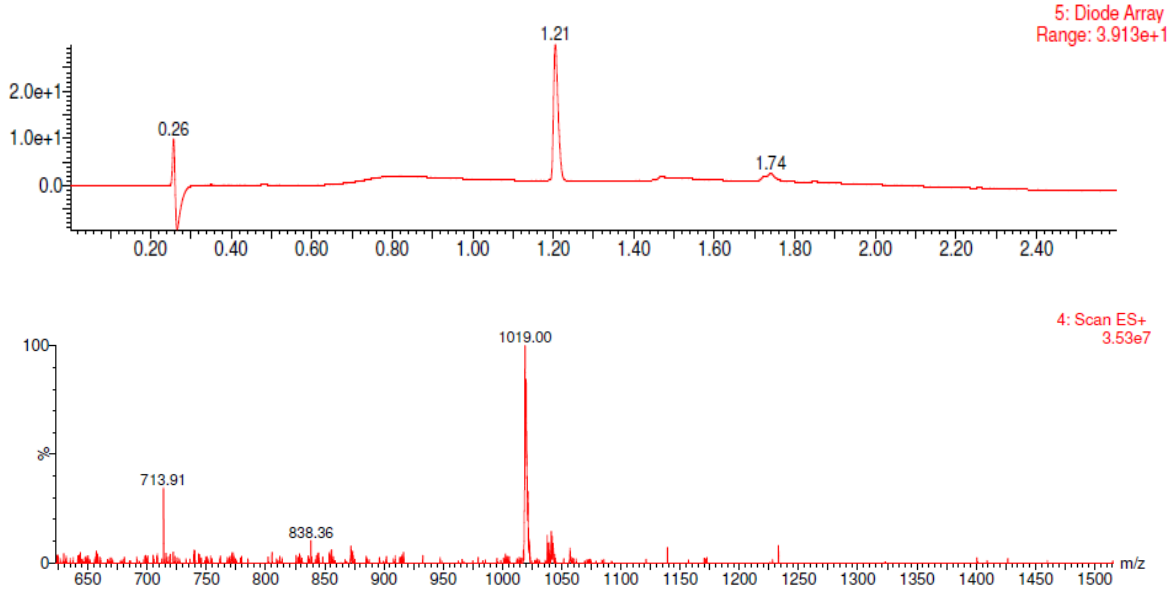
General considerations for solid-phase peptide synthesis (SPPS):

Peptides were obtained by SPPS on a 2-chlorotrityl resin (2-Cl-Trt, 1.60 mmol·g-1) following the fluorenylmethoxycarbonyl (Fmoc) / *tert*-butyl (tBu) strategy. Incorporation of the first Fmoc-AA-OH (1 eq) into 2-Cl-Trt resin was achieved by nucleophilic attack in anhydrous DCM with DIEA (2.5 eq) for 1h. Remaining reactive positions were end-capped by adding MeOH (0.8 mL·g-1) for 15 min. Fmoc removal was performed by treating the peptidyl resin with Piperidine: DMF (20:80, 1 x 15 min). The next Fmoc-protected amino acid (2.5 eq) were coupled to the peptidyl resin using DIPCDI (2.5 eq) and HOBt (2.5 eq) as activating agents in DMF for 60 min. Kaiser (ninhydrin) test was used to check coupling completions[40]. Semi-preparative RP-HPLC was performed on an Agilent 1260 Infinity with a Pursuit 10 C18 (150 mm x 21.2 mm, 10 μm) column using linear gradients of H_2O (0.1% TFA) and ACN 90% (0.1% TFA). The UV detection was performed at 220 nm and 254 nm to obtain pure peptides.

H-D-Phe-Cys(Trt)-Phe-D-Trp-Lys(Boc)-Thr(tBu)-Cys(Trt)-Throl-2-Cl-Trt resin (OCT-resin) was synthesized following the general procedure starting from 5.0 g of

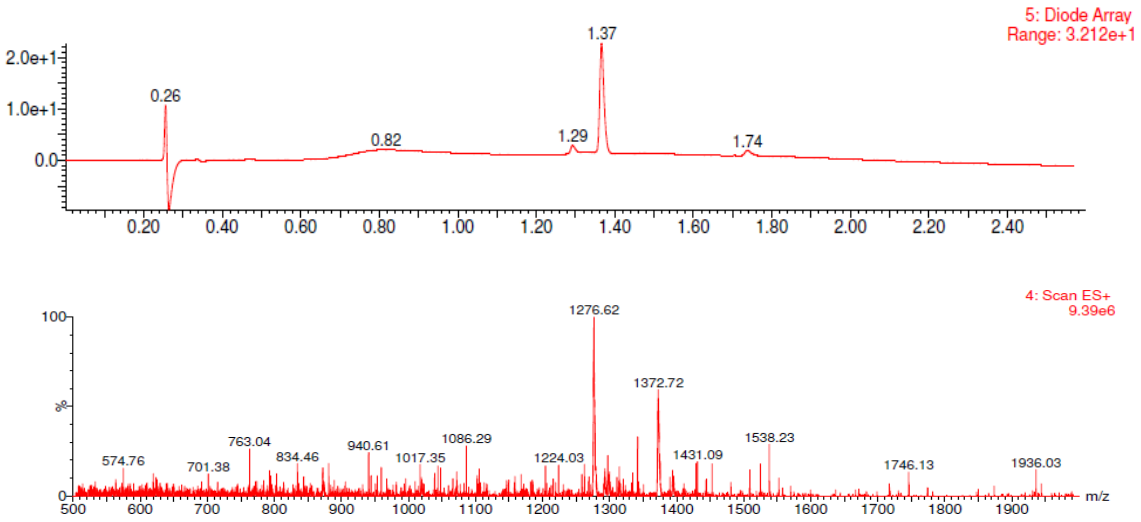
2-Cl-Trt resin (1.60 mmol·g^{-1}) using Fmoc-L-Threoninol (1 eq) as the first amino acid. After the Fmoc-removal of the last amino acid (Fmoc-D-Phe-OH) the resin was dried. Peptidyl resin (11.6 g, f = 0.338 mmol·g^{-1}) was obtained and **OCT** was cleaved from the resin using a deprotection cocktail (TFA: dodecanthiol: tioanisole: triisopropylsilane: H2O, 85:6:3:3:3 (v/v/v/v/v)) for 2h. The formation of the disulfide bridge was achieved by air oxidation dissolving the crude peptide in a mixture of H_2O:DMSO:AcOH (80:10:10 (v/v/v)) and stirring overnight. DMSO removal was achieved through a polyaromatic adsorbent resin column chromatography.

MS (ESI$^+$) calculated for $C_{49}H_{69}N_{10}O_{10}S_2$ $[M]^+$: 1021.3 / found 1021.3



H-10d-D-Phe-Cys(Trt)-Phe-D-Trp-Lys(Boc)-Thr(tBu)-Cys(Trt)-Throl-2-Cl-Trt resin (OCT-10d-resin) was synthesized from peptidyl resin (**OCT-resin**) (1 eq, 0.14 mmol, 0.42 g) and **N-Boc-10d** (1.5 eq, 0.21 mmol, 0.08 g) using PyBOP/HOBt/DIEA (1.5eq/1.5eq/3eq) as coupling reagents in DMF for 24h. Reactivation with PyBOP/HOBt/DIEA was carried out for 20h to achieve reaction completion. Peptidyl resin (**OCT-10d-resin**) (0.46 g) was obtained and subsequently cleaved and deprotected with TFA cocktail to obtain the reduced version of the peptide, which was air-oxidized in H_2O:DMSO:AcOH for 15h. Then, DMSO was removed to finally obtain **OCT-10d**.

MS (ESI$^+$) calculated for $C_{61}H_{78}O_{12}N_{15}S_2$ $[M+H]^+$: 1276.53903 / found 1276.53837



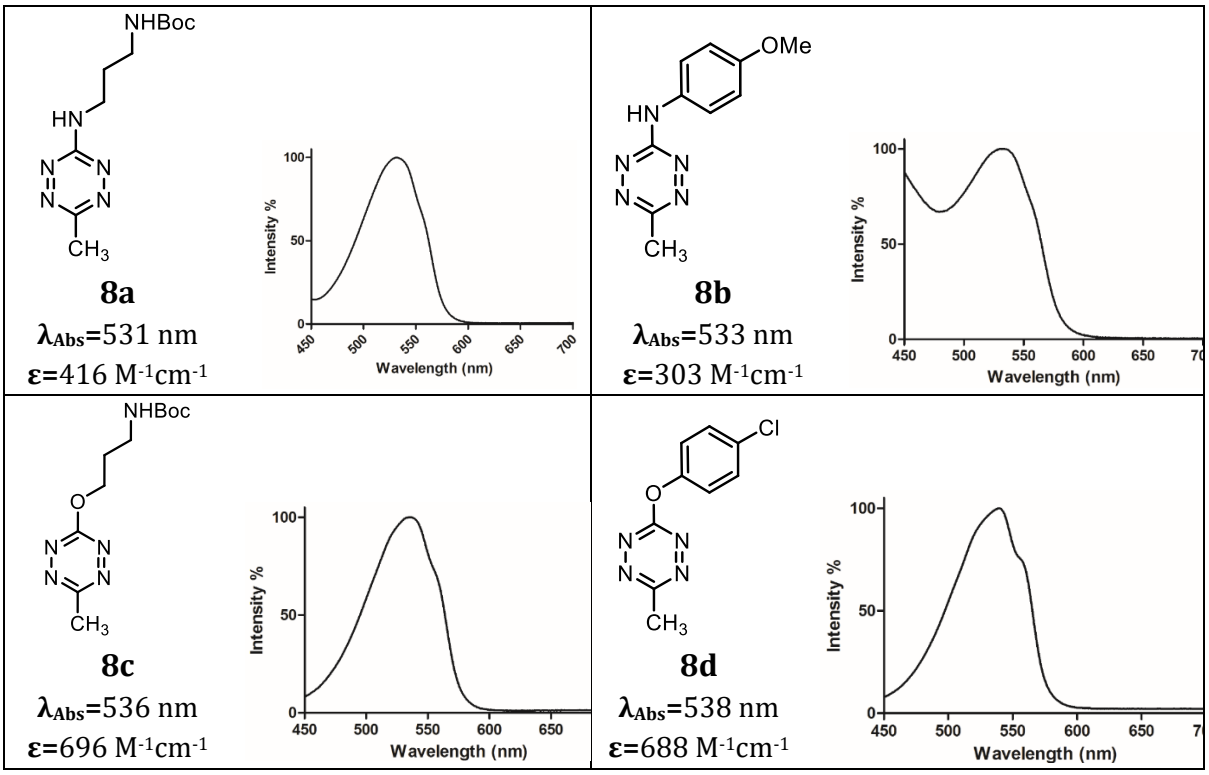
The reduced version of the tetrazine (dihydrotetrazine) was also observed and purified:

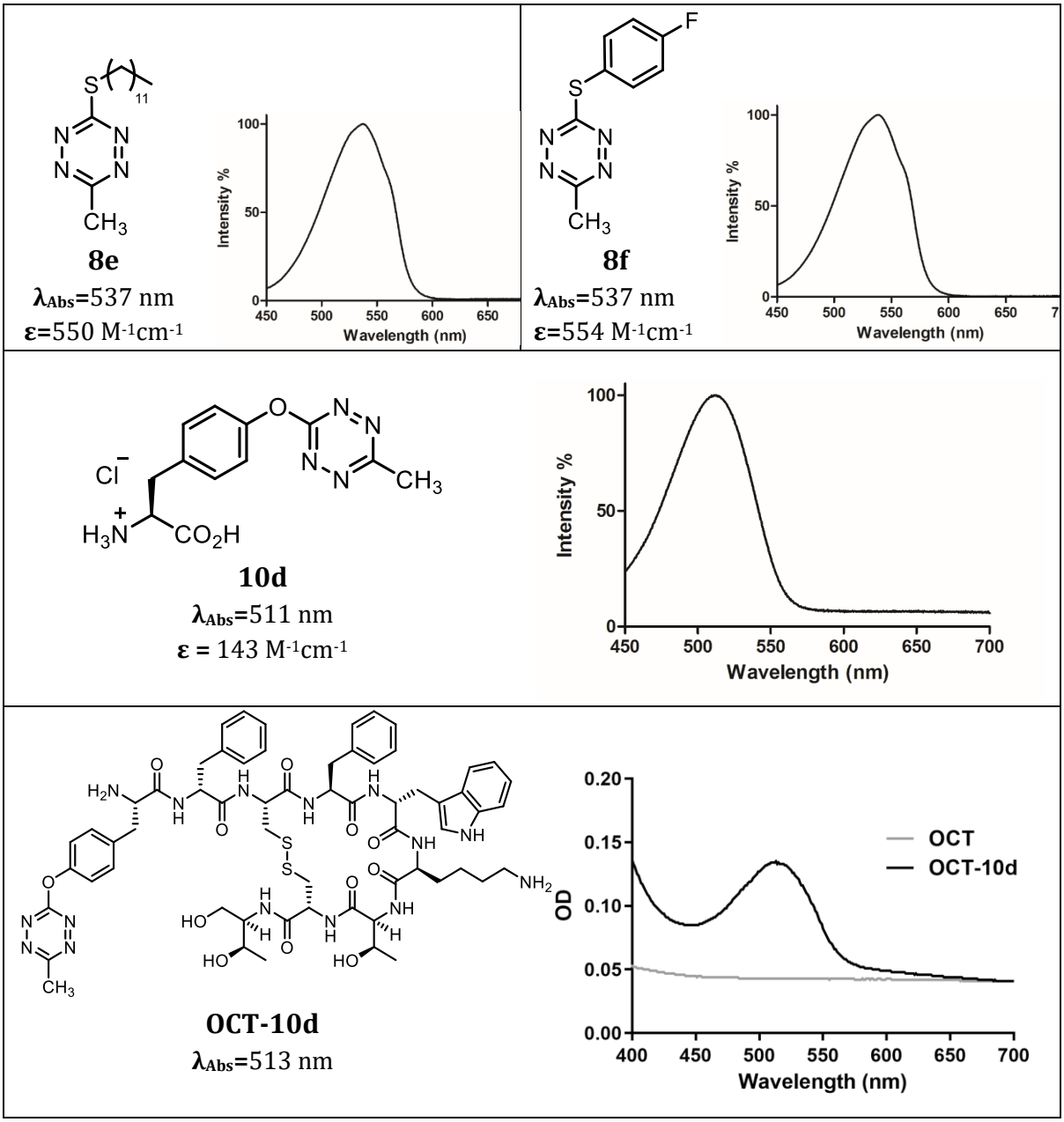
MS (ESI$^+$) calculated for $C_{61}H_{80}O_{12}N_{15}S_2$ $[M+H]^+$: 1278.55468 / found 1278.55360

3.7.2. Spectroscopic determinations

3.7.2.1. Absorbance spectra determination:

Absorbance of compounds **8a-f** (in DCM) and **10d**, **OCT** and **OCT-10d** (water) at a final concentration of 1 mM were recorded on a Varian Cary 100 Bio UV spectrophotometer:

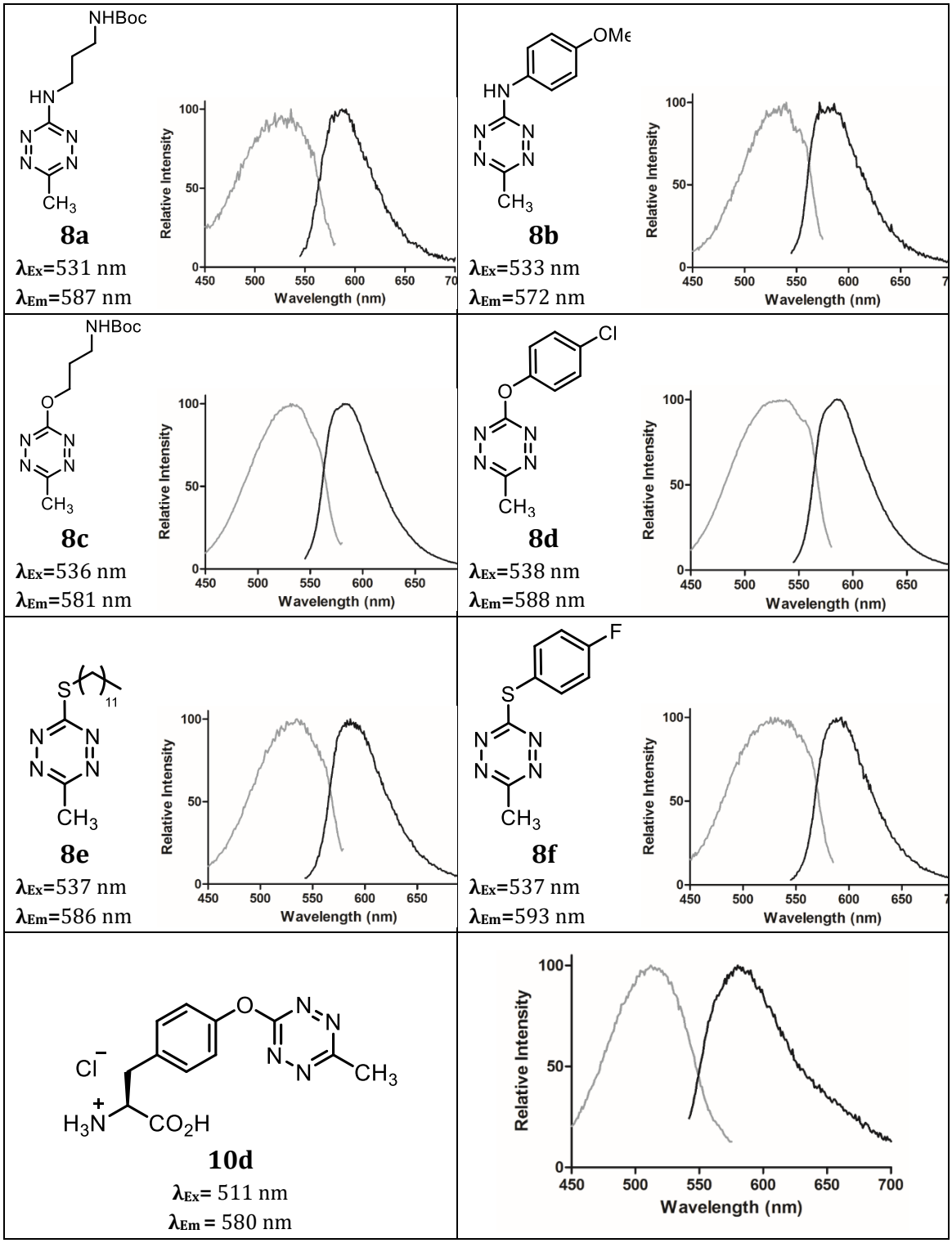




3.7.2.2. Excitation-emission spectra determination:

Fluorescence emission spectra of compounds **8a-e** in DCM and **10d** in methanol were recorded on a Horiba-Jobin-Ybon SPEX Nanolog spectrofluorimeter from dilute solutions with an absorbance value of 0.2 AU at the respective excitation wavelength (λ_{Ex}).

* λ_{Ex} is represented with a grey line. λ_{Em} is represented with a black line.



3.7.2.3. Quantum yield (Φ_F) determination:

Emission quantum yields were determined by employing Rhodamine B as a reference (Φ_F = 0.50, 25 °C, EtOH) and using the formula from Demas and Crosby[41] restricted to the comparison of compounds which can be excited at a common wavelength:

$$\Phi_x = \Phi_r \frac{A_x}{A_r}\left(\frac{\eta_x}{\eta_r}\right)^2$$

where $\boldsymbol{\eta}$ is the index of diffraction and **A** the integration of the peak.

Compounds **8a-e** and **10d** were dissolved in DCM and diluted to achieve an absorbance value of 0.2 AU at the excitation wavelength ($\boldsymbol{\lambda_{Ex}}$). Measurements were taken in duplicates. The obtained results are summarized in **Table 3-5**.

Table 3-5: Raw values of peak areas for each of the fluorescence measurements of compounds 8a-f and 10d. The calculated quantum yield (Φ_F) percentage is also annotated.

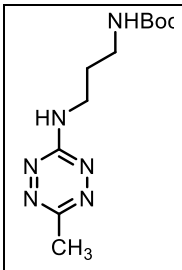
Compound	A_{x1}	A_{r1}	A_{x2}	A_{r2}	Φ_F (%)**
8a	2.71E+06	4.53E+08	2.85E+06	4.65E+08	**0.3**
8b	1.12E+08	4.89E+08	9.59E+07	4.29E+08	**12.2**
8c	1.43E+07	4.87E+08	1.41E+07	4.68E+08	**1.6**
8d	1.15E+08	4.99E+08	1.01E+08	4.41E+08	**12.4**
8e	8.80E+06	4.83E+08	7.84E+06	4.45E+08	**1.0**
8f	5.83E+06	4.92E+08	5.91E+06	4.85E+08	**0.6**
10d* (in MeOH)	4.79E+08	2.30E+09	4.78E+08	2.11E+09	**11.5**
10d (in H_2O)	3.24E+07	7.79E+08	3.22E+07	7.43E+08	**2.0**

* Rhodamine B emission in methanol (Φ_F = 0.53, 25 °C).

** $\eta_x\ (DCM) = 1.413$; $\eta_x\ (EtOH) = 1.361$; $\eta_x\ (MeOH) = 1.331$

3.7.2.4. Emission after iEDDA reaction:

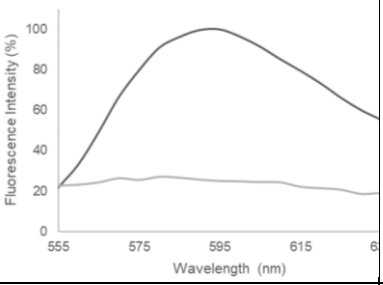
Additionally, fluorescence decay was assessed upon reaction of compounds **8a-e** with **BCN-OH** (excess) in methanol and **OCT-10d** with **BCN-PEG** (excess) in water, in both cases at room temperature. Reactions were performed in clear flat-bottomed 96-well plates, and the emission spectra from 555 to 640 nm were obtained before (blue) and after (red) adding **BCN-OH** in excess, exciting the sample at 535 nm in a Synergy H1 Hybrid Multi-Mode plate reader (BioTek™). The peaks are shown considering the initial fluorescence of compounds **8a-e** as 100% at their emission maximum.

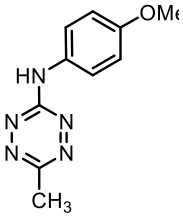


8a

λ_{Ex}=531 nm

λ_{Em}=587 nm

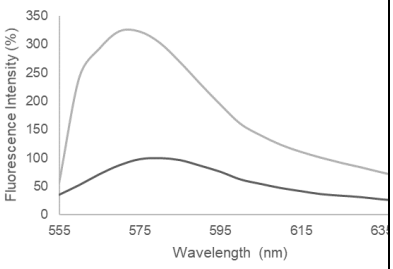


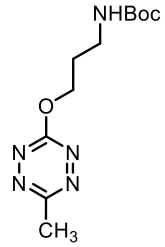


8b

λ_{Ex}=533 nm

λ_{Em}=572 nm

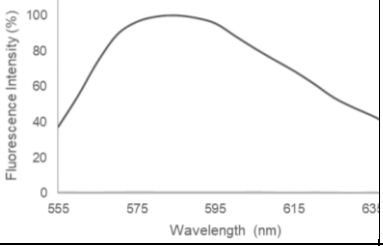


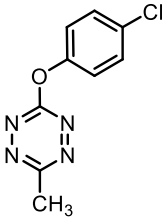


8c

λ_{Ex}=536 nm

λ_{Em}=581 nm

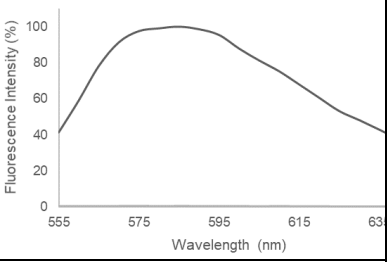


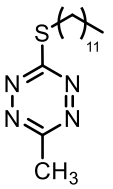


8d

λ_{Ex}=538 nm

λ_{Em}=588 nm

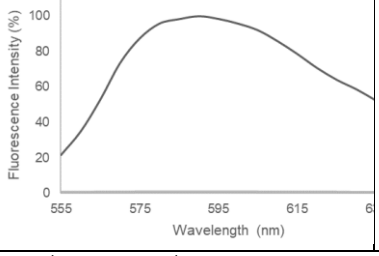


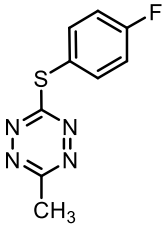


8e

λ_{Ex}=537 nm

λ_{Em}=586 nm

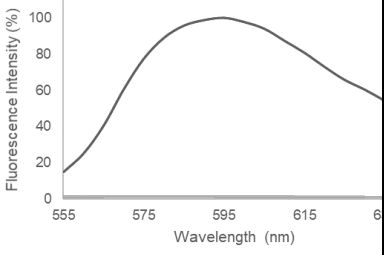


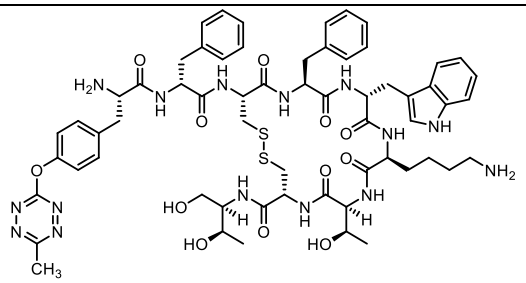


8f

λ_{Ex}=537 nm

λ_{Em}=593 nm

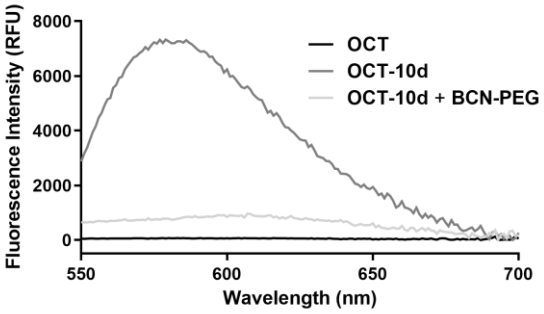




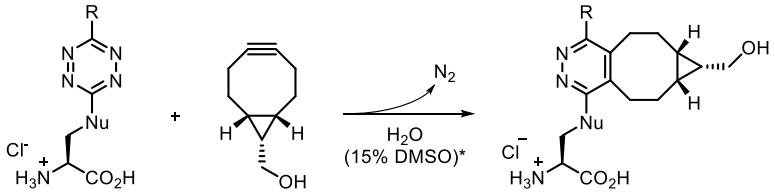
OCT-10d

λ_{Ex}=513 nm

λ_{Em}=579 nm



3.7.3. 2nd Order rate constant determinations *via* UV-Vis spectroscopy:



The second order rate constant (k_2) was approximated using a pseudo-first-order reaction, by using a controlled excess of compound **B**.

$$\frac{-d[A]}{dt} = k_2[A][B] \qquad when\ [B_0] \gg [A_0],\ then\ [B_0] \approx [B]$$

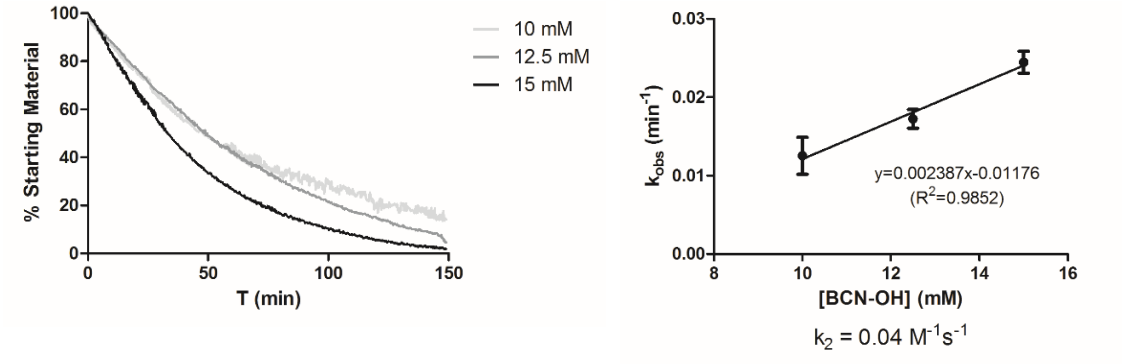
$$\frac{-d[A]}{dt} = k_{obs}[A] \qquad where \qquad k_{obs} = k_2[B]$$

The k_{obs} was determined at 37 ºC in clear flat-bottomed 96-well plates. Compounds **10a-e**, **11a** and **11e** were dissolved in water (to obtain a final concentration of 1 mM), and **BCN-OH** in DMSO was used in excess as the dienophile at a final concentration of 10, 12.5 and 15 mM respectively (except for compound **10c, 10e** and **11a**, where final concentrations of **BCN-OH** were 10, 15 and 20 mM; and **11e**, where used concentrations were 5, 7.5 and 10 mM). Final concentration of DMSO was 15%. The decay of the tetrazine absorbance was monitored in a Synergy HTX Multi-Mode plate reader (BioTek™) at 500 nm, and each different **BCN-OH** concentration for each compound was measured in triplicate. The change of absorbance over time was fitted to an exponential decay equation to determine its observed rate constant (k_{obs}).

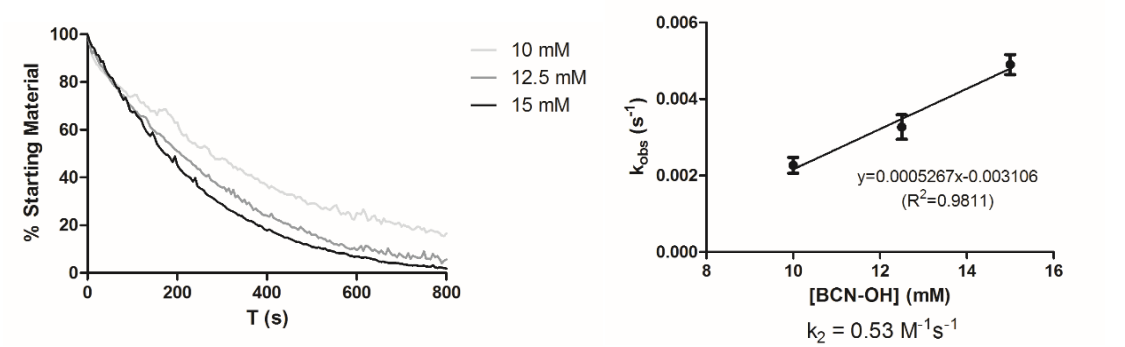
$$\frac{-d[A]}{dt} = k_{obs}[A]$$

The obtained k_{obs} of the 3 reactions were plotted against **BCN-OH** concentrations, fitted to a linear equation in order to determine the slope which corresponds to the second order rate constant (k_2).

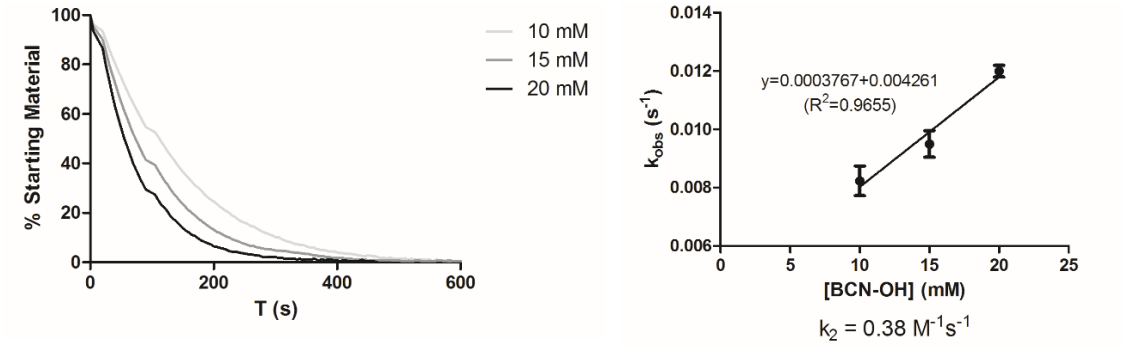
Absorbance decay and k_2 extrapolation for 10a:



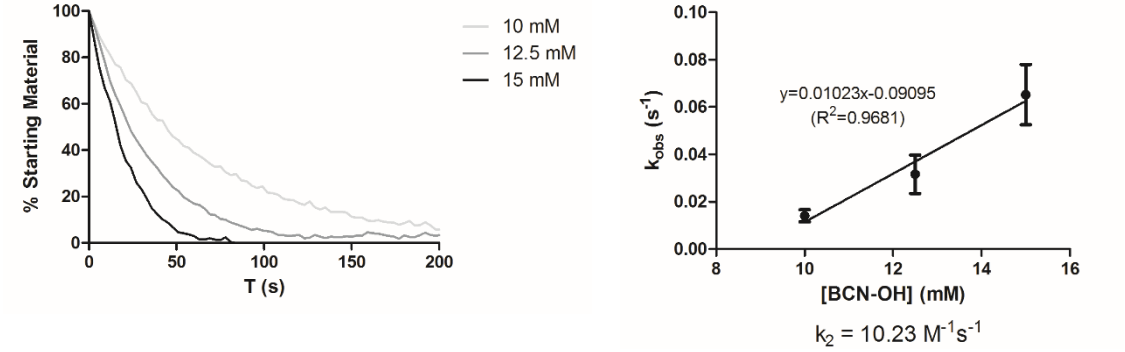
Absorbance decay and k_2 extrapolationor 10b:



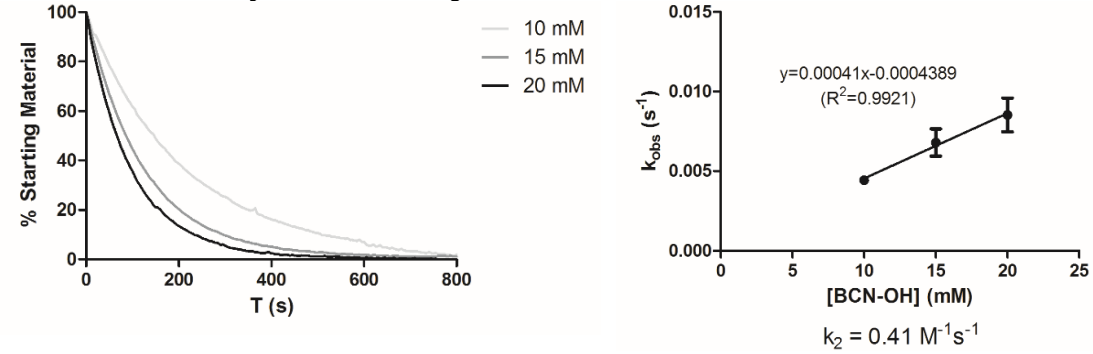
Absorbance decay and k_2 extrapolation for 10c:



Absorbance decay and k_2 extrapolation for 10d:

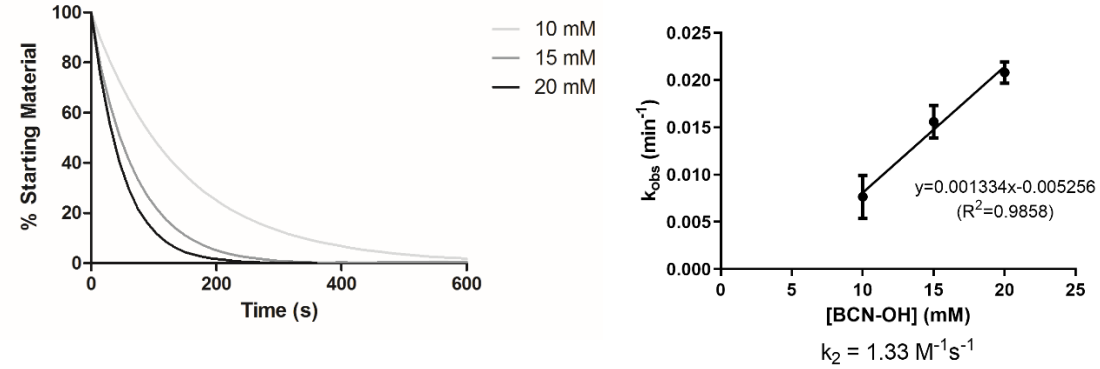


Absorbance decay and k_2 extrapolation for 10e:



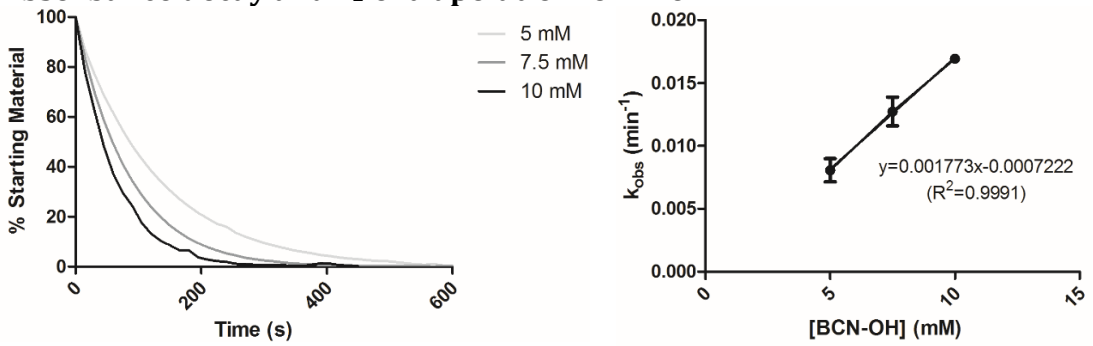
k_2 = 0.41 $M^{-1}s^{-1}$

Absorbance decay and k_2 extrapolation for 11a:



k_2 = 1.33 $M^{-1}s^{-1}$

Absorbance decay and k_2 extrapolation for 11e:



k_2 = 1.77 $M^{-1}s^{-1}$

Reaction rate constant of compounds **11b-11d** could not be determined accurately through UV-Vis spectroscopy because not enough measurements could be obtained before the reaction with different equivalents of **BCN-OH** was finished.

3.7.4. Stability of **10d** in PBS and FBS determined *via* UV-Vis spectroscopy:

10d was dissolved in Phosphate Buffered Saline (PBS) pH=7.4 and Fetal Bovine Serum (FBS) respectively at a final concentration of 1 mM, and incubated at 37 ºC for a total time of 48 h in a flat-bottomed 96-well plate. Absorbance measurements were taken at λ = 510 nm at different time points (t = 0.5 h, 1 h, 2 h, 4 h, 6 h, 8 h, 24 h and 48 h).

3.7.5. Flow cytometry experiments of **OCT-10d**:

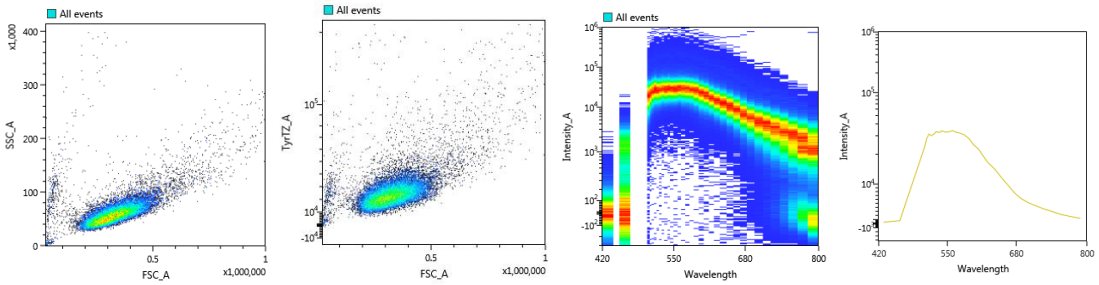
Chinese Hamster Ovary K1 (CHO-K1) cells were grown in Kaighn's modification of Ham's F-12K medium (F-12K, ThermoFisher Scientific) supplemented with 10% fetal bovine serum (FBS), 25 mM HEPES, 100 U/mL penicillin and 100 µg/mL streptomycin at 37 ºC, 5% CO_2.

For flow cytometry studies, in triplicates, cells at 70-80% confluence were incubated with solution of OCT, **OCT-10d** or **OCT-10d-BCN-PEG** in water at a final concentration of 250 µM for 14 h in flat-bottomed 24-well plates with 500 µL of culture medium. Then, the media was taken, cells washed with PBS once and fresh media added. In the cases were **BCN-PEG** was added (**OCT-10d** + **BCN-PEG**), 10 µL from a 50 mM solution in water was added (final concentration of 1 mM).

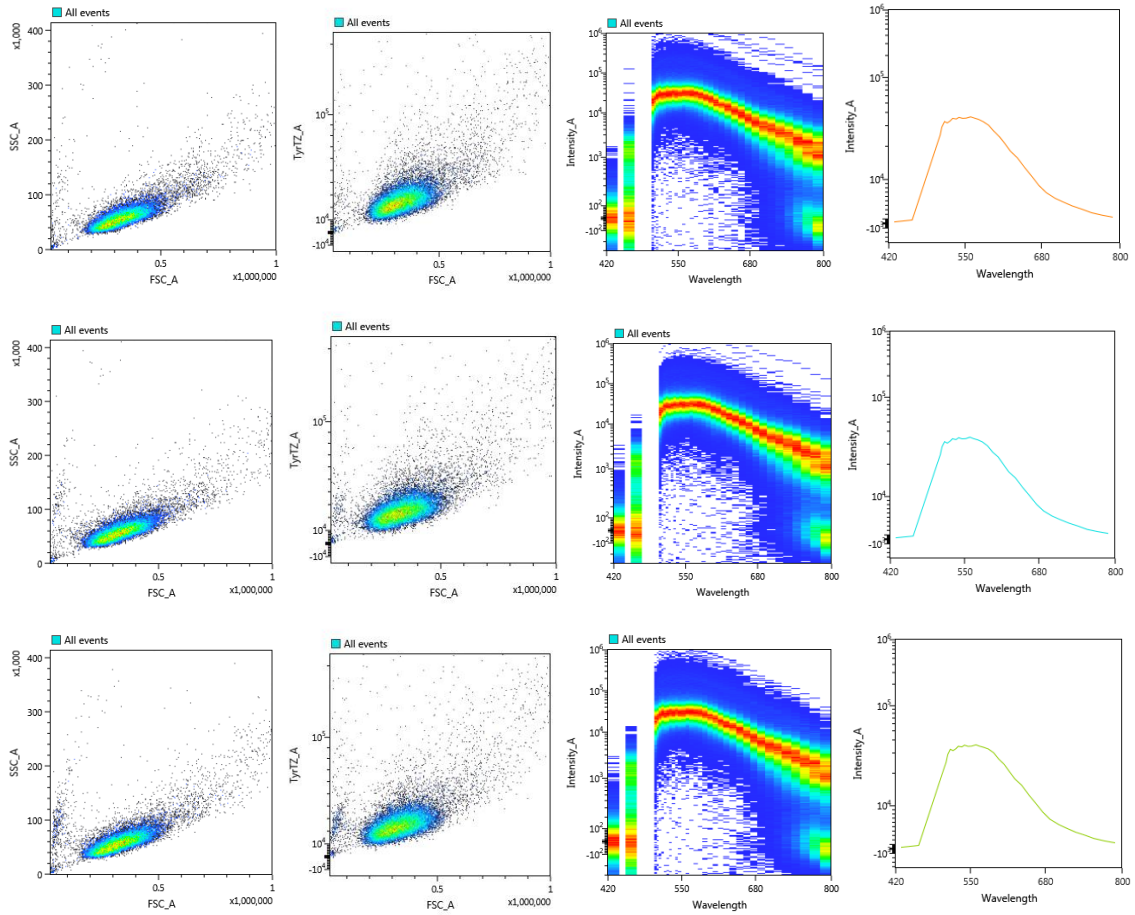
After 2 hours, cells were trypsinised, quantified ($1.1 \cdot 10^6$ total cells), washed with PBS once and resuspended in PBS at a final concentration of 1.5 cells/mL. Cells were then processed in a SA3800 Spectral Cell Analyzer (Sony Biotechnology) with excitation at 488 nm (Window Extension Normal, Threshold CH=FSC, Threshold Value 16.7%, FSC Gain=10, SSC Voltage 27%, Fluorescence PMT Voltage 85%).

The density plots of the total cells (SSC vs FSC), the density plots of the fluorescence of total cells (Fluorescence vs FSC), the fluorescence spectrum of 20000 cells treated for each condition and the mean emission spectra are:

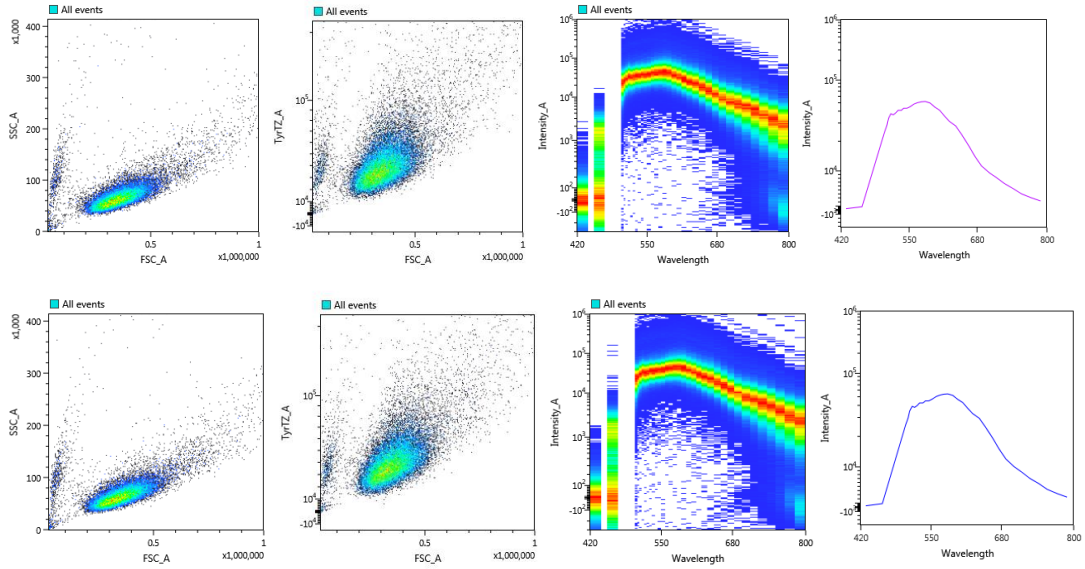
CHO-K1 cells (without treatment):

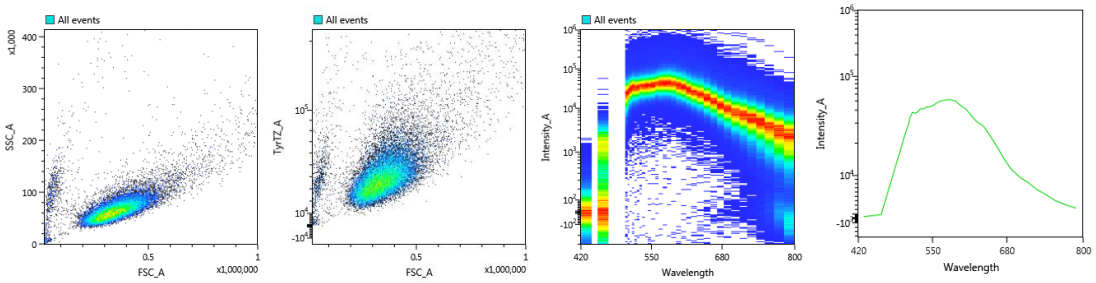


OCT treated cells:

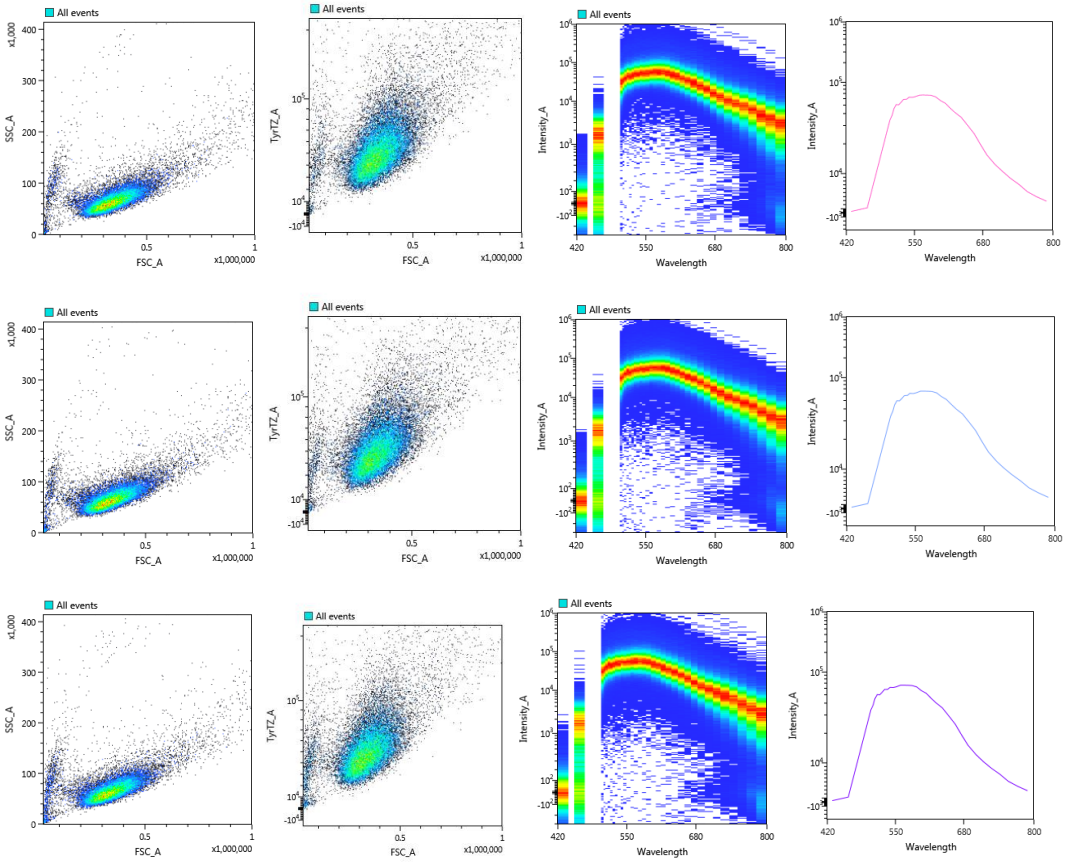


OCT-10d treated cells:

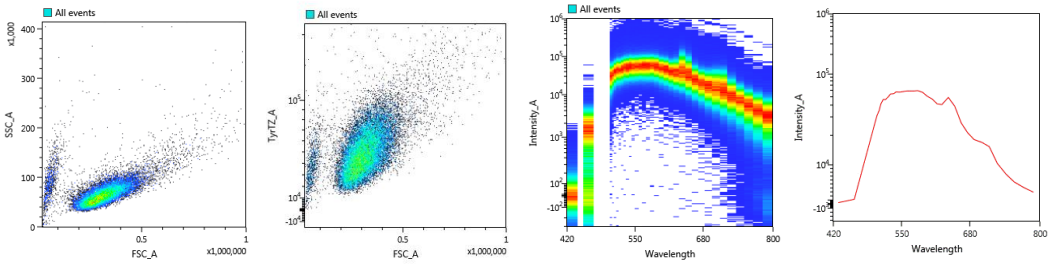




OCT-10d + BCN-PEG treated cells:



OCT-10d-BCN-PEG treated cells:



Chapter 4

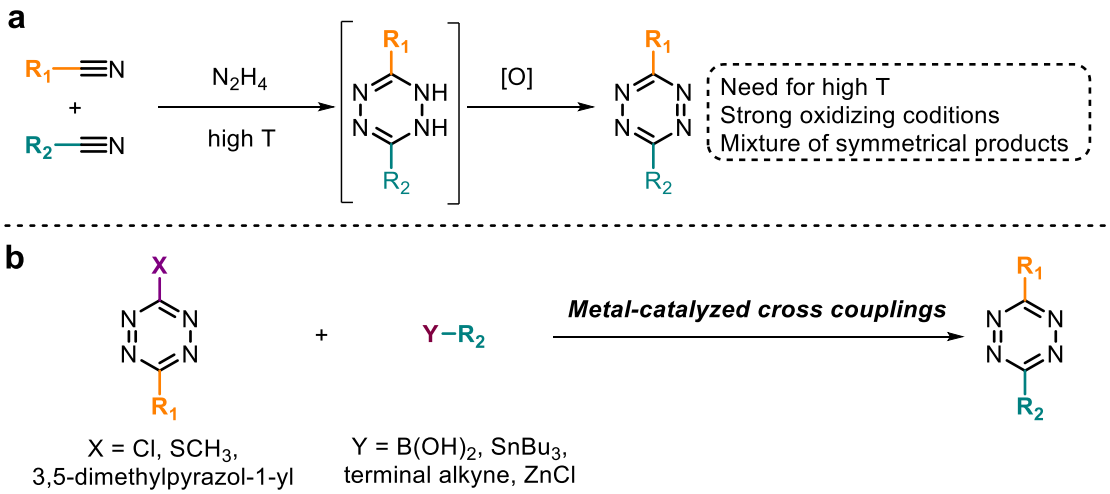
Small molecule functionalization with 3-bromo-1,2,4,5-tetrazines through Pd-catalyzed cross-coupling reactions

Synthesis, derivatization and use as CLIPTACs

4.1. Tetrazine functionalization through metal-mediated cross-coupling reactions

4.1.1. Background:

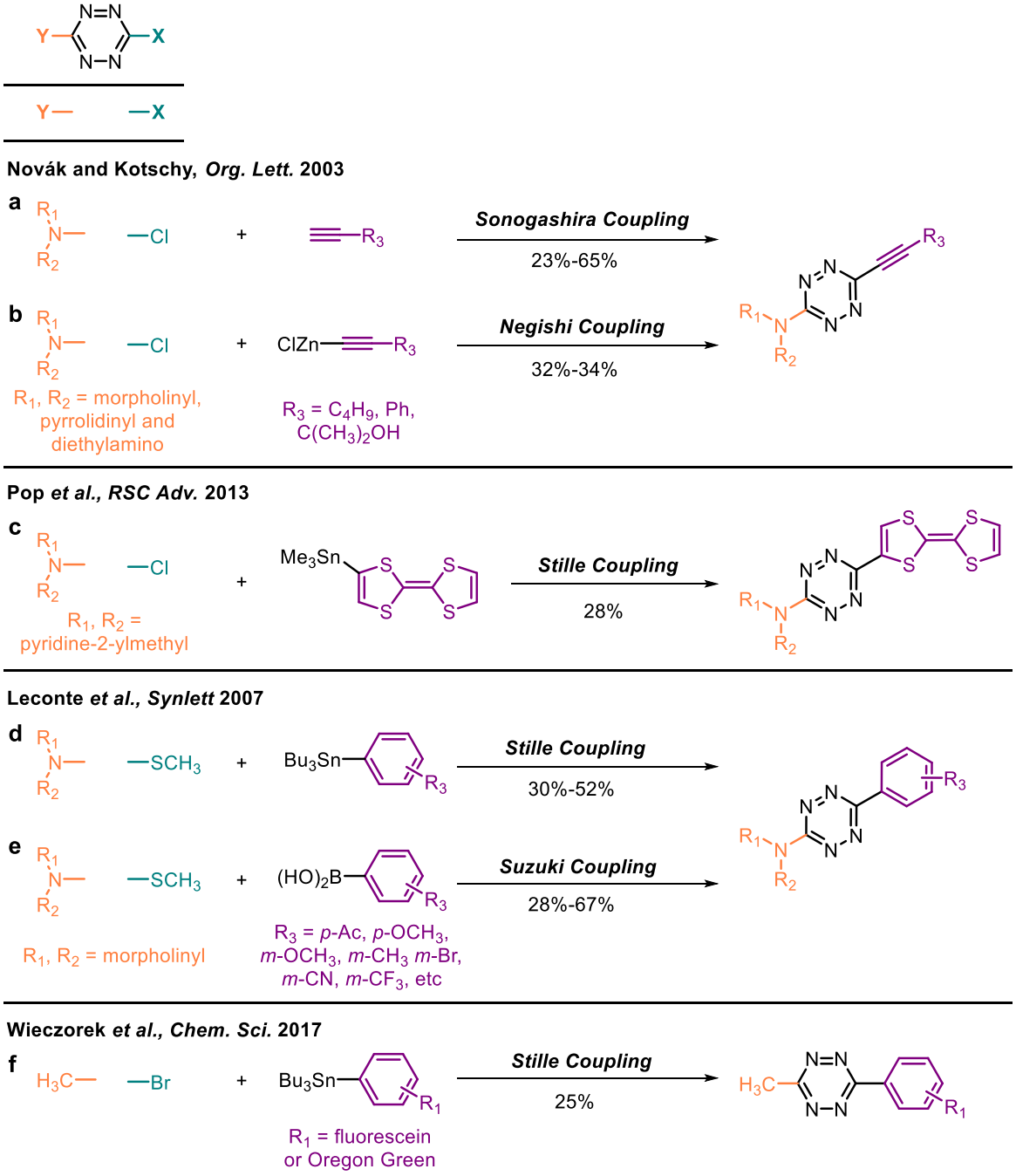
As it has been discussed in **Chapter 2**, certain substrates containing tetrazines are still difficult to synthesize. For instance, late-stage tetrazine functionalization is challenging when strong reaction conditions are required. The need for high temperatures, strong oxidants and acidic media in the nitrile condensation with hydrazine, the most usual method for tetrazine synthesis, may lead to functional group intolerances, thus in problematic syntheses.[1,2] Another challenge arising from this reaction is when the final product is an asymmetrically substituted tetrazine: the formation of a statistical mixture of products, which may be challenging to purify, leads to low yield and inefficient synthesis (**Scheme 4-1, a**).[1,2]



Scheme 4-1: Classic and modern approaches to the tetrazine core synthesis. a) Problematic late-stage tetrazine functionalization. **b)** modern methods employing Pd-catalyzed C-C bond formation reactions to overcome such challenges.

In response to these limitations, new ways to functionalize useful substrates with tetrazines have arisen in the last two decades, employing modern metal-catalyzed cross-coupling reactions with various tetrazines bearing good leaving groups (**Scheme 4-1, a**). The first report of a Pd-catalyzed C-C bond formation reaction with a tetrazine precursor dates back to 2003, when the Sonogashira coupling[3–5] between terminal alkynes and electronically enriched chlorotetrazines was described. Morpholinyl-, pyrrolidinyl- and diethylamino- chlorotetrazines were used as

precursors, affording moderate yields with a set of different aliphatic and aromatic alkynes (**Scheme 4-2, a**).[6] In the same article, a Negishi coupling[7,8] between the same chlorotetrazines and 1-hexynylzinc chloride was also reported (**Scheme 4-2, b**).[6] Despite being the first example in the literature with alkynyl-tetrazines, their reactivity in iEDDA cycloadditions and their potential use in bioorthogonal reactions has not been further explored.



Scheme 4-2: Direct metal-catalyzed cross-coupling reactions with tetrazines described to date.

Another electronically enriched chlorotetrazine, substituted by a bis(pyridine-2-ylmethyl)amine, has also been employed in Stille-type coupling[9,10] with a tetrathiafulvalene(trimethyl)stannane derivative (**Scheme 4-2, c**).[11] The products were employed as donor-acceptor systems and multifunctional ligands. Interestingly, not only halogenated tetrazines can be employed as precursors in cross-coupling reactions. 3-Methylthio-6-(morpholin-4-yl)-1,2,4,5-tetrazine can also be used in both Suzuki[12,13] (**Scheme 4-2, d**) and Stille-type[9,10] couplings (**Scheme 4-2, e**) with a wide variety of aryl boronic acids or aryl(tributyl)stannanes.[14]

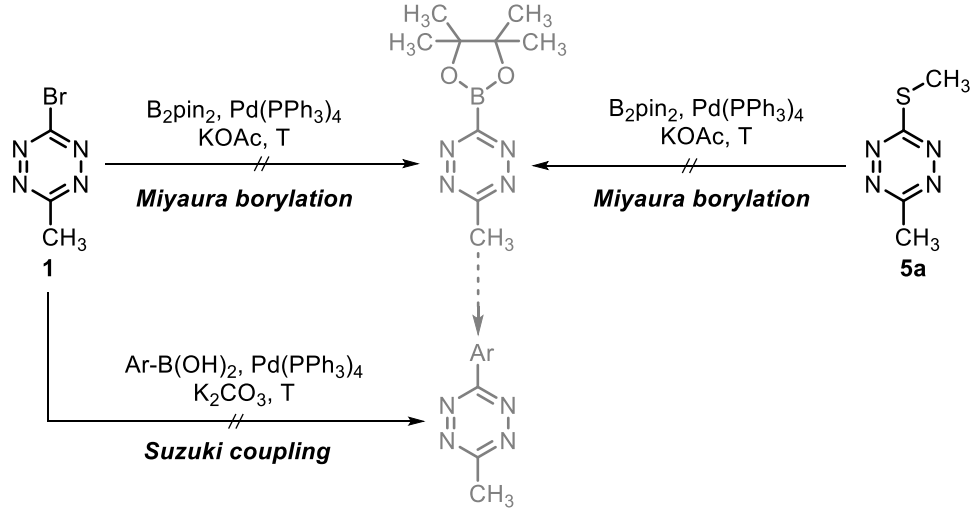
Noticeably, all abovementioned examples of cross-couplings are between tetrazines substituted with EDGs, which are not ideal for the kinetics of the iEDDA cycloaddition, thus for their bioorthogonal application (a detailed explanation is provided in **Chapter 3**, **Section 3.3.1.**). An improvement to the bioorthogonal applicability for the products of these reactions was reported when a set of fluorescein and Oregon Green derivatives were synthesized via a Stille coupling, employing 3-bromo-6-methyl-1,2,4,5-tetrazine (**1**) as precursor (**Scheme 4-2, f**).[15] The obtained dyes were successfully utilized in no-wash live-cell imaging.

However, the reduced number of substrates synthesized, the low to moderate yields obtained and the absence of methods to afford dialkyl-1,2,4,5-tetrazines are limitations for the widespread use of these metal-mediated cross-coupling reactions to generate interesting compounds to use in iEDDA bioorthogonal cycloadditions.

4.1.2. Results:

We envisioned **1** or **2** as potential tetrazine reagents to undergo metal-mediated C-C bond formation reactions for late-stage functionalization. We based our predictions on two observations: firstly, on the well-known fact that aryl bromides exhibit greater reactivity in cross-coupling reactions than aryl chlorides, which tend to be much more inert.[16] According to examples from **Section 4.1.1.**, it seems clear that chlorinated tetrazines require EDGs as substituents, whilst brominated tetrazines can participate in cross-coupling reactions without requiring any activator. Secondly, both the products of **1**, **2** or **3** would lead to stable tetrazines with fast kinetics in the iEDDA cycloaddition, which would be ideal for bioorthogonal purposes, where a trade-off between reactivity and stability is of major importance.[17]

Due to substrate availability, we screened different Pd-catalyzed reactions using 3-bromo-6-methyl-1,2,4,5-tetrazine **1** (**Scheme 4-3, left**). We initially focused on Suzuki-type cross-couplings with boronic acids. Unfortunately, these initial attempts evidenced that **1** underwent undesired reactions when using inorganic bases. Probably, the high electrophilicity of **1** prompts the formation of the S_NAr product with free hydroxyls in strong alkaline conditions, leading to the formation of 3-hydroxy-6-methyl-1,2,4,5-tetrazine. Also, we observed degradation of compound **1** when we heated the reaction mixture over 100 ºC. Next, we considered a functional group interchange between the bromine of **1** into a boronic acid, which we thought would enable its participation in Suzuki couplings. Consequently, we attempted the formation of (6-methyl-1,2,4,5-tetrazin-3-yl)boronic ester through direct Miyaura borylation of **1**. Unfortunately, we encountered the same incompatibility with inorganic bases, and when the reaction was attempted with milder organic bases such as triethylamine, no reaction was observed. Encouraged by the use of (methylthio)tetrazines in cross-coupling reactions, we attempted the Miyaura borylation on 3-methyl-6-(methylthio)-1,2,4,5-tetrazine **5a** (**Scheme 4-3, right**). After several test reactions, only traces of the desired product were observed, which prompted us to discontinue this approach and explore other reaction types.

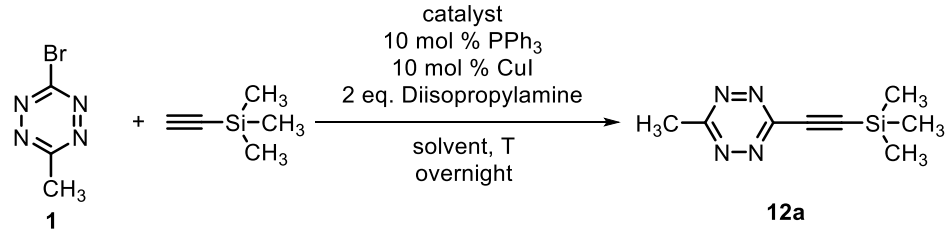


Scheme 4-3: Failed approaches for Suzuki-type cross-couplings with tetrazines 1 and 5a.

Due to the limitations encountered for **1**, which tolerates neither inorganic bases nor high temperatures, we decided to test a Sonogashira-type reaction, as it normally requires neither of these two factors. Taking as a reference previous work on this reaction with halogenated tetrazines,[6] we decided to test the coupling between **1** and

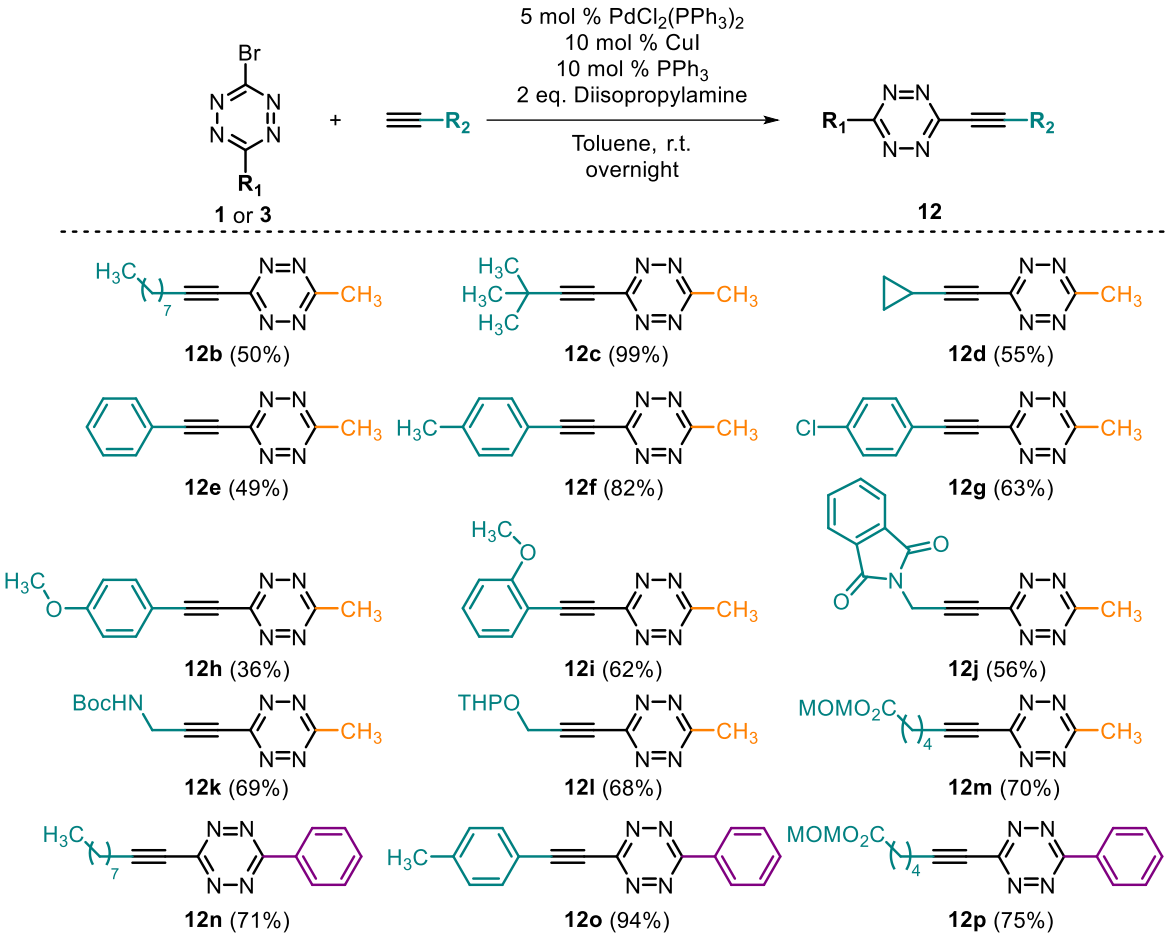
trimethylsilylacetylene, as it was reported that all attempts with this substrate proved unsuccessful and only decomposition could be observed.[6] We envisaged that if we achieved the formation of the coupled product 3-methyl-6-((trimethylsilyl)ethynyl)-1,2,4,5-tetrazine (**12a**), it would be a good starting point to explore further the stability and reactivity of other alkynyl-tetrazines (**Table 4-1**).

Table 4-1: Screening of reaction conditions for Sonogashira coupling of 1. The reaction was performed in a degassed, sealed vial with 0.057 mmol of **1** [0.15 M], using 5 mol % of catalyst loading and 2 eq of TMS-acetylene as the model alkyne. *Determined by ^{1}H NMR spectroscopy. **Yield determined by ^{1}H NMR spectroscopy using mesitylene as an internal standard.



Entry	Catalyst	T	Solvent	Conv. (%)*	Yield (%)**
1	$PdCl_2$	r.t.	Toluene	>99	57
2	$Pd(OAc)_2$	r.t.	Toluene	>99	54
3	$Pd(PPh_3)_4$	r.t.	Toluene	>99	23
4	$PdCl_2(PPh_3)_2$	r.t.	Toluene	>99	75
5	$PdCl_2(PPh_3)_2$	r.t.	DCM	>99	56
6	$PdCl_2(PPh_3)_2$	r.t.	THF	>99	5
7	$PdCl_2(PPh_3)_2$	50	Toluene	>99	66
8	$PdCl_2(PPh_3)_2$	100	Toluene	>99	53

Using 5 mol % of Pd^{II} precatalysts (either $PdCl_2$ or $Pd(OAc)_2$) (**Table 4-1,** entries **1** and **2**) in toluene, the desired product **12a** was formed in moderate yields. The use of $Pd(PPh_3)_4$ as catalyst gave lower yields (**Table 4-1,** entry **3**). In contrast, we were pleased to observe that an excellent 75% yield was obtained using $PdCl_2(PPh_3)_2$ (**Table 4-1,** entry **4**)., which is the best yield ever obtained for a direct cross-coupling into the tetrazine ring. Using more polar solvents, like dichloromethane or tetrahydrofurane (**Table 4-1,** entries **5** and **6**) or increasing the temperature (**Table 4-1,** entries **7** and **8**), worsened the yields. We concluded that, under these mild conditions, bromotetrazine **1** is an excellent reagent to undergo Sonogashira couplings with terminal alkynes, hence we explored the substrate scope of the different bromotetrazines using a range of terminal alkynes (**Scheme 4-4**).



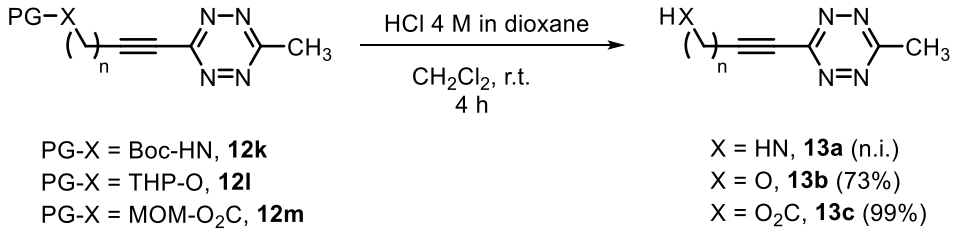
Scheme 4-4: Substrate scope of the Sonogashira reaction between 1 or BrTzPh with aliphatic and aromatic alkynes.

Coupling between **1** and aliphatic alkynes afforded acetylenic tetrazines **12b-d** in good to excellent yields. Likewise, aromatic alkynes with different substitution patterns (phenyl, tolyl, *p*-chlorophenyl, *o*- and *p*-methoxyphenyl) were tested yielding the desired products (**12e-i**) in moderate to good yields. As it could be anticipated, the coupling with alkynes presenting free alcohols or amines did not afford the desired products, giving mainly the SNAr products, whereas couplings with free carboxylic acids led to decomposition products. Therefore, we needed to use protecting groups. Acetylenic amines were protected either as phthalimides (**12j**) or *tert*-butyl carbamates (**12k**); hydroxyls were protected as tetrahydropyranyl ethers (**12l**) and acetylenic esters were tested with good results (**12m**). In all cases the coupling products were obtained in good yields.

Importantly, when we attempted the same coupling with 3-bromo-1,2,4,5-tetrazine **2**, we were unable to repeat the results obtained with **1**, and only degradation products were recovered. Whether **2** could be a suitable reagent to undergo

Sonogashira cross-coupling reactions under a different set of conditions, or even other type of cross-couplings, is still unknown. However, when we attempted different couplings with bromotetrazine **3**, we afforded the desired products (**12n-p**) with comparable yields, expanding the possibility to perform Sonogashira cross-couplings with other bromotetrazines beyond **1** or **3**.

The next step was to find good deprotection conditions for the products bearing protective groups. Since strong alkaline conditions could not be employed because they led to degradation of the tetrazine, we selected a set of acid-labile protecting groups (**Scheme 4-5**). Using a diluted solution of HCl in dioxane, compounds **12k-m** led to the desired free amine, alcohol and carboxylic acid (**13a-c**).



Scheme 4-5: Deprotection of acid-labile protective groups for alcohols, amines and carboxylic acids (n.i. = not isolated).

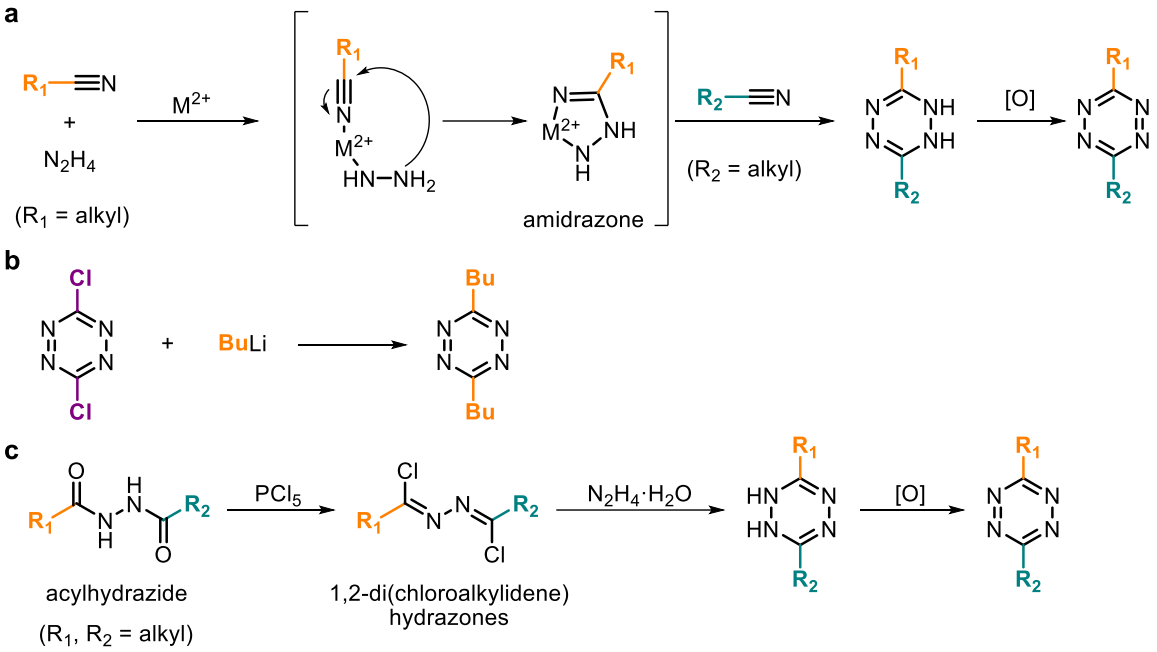
Even though alkynyl-tetrazines **12a-p** are a potentially interesting new class of compounds which have never employed in bioorthogonal iEDDA reactions, we envisaged they could be of greater utility to access a synthetically challenging class of tetrazines, namely 3,6-(dialkyl)tetrazines.

4.2. Synthesis of dialkyl-1,2,4,5-tetrazines

4.2.1. Background:

As it has been previously highlighted in **Chapter 2** (**Section 2.1.1.**), the synthesis of 3,6-(dialkyl)tetrazines is not straightforward. Aliphatic nitriles are less activated than aromatic ones, hence their condensation with hydrazine requires the presence of auxiliary reagents to afford the desired products in decent yields. In 2012, Devaraj's group reported the use of Ni^{+2} or Zn^{+2} salts (primarily $Ni(OTf)_2$ or $Zn(OTf)_2$) as catalysts to favor the reaction.[18] These metal ions act as Lewis acids, coordinating to the starting alkyl nitriles and activating them, thereby facilitating the generation of the amidrazone intermediate (**Scheme 4-6, a**).[18] The obtained dihydrotetrazines

must be oxidized *in situ* to afford the resulting tetrazines in moderate yields. However, the typical drawbacks for this synthesis remain: if asymmetric tetrazines are the desired products, the formation of the symmetrical ones is inevitable. Moreover, the presence of hydrazine, that can act both as a good nucleophile and a reductant, may bring some functional group intolerabilities, specially under heating.[19]

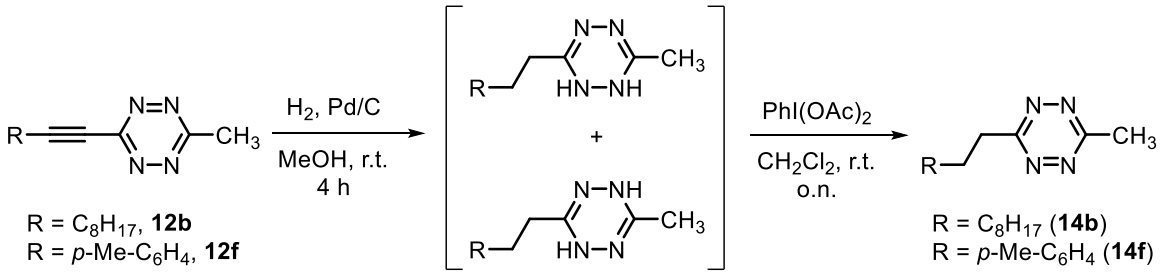


Scheme 4-6: Synthesis of 3,6-(dialkyl)tetrazines. a) Aliphatic nitrile condensation with hydrazine through metal catalysis (M^{2+} = Zn^{+2}, Ni^{+2}). **b)** S_NAr between an electrophilic tetrazine (3,6-dichlorotetrazine) and butyllithium. **c)** Stollé-type[20] synthetic route.

Other methods can also be employed to achieve tetrazines with aliphatic substituents, albeit with variable degrees of success. For instance, the S_NAr reaction with electrophilic tetrazines and organometallic reagents has been reported in the literature. Specifically, the reaction between a set of chlorotetrazines and butyllithium, that yielded the corresponding substitution products in moderate yields (**Scheme 4-6, b**).[21] Alternatively, acylhydrazides can be treated with phosphorous pentachloride to convert them into 1,2-di(chloroalkylidene)hydrazones, which can undergo a condensation reaction with hydrazine and a subsequent oxidation. If the initial acylhydrazide substituents are different aliphatic chains, the generated product will be an asymmetrically substituted 3,6-(dialkyl)tetrazine (**Scheme 4-6, c**).[22]

4.2.2. Results:

Due to limited synthetic strategies available for 3,6-(dialkyl)tetrazines, we decided to use the Sonogashira as synthetic precursors for these types of compounds. To that end, we tested the hydrogenation of the triple bond on two of the Sonogashira coupling products (**12b** and **12f, Scheme 4-7**), which were used as model compounds. When each of them was subjected to hydrogenation at 1 bar of pressure, using Pd/C as catalyst, the disappearance of the characteristic pink coloration of tetrazines from the reaction mixture was observed almost immediately. After 4 hours, TLC chromatography showed a single spot, whereas ^{1}H NMR spectroscopy of the reaction crude showed a mixture of compounds present, which can be pointing towards the formation of the different dihydrotetrazine isomers. When we attempted the re-oxidation to the fully functional tetrazine on the reaction crude employing sodium nitrite in acidic media, the most usual way to oxidize dihydrotetrazines,[1] we observed the formation of the desired products with moderate yields, together with degradation products. Gratifyingly, when the re-oxidation was done with an hypervalent iodine reagent previously reported for dihydrotetrazine oxidation, namely [(diacetoxyiodo)benzene],[30] full conversion to the desired products **14b** and **14f** was obtained under mild conditions. The two-step sequence of hydrogenation/oxidation gave near quantitative yields.

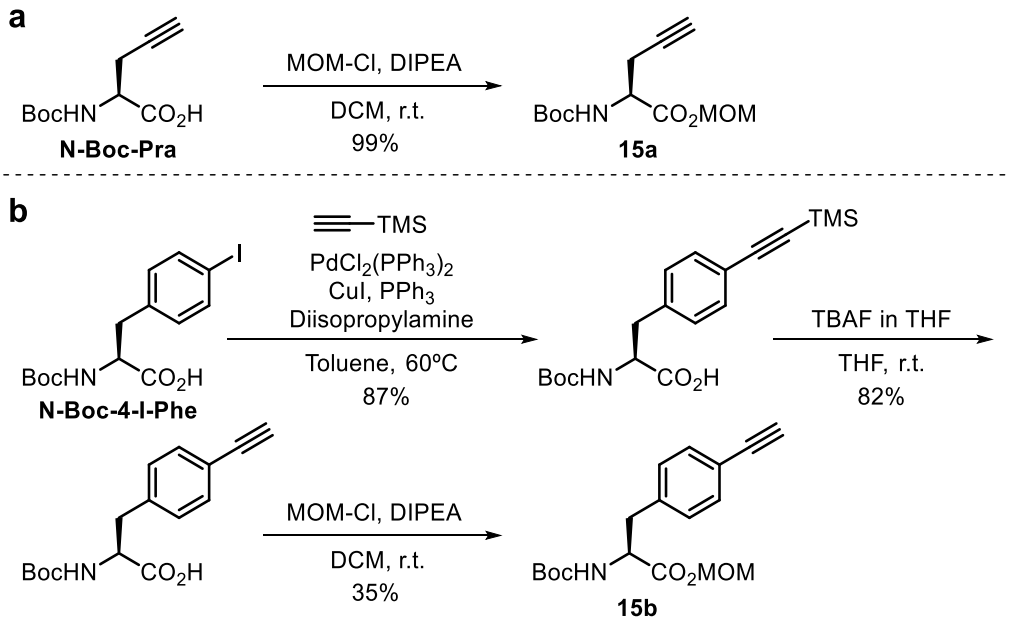


Scheme 4-7: Stepwise hydrogenation/re-oxidation of alkynyltetrazines 12c and 12f taken as model (aliphatic and aromatic) compounds.

Finally, to showcase the versatility of the reported synthetic strategy on biologically relevant molecules, the synthesis of different unnatural amino acids was performed. We pursued two objectives simultaneously: we wanted to study the kinetics of different alkynyl-tetrazines in the iEDDA reaction, and also their stability in physiological media. These two parameters are essential to determine the bioorthogonal applicability of this class of compounds. Just like we previously did

with the S_NAr products (**Chapter 3, Section 3.3.2.**), we preferred to use amino acids due to their excellent aqueous solubility. Secondly, we wanted to test their potential genetic incorporation through GCE techniques.

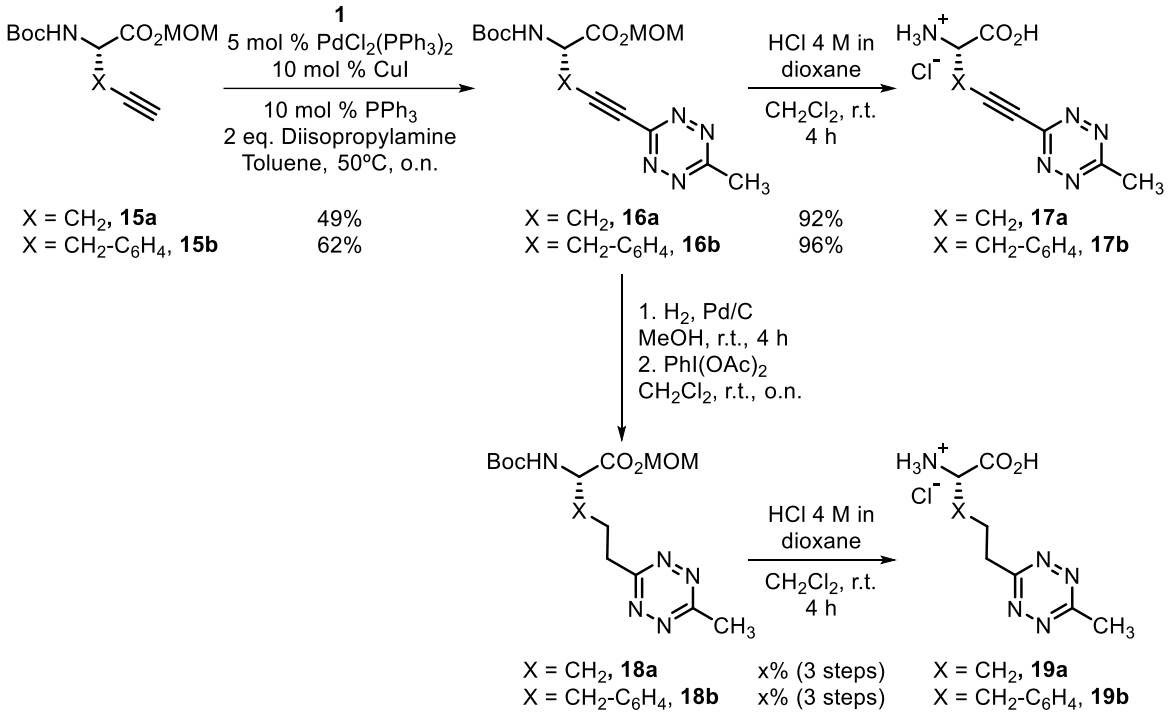
Accordingly, since we could not perform the Sonogashira coupling with the free amino or carboxylic acid groups, we selected N-Boc and MOM protected amino acids as appropriate precursors (**15a-b**, **Scheme 4-8**). For **15a**, we started from commercially available Nα-Boc propargylglycine (**N-Boc-Pra**), which was reacted with methoxymethyl chloride (MOM-Cl) to afford the desired double protected amino acid. For **15b**, the synthesis started from Nα-Boc-4-iodophenylalanine (**N-Boc-4-I-Phe**), which was functionalized with TMS acetylene and the product deprotected with TBAF to yield Nα-Boc-4-acetylenephenylalanine. Then, MOM-Cl was used to generate the MOM ester from the corresponding carboxylic acid.



Scheme 4-8: Synthesis of double protected amino acid precursors 15a-b.

Thus, alkynyl amino esters **15a** (alanine derivative) and **15b** (phenylalanine derivative) were reacted with **1** under the optimized Sonogashira conditions to afford the methyltetrazinyl-amino acid derivatives **16a-b**. Interestingly, heating the reaction mixture at 50 °C was necessary in both cases, as only starting material was recovered when performing this reaction at room temperature. The direct deprotection of these intermediates under acidic conditions yielded the hydrochloride salts **17a-b**, containing an alkynyl-tetrazine functionality. By contrast, when intermediates **16**

were subjected to the sequence of hydrogenation and re-oxidation, followed by deprotection of the Boc carbamate and MOM ester, the dialkyl-tetrazine amino acids **19a-b** were afforded (**Scheme 4-9**).



Scheme 4-9: Synthesis of unnatural amino acids bearing the 4-methyltetrazinyl fragment.

The good overall yields obtained and the straightforward syntheses for both alkyl- and alkynyl- tetrazine-containing amino acids showcase the versatility of the developed reactions, which enabled the generation of compounds that would otherwise be hardly accessible.

4.3. Application of the Sonogashira-coupled products with **1** as bioorthogonal PROTACs

4.3.1. Background:

In order to showcase a potential application for these class of compounds, we envisaged that the synthesis of proteolysis targeting chimeras (PROTACs) would be a suitable area of research. Initially reported in 2015,[23] PROTACs are bifunctional bioactive compounds that promote the degradation of a target protein exploiting the ubiquitin-proteasome system (UPS), by hijacking the enzymes responsible for protein recycling that are inherent in mammalian cells (**Fig. 4-1**).[24]

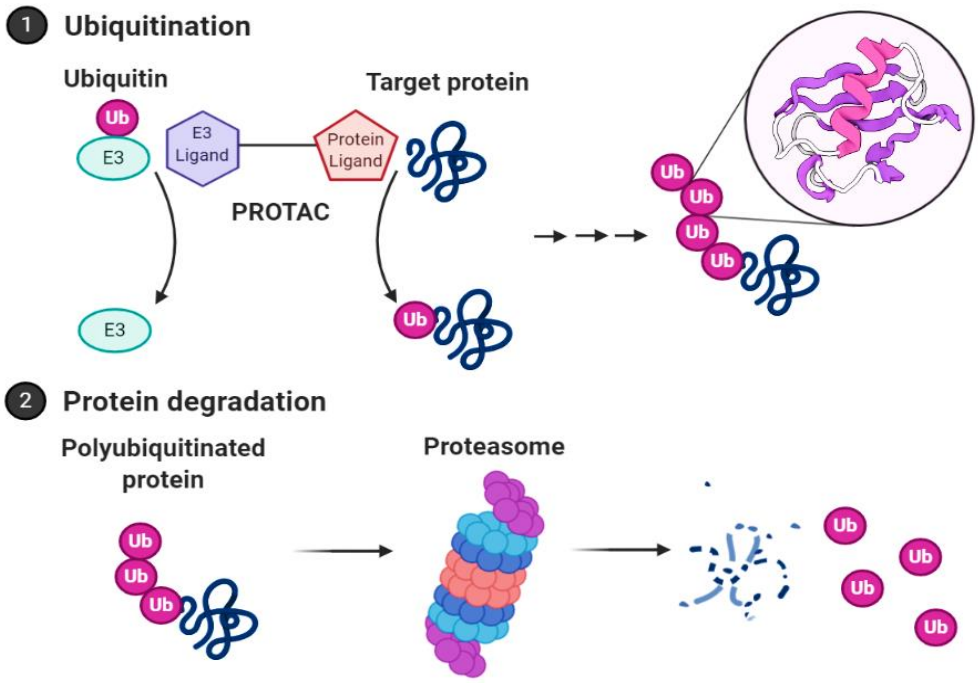
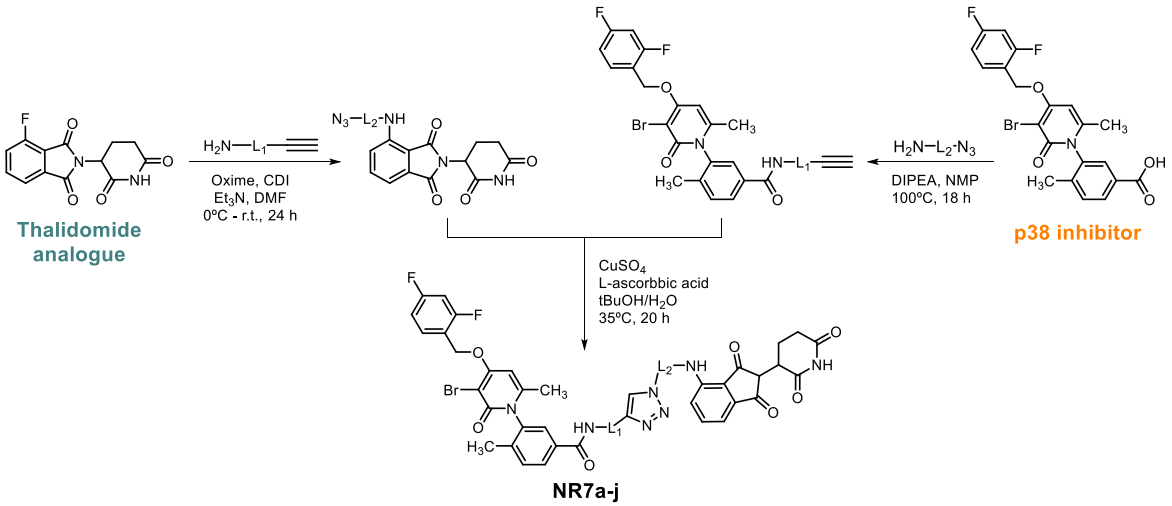


Figure 4-1: Two-step mechanism of action of a PROTAC. Initially, the bifunctional PROTAC brings in close proximity the E3 ligase and the protein of interest, which is ubiquitinated. After several rounds, the polyubiquitinated protein is recognized and degraded by the proteasome.

Structurally, PROTACs are heterobifunctional molecules, formed by a ligand of the target protein linked to a E3 ubiquitin ligase ligand. E3 ligases are endogenous proteins responsible for linking ubiquitin domains on those proteins that need to be degraded and recycled by the cell. By exploiting this system, an active PROTAC exerts its function bringing in close proximity and in the correct orientation the target protein and the E3 ligase, thereby forming a ternary complex which induces ubiquitination of the target protein. Once the protein is polyubiquitinated, it is disassembled from the complex and is eventually degraded by the 26S proteasome.[25] Importantly, achieving protein degradation is an alternative method to the classic enzyme inhibition, which has traditionally been the preferred therapeutic strategy in medicinal chemistry.

Currently, our laboratory is devoted towards the synthesis of PROTACs targeting p38α, a protein involved in the mitogen-activated protein kinase (MAPK) signaling pathway. This pathway comprises the p38 family members (α, β, χ, δ), the ERK and the JNK kinases, and is known to play a key role in the onset and development of cancer. The project involving the synthesis of p38α protein degraders is carried out in collaboration with the Signalling and Cell Cycle Laboratory of the IRB Barcelona,

led by Dr. Angel Nebreda, and aims at finding a potential new therapeutic strategy for certain types of cancer. After a thorough optimization of the synthesis and subsequent screening of potential PROTACs against p38α (a project comprised in Dr. Craig Donoghue's doctoral thesis), several active compounds were found with various degrees of efficacy.[26] Their synthesis is convergent: on the one hand, the ligand for p38α (a known p38α inhibitor initially developed by Pfizer for inflammatory diseases and discontinued after Phase II clinical trials)[27] is synthesized and derivatized with a terminal alkyne; on the other hand, the ligand for the E3 ligase (specifically a thalidomide analogue able to recruit the Cereblon E3 ligase) is derivatized with an azide. Ultimately, both molecules are coupled together through a CuAAC reaction, achieving a straightforward synthesis that enables the obtention of an array of derivatives differing on the linker length (**Scheme 4-10**). In a recent publication, a description of the linker optimization process is presented, showcasing how critical this parameter is.[26] Ultimately, compound **NR-7h**, with a linker length of 16 atoms in total, was selected as the most active analogue.[26]



Scheme 4-10: Synthesis of PROTACs developed in the Research Unit on Asymmetric Synthesis. Both a known p38α inhibitor and a thalidomide analogue are linked together employing copper-catalyzed click chemistry. (Adapted from Dr. Craig Donoghue's doctoral thesis).

Due to the requirement to incorporate both the E3 ligase recruiter and the target protein ligand, PROTACs tend to be large molecules, with molecular weights typically oscillating between 800-1000 Da.[28] This factor, combined with a high polar surface area (~200 Å^2),[28] can hamper the therapeutic applicability of PROTACs because they limit cellular permeation and solubility, thereby compromising biodistribution and

pharmacokinetics. A potential strategy to overcome such limitations relies on the use of in-cell click-formed proteolysis targeting chimeras (CLIPTACs),[28] which are PROTACs coupled directly in the cellular environment employing bioorthogonal reactions. In CLIPTACs, both recruiting functionalities are administered sequentially, each of them bearing a bioorthogonal handle that reacts intracellularly, forming the active PROTAC *in situ* (**Fig. 4-2**). Theoretically, CLIPTACs present an improved solubility and permeability profile because they involve the administration of smaller molecules with better drug-like properties.

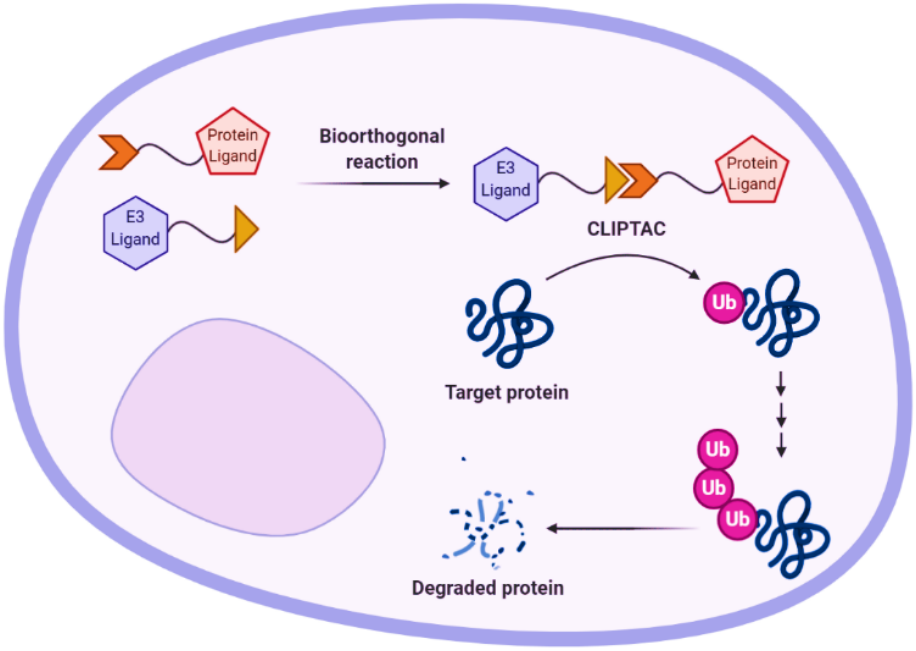


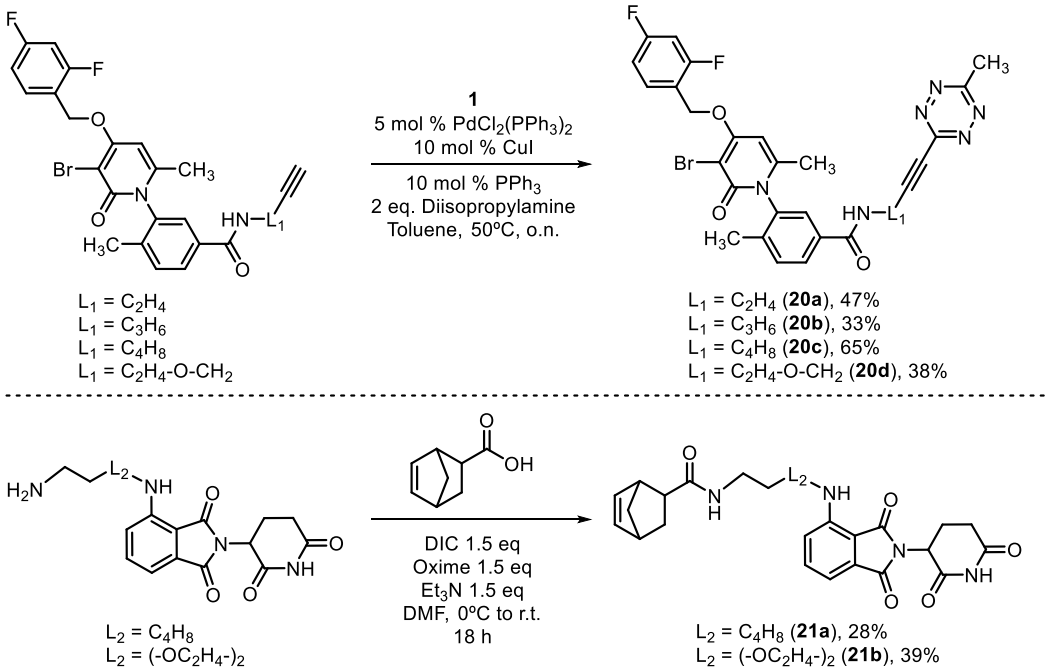
Figure 4-2: CLIPTACs mechanism of action. Both subunits bearing a reactive handle are sequentially administered, and undergo the bioorthogonal reaction intracellularly, forming the covalently linked active CLIPTAC that recruits both the target protein and the E3 ligase. The target protein undergoes polyubiquitination and is eventually degraded by the proteasome.

4.3.2. Results:

We envisaged we could adapt the PROTACs synthesis developed in our laboratory to switch from a copper-catalyzed click chemistry reaction into a copper-free reaction, aiming at generating a series of CLIPTACs. Our objective was to turn the terminal alkyne present in the p38α recruiter into a tetrazine suitable to undergo iEDDA bioorthogonal cycloadditions by the Sonogashira coupling. Importantly, not only a new series of interesting p38α degraders would be generated, but it would also showcase the ability of the Sonogashira reaction between **1** and terminal alkynes to

turn a functional group not suitable to undergo bioorthogonal reactions (terminal alkynes) into one able to do so (1,2,4,5-tetrazines).

Therefore, a series of Sonogashira products from the terminal alkyne on the p38α recruiter part were synthesized, varying on the linker (L_1) length (**Scheme 4-11, top**). According to previous results,[26] we knew linkers with 15 and 20 atoms of separation between both functional molecules were desirable to achieve significant p38α degradation. Therefore, we synthesized tetrazines with ethyl (**20a**), propyl (**20b**), butyl (**20c**) and ethyl methyl ether (**20d**) chains as linkers with moderate yields. Alternatively, the thalidomide analogue was functionalized with a norbornene derivative through an amide coupling (**Scheme 4-11, bottom**). In this case, two different linkers lengths with distinct chemistries were employed. An aliphatic chain (butyl, **21a**) and another one with 2 polyethylene glycol (PEG) units (**21b**). Norbornene was selected as the strained alkene due to availability of the starting material. Furthermore, since the Sonogashira products presented faster kinetics in the iEDDA cycloaddition than the S_NAr products used in **Chapter 3**, the use of fast cycloaddends (like the previously employed **BCN** derivatives) was not necessary to achieve significant reactions *in vivo*.



Scheme 4-11: Synthesis of each of the recruiting domains bearing bioorthogonal handles. Precursors were obtained following experimental procedures detailed in Dr. Craig Donoghue's doctoral thesis. Different linker lengths and functionalities were synthesized to test an array of combinations.

Then, the compounds were tested in the NCI-H358 cell line, mimicking the incubation conditions typically employed for CLIPTACs.[28] The experiment was performed in collaboration with Dr. Mónica Cubillos, from the Signaling and Cell Cycle Laboratory of the IRB Barcelona. Cells were grown overnight, and then they were incubated with thalidomide norbornenyl compounds (**21a** or **21b**) at 10 μM concentration for 18 hours. After media removal, the corresponding tetrazinyl-p38α ligands (**20a-d**) were added at 10 μM concentration and left for 18 hours. This period of time was considered sufficient to allow the two reactive counterparts to form the CLIPTAC *in situ* and trigger p38α degradation. Unfortunately, western blot results did not show degradation of the target protein compared to control conditions (**Fig. 4-3**).

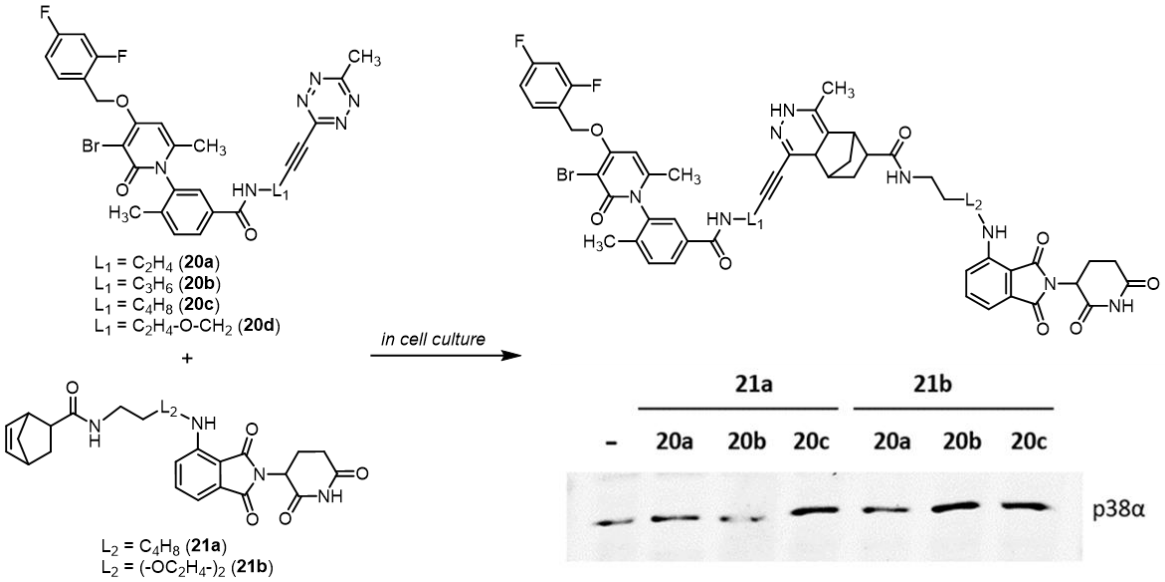


Figure 4-3: Cellular assay of CLIPTACs designed against p38α. Sequential administration of tetrazinyl-p38α ligands (**20a-d**) and norbornenyl-thalidomide analogues (**21a,b**) in NCI-H358 cell culture to generate the CLIPTACs *in situ*. Western blot results show no significant differences between any combination of ligands and control. *For simplicity reasons, only one of the possible isomers of the final CLIPTAC is represented. **Compound **20d** lead to cell death, and the corresponding western blot could not be obtained.

There are several potential reasons for the unsuccessful degradation of p38α by our CLIPTACs. For instance, inefficient cell permeability or stability of the compounds can always be an issue in cellular experiments. Alternatively, the bioorthogonal reactivity of the tetrazine-norbornene pair might be too slow to take place in a quantitative level in the used incubation times. Also, the use of norbornene leads to the formation of different pyridazine isomers in the final CLIPTACs, which can have variable activities. Or it could be simply that the generated CLIPTACs are not active in recruiting the

ternary complex to trigger ubiquitination. In an attempt to overcome some of these potential hurdles, new approaches involving different incubation conditions, thalidomide analogues bearing different strained dienophiles, or tetrazinyl-p38α ligands with a reduced triple bond, will be explored in the future.

4.4. Summary:

3-Bromotetrazines **1** and **3** are suitable reagents to undergo Sonogashira-type cross-coupling reactions directly at position 3 of the tetrazine ring. The corresponding alkynyl-tetrazines (**12a-p**) are underrepresented compounds in the literature that can be exploited to develop new interesting applications. However, we particularly used them as intermediates for the synthesis of alkyl-tetrazines, a highly challenging class of compounds, via a hydrogenation and re-oxidation protocol (**14b,f**). Using this methodology, different unnatural amino acids bearing the alkynyl- and alkyl-tetrazine moiety (**17a,b** and **19a,b**) have been prepared.

To showcase a potential therapeutic application, compounds suitable to function as CLIPTACs (**20a-d**) have been synthesized following previous work on the synthesis of PROTACs developed in the research group. Despite the lack of desired activity, further development of more stable and reactive analogues is underway, aiming at finding a suitable combination of tetrazine-dienophile that could be used to improve the pharmacokinetic behavior of these type of compounds.

4.5. Bibliography

1 B. L. Oliveira, Z. Guo and G. J. L. Bernardes. Inverse electron demand Diels–Alder reactions in chemical biology. *Chem. Soc. Rev.* 2017, **46**, 4895–4950.

2 S. Mayer and K. Lang. Tetrazines in Inverse-Electron-Demand Diels–Alder Cycloadditions and Their Use in Biology. *Synthesis* 2016, **49**, 830–848.

3 K. Sonogashira, Y. Tohda and N. Hagihara. A convenient synthesis of acetylenes: catalytic substitutions of acetylenic hydrogen with bromoalkenes, iodoarenes and bromopyridines. *Tetrahedron Lett.* 1975, **16**, 4467–4470.

4 L. Cassar. Synthesis of aryl- and vinyl-substituted acetylene derivatives by the use of nickel and palladium complexes. *J. Organomet. Chem.* 1975, **93**, 253–257.

5 H. A. Dieck and F. R. Heck. Palladium catalyzed synthesis of aryl, heterocyclic and vinylic acetylene derivatives. *J. Organomet. Chem.* 1975, **93**, 259–263.

6 Z. Novák and A. Kotschy. First Cross-Coupling Reactions on Tetrazines. *Org. Lett.* 2003, **5**, 3495–3497.

7 S. Baba and E. Negishi. A novel stereospecific alkenyl-alkenyl cross-coupling by a palladium- or nickel-catalyzed reaction of alkenylalanes with alkenyl halides. *J. Am. Chem. Soc.* 1976, **98**, 6729–6731.

8 A. O. King, N. Okukado and E. Negishi. Highly general stereo-, regio-, and chemo-selective synthesis of terminal and internal conjugated enynes by the Pd-catalysed reaction of alkynylzinc reagents with alkenyl halides. *J. Chem. Soc. Chem. Commun.* 1977, 683.

9 M. Kosugi, Y. Shimizu and T. Migita. Reaction of allyltin compounds. *J. Organomet. Chem.* 1977, **129**, C36–C38.

10 D. Milstein and J. K. Stille. A general, selective, and facile method for ketone synthesis from acid chlorides and organotin compounds catalyzed by palladium. *J. Am. Chem. Soc.* 1978, **100**, 3636–3638.

11 F. Pop, J. Ding, L. M. L. Daku, A. Hauser and N. Avarvari. Tetrathiafulvalene-s-tetrazine: Versatile platform for donor-acceptor systems and multifunctional ligands. *RSC Adv.* 2013, **3**, 3218–3221.

12 N. Miyaura, K. Yamada and A. Suzuki. A new stereospecific cross-coupling by the palladium-catalyzed reaction of 1-alkenylboranes with 1-alkenyl or 1-alkynyl halides. *Tetrahedron Lett.* 1979, **20**, 3437–3440.

13 N. Miyaura and A. Suzuki. Stereoselective synthesis of arylated (E)-alkenes by the reaction of alk-1-enylboranes with aryl halides in the presence of palladium catalyst. *J. Chem. Soc. Chem. Commun.* 1979, 866.

14 N. Leconte, A. Keromnes-Wuillaume, F. Suzenet and G. Guillaumet. Efficient palladium-catalyzed synthesis of unsymmetrical (het)aryltetrazines. *Synlett* 2007, 204–210.

15 A. Wieczorek, P. Werther, J. Euchner and R. Wombacher. Green- to far-red-emitting fluorogenic tetrazine probes – synthetic access and no-wash protein imaging inside living cells. *Chem. Sci.* 2017, **8**, 1506–1510.

16 P. Fitton and E. A. Rick. The addition of aryl halides to tetrakis(triphenylphosphine)palladium(0). *J. Organomet. Chem.* 1971, **28**, 287–291.

17 A. C. Knall and C. Slugovc. Inverse electron demand Diels-Alder (iEDDA)-initiated conjugation: A (high) potential click chemistry scheme. *Chem. Soc. Rev.* 2013, **42**, 5131–5142.

18 J. Yang, M. R. Karver, W. Li, S. Sahu and N. K. Devaraj. Metal-Catalyzed One-Pot

Synthesis of Tetrazines Directly from Aliphatic Nitriles and Hydrazine. *Angew. Chem., Int. Ed.* 2012, **51**, 5222–5225.

19 H. Wu and N. K. Devaraj. Advances in Tetrazine Bioorthogonal Chemistry Driven by the Synthesis of Novel Tetrazines and Dienophiles. *Acc. Chem. Res.* 2018, **51**, 1249–1259.

20 R. Stollé. Über die Überführung von Hydrazinabkömmlingen in heterocyklische Verbindungen. *J. für Prakt. Chemie* 1906, **73**, 277–287.

21 Q. Zhou, P. Audebert, G. Clavier, F. Miomandre and J. Tang. New unsymmetrical alkyl-s-tetrazines: original syntheses, fluorescence and electrochemical behaviour. *RSC Adv.* 2014, **4**, 7193.

22 D. S. Liu, A. Tangpeerachaikul, R. Selvaraj, M. T. Taylor, J. M. Fox and A. Y. Ting. Diels–Alder Cycloaddition for Fluorophore Targeting to Specific Proteins inside Living Cells. *J. Am. Chem. Soc.* 2012, **134**, 792–795.

23 G. E. Winter, D. L. Buckley, J. Paulk, J. M. Roberts, A. Souza, S. Dhe-paganon and J. E. Bradner. Phthalimide conjugation as a strategy for in vivo target protein degradation. 2015, **348**, 1376–1382.

24 S. An and L. Fu. Small-molecule PROTACs: An emerging and promising approach for the development of targeted therapy drugs. *EBioMedicine* 2018, **36**, 553–562.

25 K. G. Coleman and C. M. Crews. Proteolysis-Targeting Chimeras: Harnessing the Ubiquitin-Proteasome System to Induce Degradation of Specific Target Proteins. *Annu. Rev. Cancer Biol.* 2018, **2**, 41–58.

26 C. Donoghue, M. Cubillos-Rojas, N. Gutierrez-Prat, C. Sanchez-Zarzalejo, X. Verdaguer, A. Riera and A. R. Nebreda. Optimal linker length for small molecule PROTACs that selectively target p38α and p38β for degradation. *Eur. J. Med. Chem.* 2020, **201**, 112451.

27 S. Bühler and S. A. Laufer. p38 MAPK inhibitors: a patent review (2012 – 2013). *Expert Opin. Ther. Pat.* 2014, **24**, 535–554.

28 H. Lebraud, D. J. Wright, C. N. Johnson and T. D. Heightman. Protein degradation by in-cell self-assembly of proteolysis targeting chimeras. *ACS Cent. Sci.* 2016, **2**, 927–934.

4.6. Experimental section

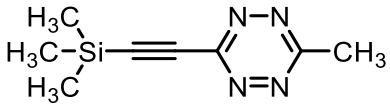
4.6.1. Screening of Sonogashira reaction conditions:

To a degassed reaction vessel with **1** (10.0 mg, 0.057 mmol, 1 eq), PPh_3 (1.5 mg, 0.006 mmol, 0.1 eq), CuI (5.5 mg, 0.006 mmol, 0.1 eq) and the corresponding Pd-catalyst (0.003 mmol, 0.05 eq), the required solvent (0.4 mL, for a final reaction concentration of 0.15 M), TMS-acetylene (15.8 µL, 0.114 mmol, 2 eq) and diisopropylamine (16.0 µL, 0.114 mmol, 2 eq) were added.

The reaction mixture was stirred overnight at the corresponding temperature, after which the suspension was filtered through Celite, the filtrate concentrated under reduced pressure, resuspended in $CDCl_3$ and mesitylene (7.9 µL, 0.057 mmol, 1 eq) added.

^{1}H NMR yield was determined through the ratio of the integration between mesitylene peak at 6.79 ppm (s, 3H) and the product 3-methyl-6-((trimethylsilyl)ethynyl)-1,2,4,5-tetrazine (**12a**) methyl peak at 3.07 ppm (s, 3H). Compound **12a** was also purified and characterized following reaction conditions specified in **section 4.5.2.**

3-methyl-6-((trimethylsilyl)ethynyl)-1,2,4,5-tetrazine (12a)



Pink oil (39.6 mg, 72%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 3.07 (s, 3H), 0.34 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.66, 156.49, 106.10, 96.93, 21.65, -0.60 ppm

HRMS (ESI$^+$) calculated for $C_{13}H_{13}N_4Si$ $[M+H]^+$: 193.0906 / found 193.0904

IR U_{max}: 2960, 2900, 2066, 1390, 1349, 1274, 1250, 1076 cm^{-1}

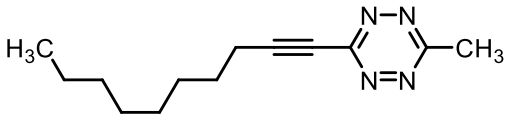
4.6.2. Substrate scope of Sonogashira coupling between terminal alkynes and **1** or **3**:

To a degassed reaction vessel with **1** (50.0 mg, 0.286 mmol, 1 eq) or **3** (67.8 mg, 0.286 mmol, 1 eq), the corresponding alkyne (0.572 mmol, 2 eq), $PdCl_2(PPh_3)_2$ (10.1 mg,

0.014 mmol, 0.05 eq), PPh_3 (7.5 mg, 0.029 mmol, 0.1 eq) and CuI (5.5 mg, 0.029 mmol, 0.1 eq), toluene (2 mL, for a final reaction concentration of 0.15 M) and diisopropylamine (80.2 µL, 0.572 mmol, 2 eq) were added. The reaction mixture was stirred overnight at room temperature.

The reaction crude was evaporated under reduced pressure, and the resulting residue extracted from water with ethyl acetate (3x). The organic phases were combined, dried with $MgSO_4$, filtered and concentrated under vacuum. The desired products were then purified with flash chromatography in 12 g of silica, using hexane with increasing concentrations of ethyl acetate.

3-(dec-1-yn-1-yl)-6-methyl-1,2,4,5-tetrazine (12b)



Pink oil (43.5 mg, 65%).

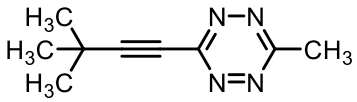
^{1}H NMR (400 MHz, $CDCl_3$) δ 3.05 (s, 3H), 2.58 (t, *J* = 7.2 Hz, 2H), 1.75 – 1.66 (m, 2H), 1.57 – 1.40 (m, 2H), 1.38 – 1.19 (m, 8H), 0.92 – 0.81 (m, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.57, 157.06, 101.44, 75.26, 32.02, 29.33, 29.26, 29.18, 27.98, 22.86, 21.63, 19.90, 14.30 ppm

HRMS (ESI$^+$) calculated for $C_{13}H_{21}N_4$ $[M+H]^+$: 233.176 / found 233.1761

IR U_{max}: 2924, 2854, 2235, 1458, 1402, 1362, 1273, 1078, 1035 cm^{-1}

3-(3,3-dimethylbut-1-yn-1-yl)-6-methyl-1,2,4,5-tetrazine (12c)



Pink oil (49.8 mg, 99%)

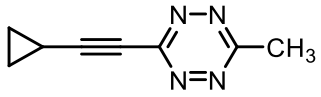
^{1}H NMR (400 MHz, $CDCl_3$) δ 3.05 (s, 3H), 1.42 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.46, 157.01, 108.39, 73.95, 30.28, 28.58, 21.53 ppm

HRMS (ESI$^+$) calculated for $C_9H_{13}N_4$ $[M+H]^+$: 177.1135 / found 177.1133

IR U_{max}: 2935, 2830, 2230, 1443, 1405, 1355, 1050, cm^{-1}

3-(cyclopropylethynyl)-6-methyl-1,2,4,5-tetrazine (12d)



Pink oil (25.2 mg, 55%).

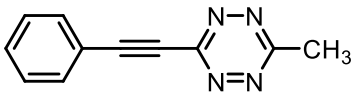
^{1}H NMR (400 MHz, $CDCl_3$) δ 3.03 (s, 3H), 1.68 – 1.57 (m, 1H), 1.12 – 1.01 (m, 4H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.23, 156.94, 104.60, 70.42, 21.51, 9.70, 0.57 ppm

HRMS (ESI$^+$) calculated for $C_8H_9N_4$ $[M+H]^+$: 161.0818 / found 161.0822

IR υ_{max}: 3011, 2923, 2234, 1410, 1336, 1274, 1182, 1077, 1058, 1033 cm^{-1}

3-methyl-6-(phenylethynyl)-1,2,4,5-tetrazine (12e)



Pink solid (27.5 mg, 49%). **Mp =** 119 °C

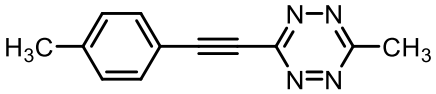
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.73 (dd, *J* = 8.3, 1.5 Hz, 2H), 7.57 – 7.39 (m, 3H), 3.10 (s, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.45, 157.34, 132.89, 132.84, 128.83, 120.38, 97.92, 83.01, 21.62 ppm

HRMS (ESI$^+$) calculated for $C_{13}H_{13}N_4Si$ $[M+H]^+$: 193.0906 / found 193.0904

IR υ_{max}: 2960, 2900, 2066, 1390, 1349, 1274, 1250, 1076 cm^{-1}

3-methyl-6-(*p*-tolylethynyl)-1,2,4,5-tetrazine (12f)



Pink solid (49.3 mg, 82%). **Mp =** 157 °C

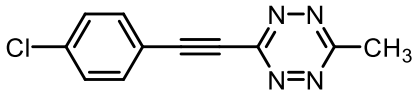
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.61 (d, *J* = 8.2 Hz, 2H), 7.25 – 7.20 (m, 2H), 3.08 (s, 3H), 2.40 (s, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.26, 157.36, 141.52, 132.82, 129.58, 117.27, 98.48, 82.68, 21.88, 21.57 ppm

HRMS (ESI$^+$) calculated for $C_{12}H_{11}N_4$ $[M+H]^+$: 211.0978 / found 211.0977

IR υ_{max}: 3040, 2953, 2923, 2850, 2213, 1509, 1404, 1365, 1274, 1158, 1036 cm^{-1}

3-((4-chlorophenyl)ethynyl)-6-methyl-1,2,4,5-tetrazine (12g)



Pink solid (41.5 mg, 63%). **Mp =** 166 °C

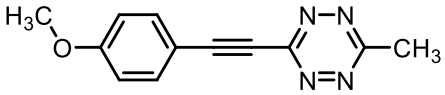
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.69 – 7.60 (m, 2H), 7.46 – 7.38 (m, 2H), 3.09 (s, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.55, 157.20, 137.26, 134.04, 129.30, 118.81, 96.49, 83.81, 21.64 ppm

HRMS (ESI$^+$) calculated for $C_{11}H_8ClN_4$ $[M+H]^+$: 231.0432 / found 231.0433

IR U_{max}: 2233, 2205, 1701, 1482, 1402, 1365, 1086 cm^{-1}

3-((4-methoxyphenyl)ethynyl)-6-methyl-1,2,4,5-tetrazine (12h)



Orange solid (23.3 mg, 36%). **Mp =** 163 °C

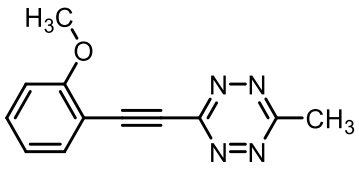
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.66 (dd, *J* = 9.0, 0.8 Hz, 2H), 6.93 (dd, *J* = 9.0, 0.9 Hz, 2H), 3.85 (s, 3H), 3.07 (s, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.12, 161.69, 157.41, 134.71, 114.53, 112.23, 98.80, 82.46, 55.56, 21.55 ppm

HRMS (ESI$^+$) calculated for $C_{12}H_{11}N_4O$ $[M+H]^+$: 227.0927 / found 227.0929

IR U_{max}: 3098, 2965, 2834, 2204, 1600, 1418, 1403, 1366, 1252, 1197, 1154 cm^{-1}

3-((2-methoxyphenyl)ethynyl)-6-methyl-1,2,4,5-tetrazine (12i)



Pink solid (40.1 mg, 62%). **Mp =** 134 °C

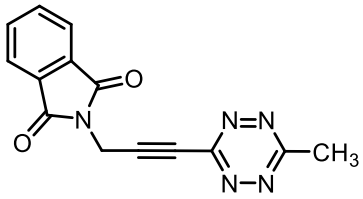
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.66 (dd, *J* = 7.6, 1.7 Hz, 1H), 7.45 (ddd, *J* = 8.5, 7.5, 1.8 Hz, 1H), 7.06 – 6.91 (m, 2H), 3.96 (s, 3H), 3.09 (s, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.21, 161.47, 157.46, 134.71, 132.55, 120.75, 111.03, 109.66, 95.16, 86.84, 56.04, 21.58 ppm

HRMS (ESI$^+$) calculated for $C_{12}H_{11}N_4O$ $[M+H]^+$: 227.0927 / found 227.0928

IR U_{max}: 3406, 2971, 2213, 1714, 1694, 1519, 1494, 1455, 1392, 1274, 1248, 1167, 1075, 1044 cm^{-1}

2-(3-(6-methyl-1,2,4,5-tetrazin-3-yl)prop-2-yn-1-yl)isoindoline-1,3-dione (12j)



Pink oil (44.7 mg, 56%).

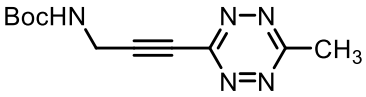
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.91 (dd, *J* = 5.5, 3.1 Hz, 2H), 7.77 (dd, *J* = 5.5, 3.1 Hz, 2H), 4.84 (s, 2H), 3.06 (s, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 166.86, 166.02, 156.49, 134.60, 132.06, 123.94, 91.77, 27.77, 21.65 ppm

HRMS (ESI$^+$) calculated for $C_{14}H_9N_5O_2$ $[M+H]^+$: 280.0830 / found 280.0829

IR U_{max}: 3203, 3078, 2975, 2924, 2216, 1753, 1598, 1475, 1387, 1288, 1053 cm^{-1}

Tert-butyl (6-(6-methyl-1,2,4,5-tetrazin-3-yl)hex-5-yn-1-yl) carbamate (12k)



Pink oil (49.2 mg, 69%).

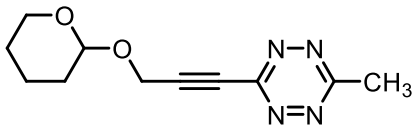
^{1}H NMR (400 MHz, $CDCl_3$) δ 4.93 (s, 1H), 4.33 (d, *J* = 5.7 Hz, 2H), 3.08 (s, 3H), 1.47 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.95, 156.62, 155.26, 95.10, 80.74, 77.39, 31.24, 28.46, 21.63 ppm

HRMS (ESI$^+$) calculated for $C_{11}H_{16}N_5O_2$ $[M+H]^+$: 250.12985 / found 250.13180

IR U_{max}: 3338, 2978, 2932, 2255, 1698, 1519, 1403, 1366, 1250, 1167, 1048 cm^{-1}

3-methyl-6-(3-((tetrahydro-2H-pyran-2-yl)oxy)prop-1-yn-1-yl)-1,2,4,5-tetrazine (12l)



Pink oil (45.5 mg, 68%).

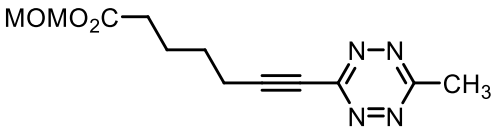
^{1}H NMR (400 MHz, $CDCl_3$) δ 4.94 (t, *J* = 3.1 Hz, 1H), 4.63 (s, 2H), 3.89 (ddd, *J* = 11.1, 9.3, 3.1 Hz, 1H), 3.63 – 3.56 (m, 1H), 3.08 (s, 3H), 1.89 – 1.61 (m, 6H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.94, 156.68, 97.51, 94.92, 79.43, 62.20, 54.41, 30.26, 25.42, 21.65, 18.97 ppm

HRMS (ESI$^+$) calculated for $C_{11}H_{14}N_4O_2$ $[M+H]^+$: 235.11908 / found 235.11895

IR U_{max}: 3326, 2943, 2831, 1448 cm^{-1}

Methoxymethyl 7-(6-methyl-1,2,4,5-tetrazin-3-yl)hept-6-ynoate (12m)



Pink oil (52.8 mg, 70%).

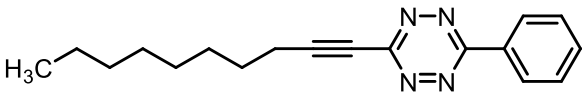
^{1}H NMR (400 MHz, $CDCl_3$) δ 5.24 (s, 2H), 3.47 (s, 3H), 3.06 (s, 3H), 2.64 (t, J = 6.9 Hz, 2H), 2.44 (t, J = 7.2 Hz, 2H), 1.97 – 1.68 (m, 4H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 172.84, 165.59, 156.89, 100.17, 90.53, 75.56, 57.81, 33.80, 27.24, 24.13, 21.55, 19.57 ppm

HRMS (ESI$^+$) calculated for $C_{12}H_{17}N_4O_3$ $[M+H]^+$: 265.12952 / found 265.13156

IR U_{max}: 2941, 2237, 1741, 1590, 1405, 1364, 1137, 1093 cm^{-1}

3-(dec-1-yn-1-yl)-6-phenyl-1,2,4,5-tetrazine (12n)



Pink oil (59.8 mg, 71%).

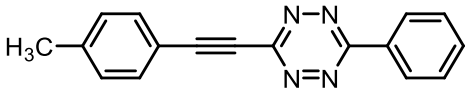
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.60 (dd, J = 8.2, 1.6 Hz, 2H), 7.66 – 7.54 (m, 3H), 2.62 (t, J = 7.1 Hz, 2H), 1.74 (p, J = 7.2 Hz, 2H), 1.56 – 1.46 (m, 2H), 1.37 – 1.23 (m, 8H), 0.88 (t, J = 7.0 Hz, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 161.74, 156.89, 133.11, 131.61, 129.47, 128.49, 102.03, 75.65, 31.96, 29.27, 29.20, 29.13, 27.94, 22.79, 19.95, 14.23 ppm

HRMS (ESI$^+$) calculated for $C_{18}H_{23}N_4$ $[M+H]^+$: 295.1917 / found 295.1914

IR υ_{max}: 2923, 2853, 2233, 1727, 1600, 1463, 1391, 1087 cm^{-1}

3-phenyl-6-(p-tolylethynyl)-1,2,4,5-tetrazine (12o)



Pink oil (73.2 mg, 94%).

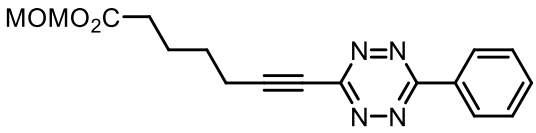
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.64 (dd, *J* = 8.2, 1.5 Hz, 2H), 7.69 – 7.54 (m, 5H), 7.26 (d, *J* = 8.0 Hz, 2H), 2.43 (s, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 161.48, 157.31, 141.60, 133.20, 132.92, 131.60, 131.04, 129.65, 129.51, 128.57, 117.45, 99.22, 83.28, 21.95 ppm

HRMS (ESI$^+$) calculated for $C_{17}H_{13}N_4$ $[M+H]^+$: 273.1135 / found 273.1131

IR υ_{max}: 3066, 2952, 2920, 2213, 1723, 1598, 1509, 1395, 1166, 1076 cm^{-1}

Methoxymethyl 7-(6-phenyl-1,2,4,5-tetrazin-3-yl)hept-6-ynoate (12p)



Pink oil (70.0 mg, 75%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.67 – 8.55 (m, 2H), 7.70 – 7.49 (m, 3H), 5.24 (d, *J* = 0.5 Hz, 2H), 3.47 (d, *J* = 0.5 Hz, 3H), 2.67 (t, *J* = 6.9 Hz, 2H), 2.46 (t, *J* = 7.2 Hz, 2H), 1.98 – 1.73 (m, 4H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 172.83, 161.76, 156.77, 133.15, 131.53, 129.46, 128.50, 100.75, 90.51, 76.02, 57.80, 33.80, 27.26, 24.14, 19.67 ppm

HRMS (ESI$^+$) calculated for $C_{17}H_{19}N_4O_3$ $[M+H]^+$: 327.1452 / found 327.1452

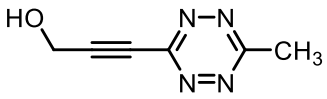
IR υ_{max}: 2938, 2360, 2233, 1735, 1600, 1391, 1133, 1089 cm^{-1}

4.6.3. Deprotection of protecting groups:

The corresponding amount of protected alkynyl-tetrazine (0.100 mmol) was placed in a glass vial, which was sealed, degassed and placed under N_2 atmosphere. DCM was then added (to reach a final concentration of 0.1 M), and HCl 4 M in dioxane (50 µL, 2 eq) was added. The reaction mixture was stirred at room temperature for 4 h, after which the reaction mixture was concentrated under reduced pressure, diluted in water and extracted with ethyl acetate (3x). The combined organic phases were washed once with brine, dried with $MgSO_4$ and concentrated under high vacuum to afford the corresponding products.

* Deprotection of **12k** was verified, but compound **13a** could not be isolated.

3-(6-methyl-1,2,4,5-tetrazin-3-yl)prop-2-yn-1-ol (13b)



Pink oil (11.0 mg, 73%).

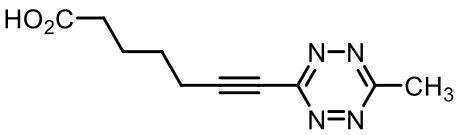
^{1}H NMR (400 MHz, $CDCl_3$) δ 4.66 (s, 2H), 3.09 (s, 3H), 2.00 (s, 1H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 166.08, 156.61, 96.17, 79.36, 51.50, 21.66 ppm

HRMS (ESI$^+$) calculated for $C_6H_6N_4O$ $[M+H]^+$: 151.06101 / found 151.06144

IR U_{max}: 3189 (broad), 2991, 2926, 2843, 2156, 1497, 1302, 1178 cm^{-1}

7-(6-methyl-1,2,4,5-tetrazin-3-yl)hept-6-ynoic acid (13c)



Pink oil (21.8 mg, 99%). **Mp** = 102.4 °C

^{1}H NMR (400 MHz, $CDCl_3$) δ 3.06 (s, 3H), 2.64 (t, *J* = 6.8 Hz, 2H), 2.45 (t, *J* = 7.1 Hz, 2H), 1.92 – 1.69 (m, 4H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 178.62, 165.60, 156.88, 100.16, 75.56, 33.40, 27.18, 24.02, 21.56, 19.56 ppm

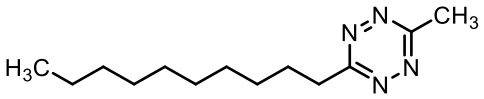
HRMS (ESI$^+$) calculated for $C_{10}H_{13}N_4O_2$ $[M+H]^+$: 221.10330 / found 221.10503

IR U_{max}: 2926, 2860, 2231, 1703, 1406, 1263 cm^{-1}

4.6.4. Tandem hydrogenation + oxidation of substrates:

Glass vials equipped with a PTFE-coated stirring bar were charged with the corresponding substrate (0.187 mmol, 1.0 eq.). Commercially available 10% palladium on carbon, 50% wet (Fluorochem) was added (2 mol % final Pd concentration), and MeOH (1 mL/0.1 mmol substrate) was added as the reaction solvent. Each reaction vial was placed on a reactor, which was purged with N_2, charged with 1 bar(g) of H_2 and left stirring overnight at room temperature. Afterwards, the reaction mixture was filtrated through a short plug of Celite and concentrated under reduced pressure to afford the corresponding crude mixture, which was dissolved in DCM. $PhI(OAc)_2$ (90.4 mg, 0.281 mmol, 1.5 eq.) was added, and the reaction mixture was stirred for 4 h at room temperature until TLC showed completion. The reaction mixture was washed with water (3x), the organic phase was dried with $MgSO_4$ and evaporated under vacuum. The desired product was then purified with flash chromatography in 12 g of silica, using hexane with increasing concentrations of ethyl acetate as the mobile phase.

3-decyl-6-methyl-1,2,4,5-tetrazine (14b)



Pink oil (44.1 mg, >99%).

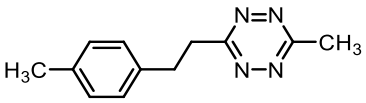
^{1}H NMR (400 MHz, $CDCl_3$) δ 3.27 (t, *J* = 91.1 Hz, 2H), 3.03 (s, 3H), 1.92 (tt, *J* = 7.8, 6.5 Hz, 2H), 1.47 – 1.19 (m, 14H), 0.95 – 0.82 (m, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 170.39, 167.47, 34.88, 32.02, 29.68, 29.58, 29.43, 29.39, 29.29, 28.50, 22.81, 21.22, 14.24 ppm

HRMS (ESI$^+$) calculated for $C_{13}H_{25}N_4$ $[M+H]^+$: 237.20737 / found 237.20925

IR U_{max}: 2957, 2926, 2855, 1730, 1464, 1405, 1284, 1125, 1074 cm^{-1}

3-methyl-6-(4-methylphenethyl)-1,2,4,5-tetrazine (14f)



Pink solid (40.0 mg, >99%). **Mp** = 84 °C

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.11 (q, J = 8.1 Hz, 4H), 3.68 – 3.50 (m, 2H), 3.29 – 3.19 (m, 2H), 3.03 (s, 3H), 2.31 (s, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 169.41, 167.54, 136.93, 136.14, 129.42, 128.42, 36.60, 33.73, 21.23, 21.14 ppm

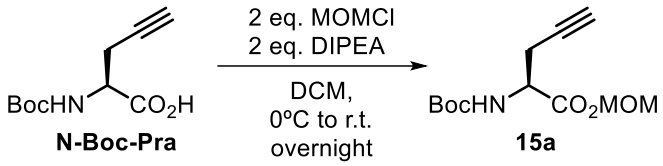
HRMS (ESI$^+$) calculated for $C_{12}H_{15}N_4$ $[M+H]^+$: 215.12912 / found 215.13082

IR U_{max}: 2925, 2848, 1593, 1516, 1431, 1408, 1348, 1045 cm^{-1}

4.6.5. Synthesis of unnatural amino acids

4.6.5.1. Preparation of amino acid precursors:

Synthesis of methoxymethyl (*S*)-2-((*tert*-butoxycarbonyl)amino) pent-4-ynoate (15a)



Commercially available **N-Boc-Pra** (Merck; 400 mg, 1.876 mmol, 1 eq) was suspended in DCM (10 mL) at 0°C, and DIPEA (654 µL, 3.752 mmol, 2 eq) was added, turning the suspension a clear solution. Chloromethyl methyl ether (285 µL, 3.752 mmol, 2 eq) was added dropwise, generating a thick white gas that was totally removed with N_2 flow. The reaction mixture was allowed to warm at room temperature and stirred overnight. The reaction mixture was then diluted with water, extracted with DCM (3x), the combined organics washed once with brine, dried over $MgSO_4$ and evaporated *in vacuo* to afford **15a**, 480 mg, 99% yield, as a colourless oil.

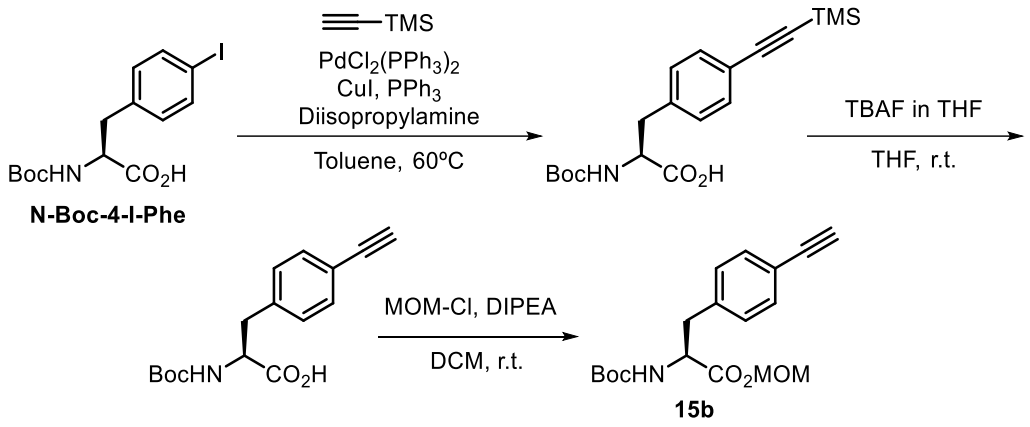
^{1}H NMR (400 MHz, $CDCl_3$) δ 5.32 (dd, *J* = 35.6, 5.9 Hz, 2H), 4.50 (dt, *J* = 8.7, 4.7 Hz, 1H), 3.48 (s, 3H), 2.93 – 2.61 (m, 2H), 2.05 (t, *J* = 2.6 Hz, 1H), 1.45 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 170.49, 155.23, 91.67, 80.43, 78.52, 71.92, 58.00, 52.16, 28.42, 22.91 ppm

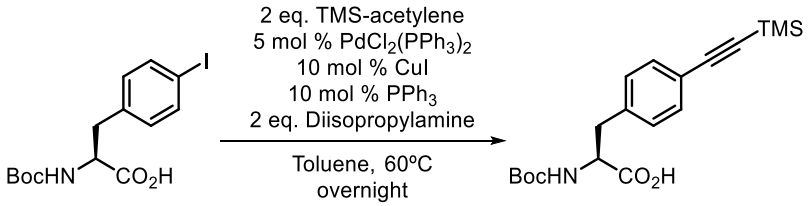
HRMS (ESI$^+$) calculated for $C_{12}H_{20}NO_5$ $[M+H]^+$: 258.13360 / found 258.13570

IR U_{max}: 3375, 3292, 2977, 2934, 1751, 1715, 1505, 1367, 1251, 1162, 1251, 1162, 1097, 1061 cm^{-1}

Synthesis of methoxymethyl (*S*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-ethynylphenyl) propanoate (15b)



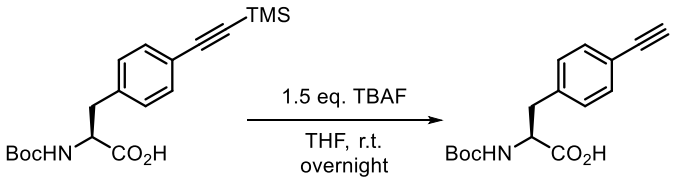
(*S*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((trimethylsilyl)ethynyl)phenyl) propanoic acid:



Commercially available **N-Boc-4-I-Phe** (Fluorochem; 500 mg, 1.278 mmol, 1 eq) was placed in a reaction vial together with $PdCl_2(PPh_3)_2$ (44.8 mg, 0.064 mmol, 0.05 eq), PPh_3 (33.6 mg, 0.128 mmol, 0.1 eq) and CuI (24.4 mg, 0.128 mmol, 0.1 eq). Toluene (8 mL, for a final reaction concentration of 0.16 M), TMS-acetylene (354 µL, 2.556 mmol, 2 eq) and diisopropylamine (358 µL, 2.556 mmol, 2 eq) were added. The reaction mixture was stirred overnight at 60 °C, and then it was filtered through Celite and the filtrate evaporated under reduced pressure. The resulting residue was re-suspended in a HCl 0.1 M aqueous solution, extracted with ethyl acetate (3x), the combined organics washed once with brine, dried over $MgSO_4$ and evaporated to yield a brown oil, from which the desired product (401 mg, 1.109 mmol, 87% yield) was purified as a colorless oil.

^{1}H NMR (400 MHz, DMSO-d_6) δ 7.78 (s, 1H), 7.17 (d, *J* = 7.7 Hz, 2H), 7.10 (d, *J* = 7.5 Hz, 2H), 6.72 (d, *J* = 7.6 Hz, 1H), 4.09 (s, 1H), 2.86 (dd, *J* = 70.6, 12.5 Hz, 2H), 1.42 (s, 9H), -0.01 (s, 9H) ppm

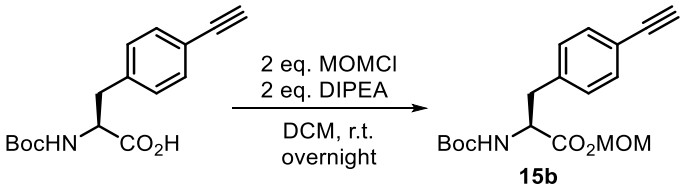
(*S*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-ethynylphenyl)propanoic acid:



The corresponding product from previous reaction (400 mg, 1.106 mmol, 1 eq) was dissolved in THF (6 mL), and tetrabutylammonium fluoride solution 1.0 M in THF (1.66 mL, 1.660 mmol, 1.5 eq) was added dropwise. The reaction mixture was stirred overnight at room temperature, moment in which TLC showed completion. The reaction mixture was concentrated *in vacuo*, the resulting residue re-suspended in HCl 0.1 M aqueous solution and extracted with ethyl acetate (3x). The combined organics were washed once with brine, dried over $MgSO_4$ and evaporated under vacuum to afford the desired product (263.8 mg, 0.912 mmol, 82%) as a brown oil, which was taken onto next step without further purification.

^{1}H NMR (400 MHz, $CDCl_3$) δ 9.62 (s, 1H), 7.41 (d, *J* = 8.1 Hz, 2H), 7.14 (d, *J* = 8.3 Hz, 2H), 5.03 (d, *J* = 9.3 Hz, 1H), 4.60 (d, *J* = 6.9 Hz, 1H), 3.20 (dd, *J* = 14.1, 5.4 Hz, 1H), 3.09 (dd, *J* = 14.1, 5.4 Hz, 1H), 3.06 (s, 1H), 1.41 (s, 9H) ppm

Methoxymethyl (*S*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-ethynylphenyl) propanoate (15b):

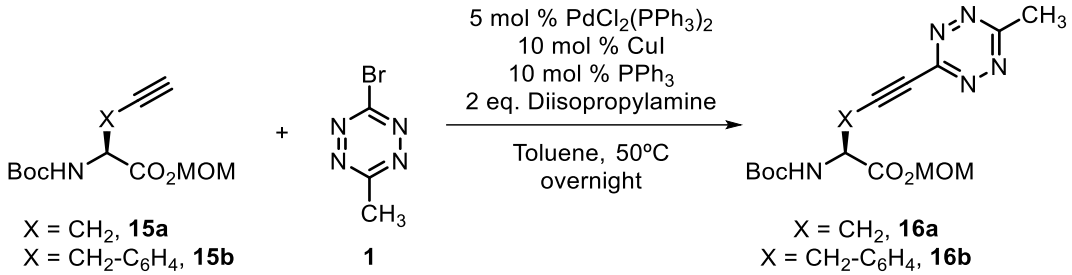


The corresponding N-Boc-4-acetylene-L-phenylalanine (263.8 mg, 0.912 mmol, 1 eq) was suspended in DCM at 0 °C, and DIPEA (318 µL, 1.824 mmol, 2 eq) was added. Chloromethyl methyl ether (139 µL, 1.824 mmol, 2 eq) was added dropwise, generating a thick white gas that was totally removed with N_2 in-flow. The reaction mixture was allowed to warm at room temperature and stirred overnight. The reaction mixture was then diluted with water, extracted with ethyl acetate (3x), the combined organics washed once with brine, dried over $MgSO_4$ and evaporated *in vacuo*. The resulting residue was purified with flash chromatography in 12 g of silica, using hexane with increasing concentrations of ethyl acetate. The relevant fractions

were combined and concentrated to obtain **15b**, 107 mg, 0.320 mmol, 35% yield, as a colorless oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.42 (d, *J* = 8.2 Hz, 2H), 7.13 (d, *J* = 8.1 Hz, 2H), 5.38 – 5.12 (m, 2H), 4.97 (d, *J* = 8.3 Hz, 1H), 4.60 (d, *J* = 7.7 Hz, 1H), 3.41 (s, 3H), 3.16 (dd, *J* = 13.9, 5.9 Hz, 1H), 3.09 (dd, *J* = 13.9, 5.9 Hz, 1H), 3.06 (s, 1H), 1.41 (s, 9H) ppm
^{13}C NMR (101 MHz, $CDCl_3$) δ 171.52, 155.16, 136.98, 132.45, 129.51, 121.05, 91.52, 83.47, 80.28, 77.48, 58.08, 54.41, 38.30, 28.41 ppm

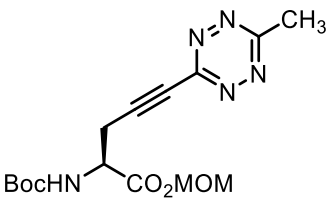
4.6.5.2. Sonogashira coupling between **15a-b** and **1**:



To a degassed reaction vessel with **1** (50.0 mg, 0.286 mmol, 1 eq), the corresponding double protected amino acid **15a** or **15b** (0.572 mmol, 2 eq), $PdCl_2(PPh_3)_2$ (10.1 mg, 0.0143 mmol, 0.05 eq), PPh_3 (7.5 mg, 0.029 mmol, 0.1 eq) and CuI (5.5 mg, 0.029 mmol, 0.1 eq), toluene (2 mL, for a final reaction concentration of 0.15 M) and diisopropylamine (80.2 µL, 0.572 mmol, 2 eq) were added. The reaction mixture was stirred overnight at 50 °C.

The reaction crude was evaporated under reduced pressure, and the resulting residue extracted from water with ethyl acetate (3x). The organic phases were combined, dried with $MgSO_4$, filtered and concentrated under vacuum. The desired tetrazine-containing protected amino acids were then purified with flash chromatography in 12 g of silica, using hexane with increasing concentrations of ethyl acetate.

Methoxymethyl (*S*)-2-((*tert*-butoxycarbonyl)amino)-5-(6-methyl-1,2,4,5-tetrazin-3-yl)pent-4-ynoate (16a)



Pink oil (49.2 mg, 49%).

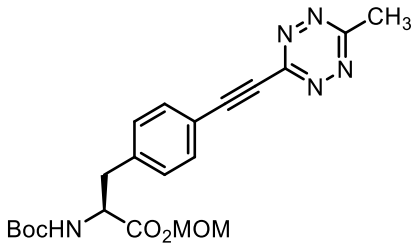
^{1}H NMR (400 MHz, $CDCl_3$) δ 5.50 (d, *J* = 7.8 Hz, 1H), 5.36 (dd, *J* = 21.6, 5.9 Hz, 2H), 4.71 – 4.63 (m, 1H), 3.51 (s, *J* = 1.3 Hz, 3H), 3.21 (d, *J* = 5.8 Hz, 2H), 3.06 (s, 3H), 1.46 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 169.97, 165.81, 156.62, 155.11, 94.41, 92.07, 80.77, 58.22, 52.03, 28.42, 24.19, 21.60 ppm

HRMS (ESI$^+$) calculated for $C_{15}H_{22}N_5O_5$ $[M+H]^+$: 352.1615 / found 352.1624

IR U_{max}: 3339, 2976, 2922, 2359, 2239, 1713, 1614, 1503, 1366, 1259, 1149 cm^{-1}

Methoxymethyl (*S*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((6-methyl-1,2,4,5-tetrazin-3-yl)ethynyl)phenyl)propanoate (16b)



Pink oil (78.2 mg, 64%).

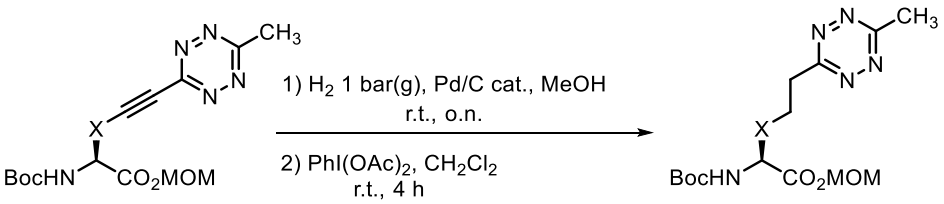
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.67 (d, 2H), 7.26 (d, 2H), 5.28 (dd, 2H), 5.02 (d, *J* = 8.3 Hz, 1H), 4.64 (m, 1H), 3.43 (s, 3H), 3.27 – 3.11 (m, 2H), 3.10 (s, 3H), 1.42 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 171.39, 165.42, 157.32, 155.12, 139.42, 133.05, 129.83, 119.12, 97.70, 91.62, 83.24, 80.39, 58.16, 54.36, 38.58, 28.41, 21.63 ppm

HRMS (ESI$^+$) calculated for $C_{42}H_{50}N_{10}NaO_{10}$ $[2M+Na]^+$: 877.3604 / found 877.3604

IR U_{max}: 3361, 2973, 2929, 2218, 1744, 1704, 1508, 1404, 1365, 1156, 1091, 1053, 1019 cm^{-1}

4.6.5.3. Tandem hydrogenation-oxidation of **16a-b**:

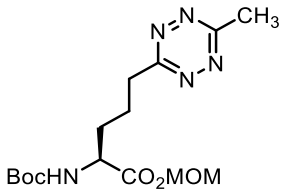


X = CH_2, **16a**
X = CH_2-C_6H_4, **16b**

X = CH_2, **18a**
X = CH_2-C_6H_4, **18b**

The same experimental protocol described in **Experimental Section 4.5.4** was followed. The amount of starting material employed is detailed in each case.

Methoxymethyl (*S*)-2-((*tert*-butoxycarbonyl)amino)-5-(6-methyl-1,2,4,5-tetrazin-3-yl)pentanoate (18a)



Pink oil (36.7 mg, 97% starting from 37.4 mg of **16a**).

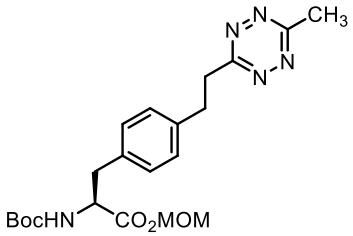
^{1}H NMR (400 MHz, $CDCl_3$) δ 5.28 (dd, *J* = 32.3, 6.0 Hz, 2H), 5.07 (d, *J* = 8.4 Hz, 1H), 4.40 (s, 1H), 3.47 (s, 3H), 3.34 (td, *J* = 7.5, 3.2 Hz, 2H), 3.04 (s, 3H), 2.11 – 1.97 (m, 2H), 1.86 – 1.77 (m, 1H), 1.60 (m, 1H), 1.44 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 172.31, 169.49, 167.70, 155.49, 91.43, 80.26, 58.05, 53.36, 34.30, 32.22, 28.45, 23.98, 21.25 ppm

HRMS (ESI$^+$) calculated for $C_{15}H_{26}N_5O_5$ $[M+H]^+$: 356.1928 / found 356.1932

IR υ_{max}: 3354, 2971, 2929, 2360, 2341, 1709, 1507, 1456, 1405, 1365, 1249, 1148, 1088, 1049 cm^{-1}

Methoxymethyl (*S*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-(2-(6-methyl-1,2,4,5-tetrazin-3-yl)ethyl)phenyl)propanoate (18b)



Pink oil (21.3 mg, 95% starting from 22.3 mg of **16b**).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.22 – 7.15 (m, 2H), 7.07 (dd, *J* = 17.2, 7.8 Hz, 2H), 5.30 (d, *J* = 5.9 Hz, 1H), 5.22 (d, *J* = 6.0 Hz, 1H), 4.95 (d, *J* = 8.3 Hz, 1H), 4.58 (m, 1H), 3.70 (s, 1.4H), 3.59 (ddd, *J* = 8.0, 7.0, 1.7 Hz, 2H), 3.41 (s, 1.6H), 3.25 (dd, *J* = 9.1, 6.9 Hz, 2H), 3.14 – 3.04 (m, 2H), 3.03 (s, 3H), 1.41 (s, 9H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 172.46, 171.74, 169.29, 167.62, 155.21, 138.85, 134.25 (d, *J* = 15.9 Hz), 129.75, 129.70, 128.77, 91.43, 80.14, 58.04, 54.53, 52.34, 38.02 (d, *J* = 16.8 Hz), 36.39, 33.69, 28.43, 21.25 ppm

HRMS (ESI$^+$) calculated for $C_{21}H_{30}N_5O_5$ $[M+H]^+$: 432.2241 / found 432.2240

IR U_{max}: 3353, 2973, 2927, 2854, 1743, 1707, 1514, 1404, 1365, 1160, 1093 cm^{-1}

4.6.5.4. Deprotection of amino acids **16a-b** and **18a-b**:

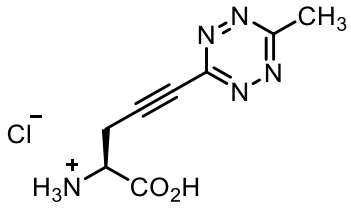
The corresponding amount of double protected amino acid (specified for each substrate) was placed in a glass vial, which was sealed, degassed and placed under N_2 atmosphere. DCM was then added (to reach a final concentration of 0.1 M), and HCl 4 M in dioxane (2 eq) was added. The reaction mixture was stirred at room temperature for 4 h, after which the reaction mixture was concentrated under reduced pressure. Two different separation methods were employed:

Method A: the resulting residue was washed with DCM, and the supernatant decanted.

Method B: the resulting residue resuspended in DCM and filtered through a C-type crucible filter.

Each final amino acid was recovered in its hydrochloride salt form.

(*S*)-1-carboxy-4-(6-methyl-1,2,4,5-tetrazin-3-yl)but-3-yn-1-aminium chloride (17a)



Purple solid (17.9 mg, 92% starting from 28.0 mg of **16a**). Isolated using **Method A**.

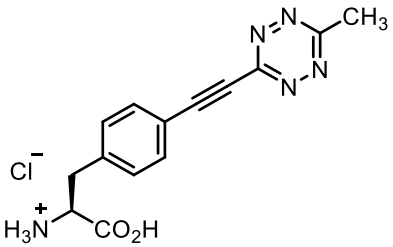
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.33 (s, 1H), 4.39 (dd, *J* = 8.7, 5.2 Hz, 1H), 3.42 (ddd, *J* = 14.9, 5.2, 1.0 Hz, 1H), 3.25 (ddd, *J* = 14.9, 8.7, 0.7 Hz, 1H), 3.03 (s, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 167.70, 165.02, 140.61, 136.47, 124.55, 52.22, 43.37, 21.23 ppm

HRMS (ESI^+) calculated for $C_8H_{11}ClN_5O_2$ $[M+H]^+$: 244.0596 / found 244.0602

IR U_{max}: 3338, 2929, 2241, 1709, 1509, 1402, 1365, 1275, 1251, 1157, 1058 cm^{-1}

(*S*)-1-carboxy-2-(4-((6-methyl-1,2,4,5-tetrazin-3-yl)ethynyl) phenyl)ethan-1-aminium chloride (17b)



Pink solid (15.8 mg, 96% starting from 22.0 mg of **16b**). Isolated using **Method B**.

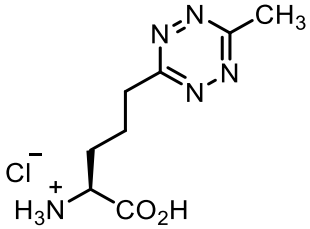
^{1}H NMR (400 MHz, CD_3OD) δ 7.76 (d, *J* = 8.5 Hz, 2H), 7.46 (d, *J* = 8.2 Hz, 2H), 4.33 (m, 1H), 3.40 (dd, *J* = 14.6, 5.8 Hz, 1H), 3.25 (dd, *J* = 14.6, 5.8 Hz, 1H), 3.04 (s, 3H) ppm

^{13}C NMR (101 MHz, CD_3OD) δ 170.99, 167.09, 158.21, 138.98, 134.21, 131.19, 121.06, 96.89, 84.27, 54.78, 37.30, 21.45 ppm

HRMS (ESI$^+$) calculated for $C_{14}H_{14}N_5O_2$ $[M\text{-}Cl]^+$: 284.1142 / found 284.1144

IR U_{max}: 3412, 2920, 2215, 1724, 1510, 1403, 1205, 1159, 1111, 1076, 1019 cm^{-1}

(*S*)-1-carboxy-4-(6-methyl-1,2,4,5-tetrazin-3-yl)butan-1-aminium chloride (19a)



Pink solid (15.6 mg, 89% starting from 25.1 mg of **18a**). Isolated using **Method A**.

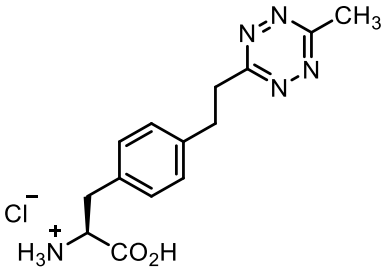
^{1}H NMR (400 MHz, CD_3OD) δ 4.06 (t, *J* = 5.7 Hz, 1H), 3.37 (t, *J* = 6.9 Hz, 2H), 2.99 (s, 3H), 2.23 – 2.04 (m, 4H) ppm

^{13}C NMR (101 MHz, CD_3OD) δ 171.58, 170.28, 169.00, 53.64, 34.76, 30.84, 24.18, 21.03 ppm

HRMS (ESI$^+$) calculated for $C_8H_{14}N_5O_2$ $[M+H]^+$: 212.11420 / found 212.11588

IR U_{max}: 3339, 3155, 2934, 1742, 1598, 1404, 1216, 1216, 1084, 1043 cm^{-1}

(*S*)-1-carboxy-2-(4-(2-(6-methyl-1,2,4,5-tetrazin-3-yl)ethyl) phenyl)ethan-1-aminium chloride (19b)



Pink solid (12.1 mg, 95% starting from 17.0 mg of **18b**). Isolated using **Method B**.

^{1}H NMR (400 MHz, CD_3OD) δ 7.31 – 7.14 (m, 7H), 4.23 (dd, *J* = 7.9, 5.3 Hz, 1H), 3.80 (d, *J* = 0.8 Hz, 1H), 3.58 (t, *J* = 8.6 Hz, 2H), 3.30 – 3.25 (m, 2H), 3.11 (dd, *J* = 14.6, 7.8 Hz, 2H), 2.97 (s, 3H) ppm

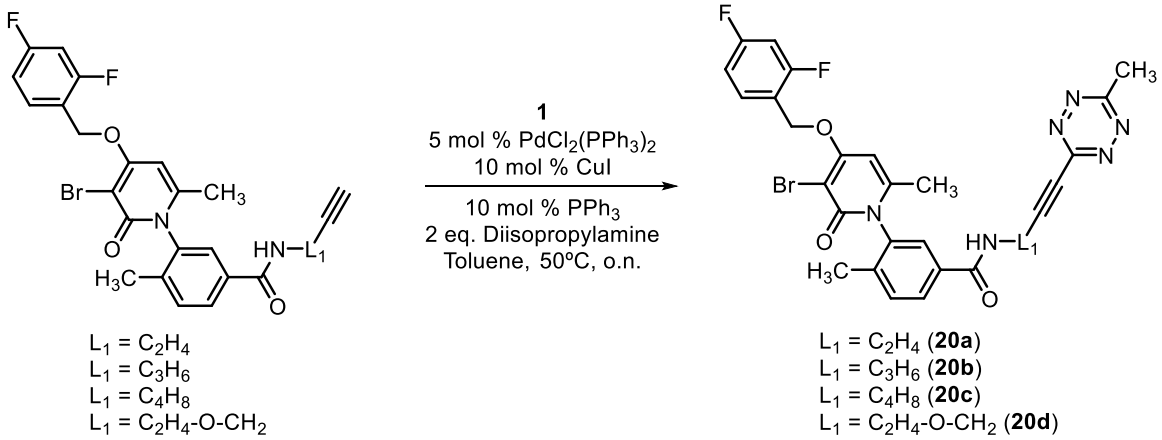
^{13}C NMR (101 MHz, CD_3OD) δ 171.18, 170.33, 168.89, 141.31, 133.62, 130.70, 130.33, 55.09, 37.15, 36.93, 34.35, 21.00 ppm

HRMS (ESI$^+$) calculated for $C_{14}H_{18}N_5O_2$ [M+H]$^+$: 288.14550 / found 288.14781

IR U_{max}: 3412, 2930, 1742, 1726, 1587, 1482, 1407 cm^{-1}

4.6.6. CLIPTACs synthesis

4.6.6.1. p38α recruiter synthesis:

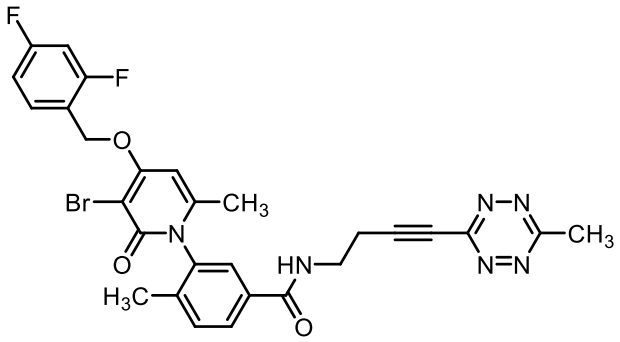


The same experimental procedure as in **Section 4.5.5.2.** was followed. Briefly, terminal alkynes with 4 different linker (L_1) lengths (0.572 mmol, 1.5 eq) were placed in a degassed reaction vessel together with **1** (50.0 mg, 0.286 mmol, 1 eq), $PdCl_2(PPh_3)_2$ (10.1 mg, 0.0143 mmol, 0.05 eq), PPh_3 (7.5 mg, 0.029 mmol, 0.1 eq) and CuI (5.5 mg, 0.029 mmol, 0.1 eq), toluene (2 mL, for a final reaction concentration of

0.15 M) and diisopropylamine (80.2 µL, 0.572 mmol, 2 eq) were added. The reaction mixture was stirred overnight at 50 ºC.

The reaction crude was evaporated under reduced pressure, and the resulting residue extracted from water with ethyl acetate (3x). The organic phases were combined, dried with $MgSO_4$, filtered and concentrated under vacuum. The desired products were then purified with flash chromatography in 12 g of silica, using hexane with increasing concentrations of ethyl acetate.

3-(3-bromo-4-((2,4-difluorobenzyl)oxy)-6-methyl-2-oxopyridin-1(2H)-yl)-4-methyl-N-(4-(6-methyl-1,2,4,5-tetrazin-3-yl)but-3-yn-1-yl)benzamide (20a)



Pink solid (81.9 mg, 47%).

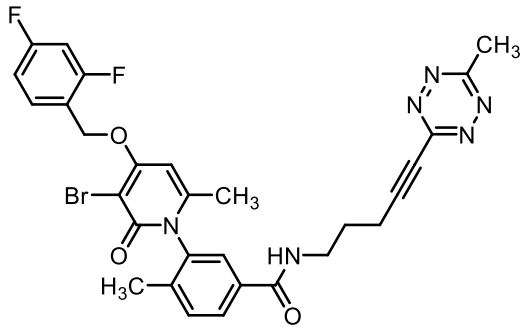
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.86 (dd, *J* = 7.9, 1.8 Hz, 1H), 7.62 – 7.56 (m, 2H), 7.52 – 7.47 (m, 1H), 7.37 (d, *J* = 8.0 Hz, 1H), 6.98 – 6.89 (m, 1H), 6.80 (tdd, *J* = 8.7, 2.5, 0.9 Hz, 1H), 6.14 (s, 1H), 5.17 (d, *J* = 3.6 Hz, 2H), 3.68 – 3.27 (m, 2H), 3.02 (d, *J* = 0.9 Hz, 3H), 2.77 (td, *J* = 7.2, 2.9 Hz, 2H), 2.12 (s, 3H), 1.92 (d, *J* = 0.8 Hz, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.7, 161.2 (d, *J* = 12.1 Hz), 160.5, 156.8, 146.9, 138.7, 137.4, 133.8, 131.4, 129.8 (dd, *J* = 9.7, 5.1 Hz), 129.1, 128.6 (d, *J* = 12.2 Hz), 126.7, 118.6 (dd, *J* = 14.1, 3.8 Hz), 111.9 (dd, *J* = 21.4, 3.6 Hz), 104.0 (t, *J* = 25.2 Hz), 96.7, 64.7, 64.6, 60.5, 38.4, 21.7, 21.5, 20.5, 17.4 ppm

HRMS (ESI$^+$) calculated for $C_{28}H_{24}O_3N_6BrF_2$ $[M+H]^+$: 609.10558 / found 609.10577

IR υ_{max}: 3318, 3052, 2917, 2231, 1642, 1524, 1401, 1348, 1099, 1037 cm^{-1}

3-(3-bromo-4-((2,4-difluorobenzyl)oxy)-6-methyl-2-oxopyridin-1(2H)-yl)-4-methyl-N-(5-(6-methyl-1,2,4,5-tetrazin-3-yl)pent-4-yn-1-yl)benzamide (20b)



Pink solid (58.8 mg, 33%).

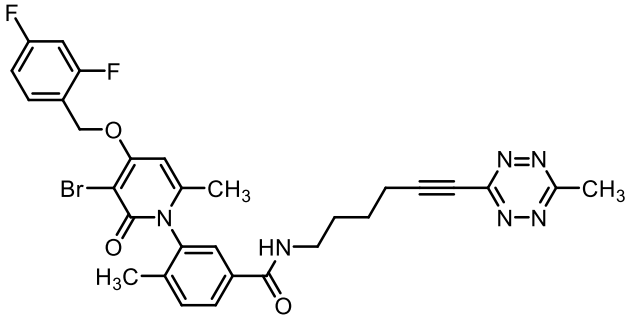
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.87 (dt, *J* = 7.9, 2.0 Hz, 1H), 7.70 – 7.63 (m, 5H), 7.59 (d, *J* = 1.9 Hz, 1H), 7.40 (d, *J* = 7.6 Hz, 1H), 7.23 – 7.17 (m, 1H), 6.99 – 6.93 (m, 1H), 6.82 (dddd, *J* = 10.2, 8.7, 2.5, 1.6 Hz, 1H), 6.15 (dd, *J* = 1.7, 0.9 Hz, 1H), 5.18 (s, 2H), 3.41 (ddq, *J* = 49.0, 13.0, 6.4 Hz, 2H), 3.04 (d, *J* = 1.7 Hz, 3H), 2.64 – 2.57 (m, 2H), 2.09 (d, *J* = 1.6 Hz, 3H), 1.93 (dd, *J* = 1.7, 0.7 Hz, 3H), 1.89 (td, *J* = 7.7, 6.2 Hz, 2H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.95, 165.51, 163.43, 160.56, 156.87, 146.81, 138.60, 137.40, 134.14, 133.16, 132.28, 132.18, 132.09, 131.59, 126.54, 112.03 (d, *J* = 21.1 Hz), 104.02, 100.45, 96.66, 64.62, 60.53, 39.35, 36.78, 27.61, 24.83, 23.53, 21.78, 21.53, 17.58, 17.46 ppm

HRMS (ESI$^+$) calculated for $C_{29}H_{26}O_3N_6BrF_2$ $[M+H]^+$: 623.12123 / found 623.12087

IR υ_{max}: 3318, 3048, 2913, 2231, 1646, 1528, 1401, 1262 cm^{-1}

3-(3-bromo-4-((2,4-difluorobenzyl)oxy)-6-methyl-2-oxopyridin-1(2H)-yl)-4-methyl-N-(6-(6-methyl-1,2,4,5-tetrazin-3-yl)hex-5-yn-1-yl)benzamide (20c)



Pink solid (118.5 mg, 65%).

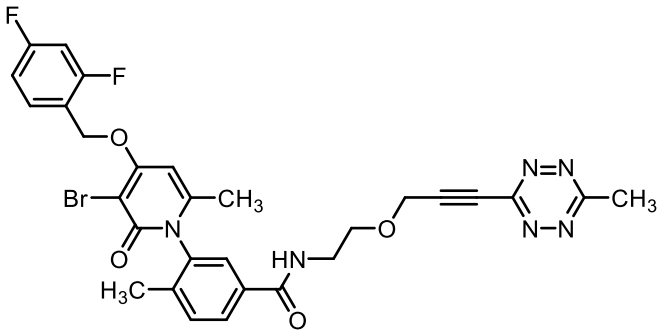
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.82 (dd, *J* = 7.9, 1.8 Hz, 1H), 7.68 – 7.57 (m, 1H), 7.57 (d, *J* = 1.9 Hz, 1H), 7.33 (dd, *J* = 8.0, 0.9 Hz, 1H), 7.13 (t, *J* = 5.7 Hz, 1H), 6.99 – 6.89 (m, 1H), 6.80 (ddd, *J* = 10.3, 8.7, 2.5 Hz, 1H), 6.14 (s, 1H), 5.17 (s, 2H), 3.40 – 3.09 (m, 2H), 3.01 (s, 3H), 2.59 (t, *J* = 6.7 Hz, 2H), 2.07 (s, 3H), 1.91 (s, 3H), 1.77 – 1.60 (m, 4H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 165.58, 165.43, 164.26 (d, *J* = 11.9 Hz), 163.39, 161.78 (d, *J* = 11.9 Hz), 161.16 (d, *J* = 12.0 Hz), 160.47, 158.68 (d, *J* = 12.1 Hz), 156.85, 146.85, 138.26, 137.29, 134.20, 131.38, 129.80 (dd, *J* = 9.9, 5.3 Hz), 128.96, 128.61 (d, *J* = 12.2 Hz), 126.50, 111.94 (dd, *J* = 21.4, 3.6 Hz), 103.95 (t, *J* = 25.2 Hz), 96.63, 64.63, 39.40, 28.56, 25.18, 23.57, 21.62 (d, *J* = 29.8 Hz), 19.43, 17.41 ppm

HRMS (ESI$^+$) calculated for $C_{30}H_{28}O_3N_6BrF_2$ $[M+H]^+$: 637.13688 / found 637.13660

IR U_{max}: 3855, 3318, 3053, 2925, 2234, 2040, 1620, 1560, 1343, 1274, 1036 cm^{-1}

3-(3-bromo-4-((2,4-difluorobenzyl)oxy)-6-methyl-2-oxopyridin-1(2H)-yl)-4-methyl-N-(2-((3-(6-methyl-1,2,4,5-tetrazin-3-yl)prop-2-yn-1-yl)oxy)ethyl)benzamide (20d)



Pink solid (69.5 mg, 38%).

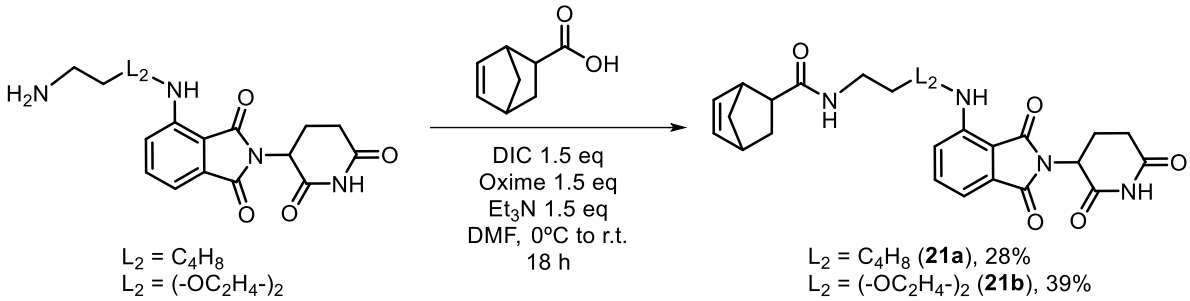
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.76 (dd, *J* = 7.9, 1.9 Hz, 1H), 7.60 (td, *J* = 8.5, 6.3 Hz, 1H), 7.52 (d, *J* = 1.9 Hz, 1H), 7.39 – 7.32 (m, 1H), 7.02 – 6.76 (m, 3H), 6.14 (d, *J* = 0.9 Hz, 1H), 5.23 (d, *J* = 1.3 Hz, 2H), 4.15 (d, *J* = 2.4 Hz, 2H), 3.69 – 3.42 (m, 4H), 2.43 (t, *J* = 2.4 Hz, 1H), 2.08 (s, 3H), 1.95 – 1.88 (m, 3H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 166.23, 165.98, 163.31, 160.62, 156.52, 146.46, 138.91, 137.66, 134.15, 131.57, 130.28, 128.43, 126.82, 118.65, 112.03 (d, *J* = 21.3 Hz), 104.05 (t, *J* = 25.2 Hz), 96.91, 96.40, 94.28, 80.10, 69.41, 64.53, 64.49, 60.52, 58.74, 39.80, 29.82, 21.67 (d, *J* = 8.6 Hz), 17.49, 14.33 ppm

HRMS (ESI$^+$) calculated for $C_{29}H_{26}O_4N_6BrF_2$ $[M+H]^+$: 639.11615 / found 639.11592

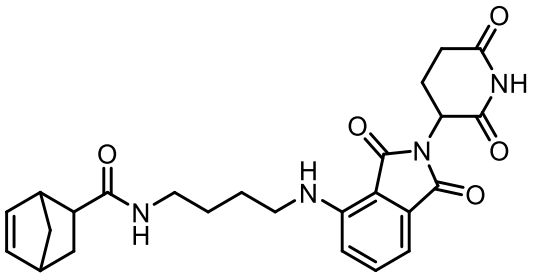
IR U_{max}: 3440 (broad), 2917, 2235, 1638, 1528, 1401, 1348, 1274, 1099, 1037 cm^{-1}

4.6.6.2. Thalidomide-norbornene synthesis:



Commercially available 5-norbornen-2-carboxylic acid, mixture of *endo* and *exo,* predominantly *endo* (Merck) (34 µL, 0.247 mmol, 1 eq) was suspended in DMF (4 mL) and cooled to 0 ºC, and oxime (52.7 mg, 0.371 mmol, 1.5 eq) and DIC (57 µL, 0.371 mmol, 1.5 eq) were then added. After 15 min, the solution turned yellow and Et_3N (66 µL, 0.371 mmol, 1.5 eq) was added, followed by the corresponding thalidomide analogues with different linker (L_2) lengths (0.247 mmol, 1 eq). The reaction was left 18 hours before quenching with water. The mixture was suspended in water and extracted with EtOAc (x3), the combined organic layers washed with brine, dried over $MgSO_4$ and concentrated under reduced pressure. The corresponding crude amide(s) *endo* products were then purified by flash chromatography.

N-(4-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl)amino) butyl)bicyclo[2.2.1] hept-5-ene-2-carboxamide (*endo* isomer) (21a)



Pink solid (32.1 mg, 28%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.03 (s, 1H), 7.49 (dd, *J* = 8.5, 7.1 Hz, 1H), 7.10 (dd, *J* = 7.1, 0.6 Hz, 1H), 6.88 (dd, *J* = 8.6, 0.6 Hz, 1H), 6.23 (dd, *J* = 5.7, 3.1 Hz, 1H), 5.96 (dd, *J* = 5.7, 2.8 Hz, 1H), 5.42 (s, 1H), 4.91 (dd, *J* = 12.1, 5.3 Hz, 1H), 3.33 – 3.22 (m, 4H), 3.12 (s, 1H), 2.91 (d, *J* = 2.7 Hz, 1H), 2.90 – 2.67 (m, 4H), 2.18 – 2.10 (m, 1H), 1.94 (ddd, *J* =

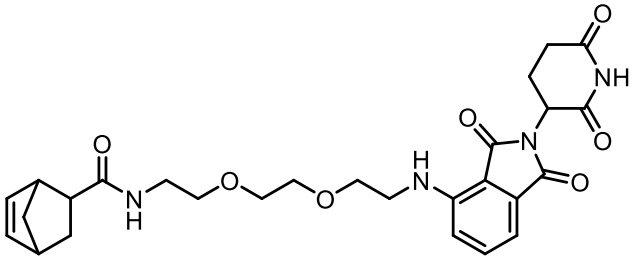
11.8, 9.4, 3.8 Hz, 1H), 1.63 (dddd, *J* = 22.9, 11.5, 7.9, 4.5 Hz, 7H), 1.45 (dq, *J* = 6.4, 2.2 Hz, 1H), 1.34 – 1.28 (m, 2H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 174.52, 171.09, 169.68, 168.47, 167.71, 147.01, 137.99, 136.35, 132.38, 116.82, 111.73, 110.13, 50.19, 49.02, 46.34, 45.01, 42.86, 42.30, 38.98, 31.55, 30.11, 27.42, 26.67, 22.95 ppm

HRMS (ESI$^+$) calculated for $C_{28}H_{24}O_3N_6BrF_2$ $[M+H]^+$: 465.21325 / found 465.21324

IR U_{max}: 3396 (broad), 2921, 2848, 1695, 1626, 1507, 1352, 1258, 1025 cm^{-1}

N-(2-(2-(2-((2-(2,6-dioxopiperidin-3-yl)-1,3-dioxoisoindolin-4-yl) amino)ethoxy)ethoxy) ethyl)bicyclo[2.2.1]hept-5-ene-2-carboxamide (*endo* isomer) (21b)



Pink solid (50.5 mg, 39%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.46 (s, 1H), 7.50 (dd, *J* = 8.5, 7.1 Hz, 1H), 7.11 (dd, *J* = 7.1, 0.7 Hz, 1H), 6.90 (d, *J* = 8.5 Hz, 1H), 6.52 (t, *J* = 5.5 Hz, 1H), 6.20 (dd, *J* = 5.7, 3.1 Hz, 1H), 5.99 – 5.85 (m, 2H), 4.94 – 4.87 (m, 1H), 3.73 (t, *J* = 5.3 Hz, 2H), 3.69 – 3.61 (m, 5H), 3.55 (t, *J* = 5.2 Hz, 2H), 3.50 – 3.45 (m, 2H), 3.40 (td, *J* = 5.8, 4.4 Hz, 2H), 3.10 (s, 1H), 2.91 – 2.70 (m, 5H), 2.17 – 2.09 (m, 1H), 1.89 (dddd, *J* = 11.5, 9.4, 3.8, 1.9 Hz, 1H), 1.65 (s, 2H), 1.42 – 1.29 (m, 2H), 1.26 – 1.22 (m, 1H) ppm

^{13}C NMR (101 MHz, $CDCl_3$) δ 174.59, 171.20, 169.50, 168.54, 167.70, 146.91, 137.82, 136.25, 132.71, 132.43, 116.82, 111.92, 110.57, 70.90, 70.22 (d, *J* = 5.0 Hz), 69.44, 50.07, 49.04, 46.29, 44.94, 42.85, 42.48, 39.34, 31.53, 29.98, 23.03 ppm

HRMS (ESI$^+$) calculated for $C_{27}H_{33}O_7N_4$ $[M+H]^+$: 525.23438 / found 525.23438

IR U_{max}: 3379, 2958, 2913, 2868, 1691, 1622, 1507, 1348, 1115 cm^{-1}

4.6.7. CLIPTACs degradation test assay:

* The experiment was carried out by Dr. Mónica Cubillos (Signalling and Cell Cycle Laboratory of the IRB Barcelona) applying the standardized protocol. Briefly, NCI

H358 cells were plated and incubated overnight, point at which they were treated with compounds **21a** or **21b** at a final concentration of 10 μM. After 18 hours, the media was removed, cells were washed once with PBS and fresh media was added. Compounds **20a-d** were added at a final concentration of 10 μM., and incubated for 18 hours. Then, cells were collected to perform a western blot analysis using a anti-p38 specific antibody.

Chapter 5

Chemoselective protein labeling with 3-bromo-1,2,4,5-tetrazines

5.1. Protein modification with 1,2,4,5-tetrazines

5.1.1. Background:

There is a growing interest for protein decoration with useful chemical handles able to modulate or expand their functions. From labelling proteins with fluorophores to study their behavior, to conjugating antibodies with anticancer agents, the potential for chemical protein modification is enormous. For that reason, over recent years the toolbox of methods to attain protein labeling has broaden with modern methodologies, including the genetic incorporation of unnatural amino acids that will be discussed in **Chapters 7**, **8** and **9**.

However, prior to the emergence of these novel methodologies, classical strategies for protein labeling normally relied on targeting the functional groups contained in proteinogenic amino acids. Specifically, the nucleophilicity of these functionalities is normally the exploited parameter, in reaction with electrophilic reagents. These strategies are still the mainly used for protein modification, partly due to their technical simplicity and the vast amount of commercially available reagents. However, a caveat when employing these methodologies is that the targeted protein needs to be purified in solution. Ideally, these nucleophilic substitution reactions should preferentially happen with one amino acid type over the others, in search of a chemical homogeneity within the modified protein. If this condition is fulfilled, then the labeling reaction can be considered *chemoselective*.[1] Furthermore, if the reaction not only discriminates a certain amino acid identity but also a particular amino acid at a specific position over the others, then the labelling can be considered *site-selective*.[2]

Although examples of tyrosine[3–5] or histidine[6] selective modification exist, the usual targeted amino acids are cysteines (through their thiols) or lysines (through their primary amines), which are the ones presenting a higher nucleophilicity.[1] The selectivity between these two residues depend on different factors. Firstly, lysines are more prevalent, representing 5.9% of amino acids in the human proteome, whereas cystines only account for 1.9%, being the second least abundant amino acid.[7,8] Furthermore, whilst cysteines tend to be buried within the protein forming disulfide bonds critical for structure and therefore inaccessible for modification, lysines are

normally found in the solvent-exposed region of proteins due to their ionic character and subsequent hydrophilicity, easily accessible for modification. It is for these reasons that lysines are, in principle, more convenient for protein modification, although usually chemoselective methods for lysine-labelling also modify the *N*-terminus.[9] Typical reagents employed for lysine-selective labelling include activated carboxylic acids, such as N-hydroxysuccinimide (NHS)-esters,[10] but also iso(thio)cyanates,[11,12] squaric acids[9] and carbonyls, that undergo a reductive amination[13] (**Scheme 5-1, a**).

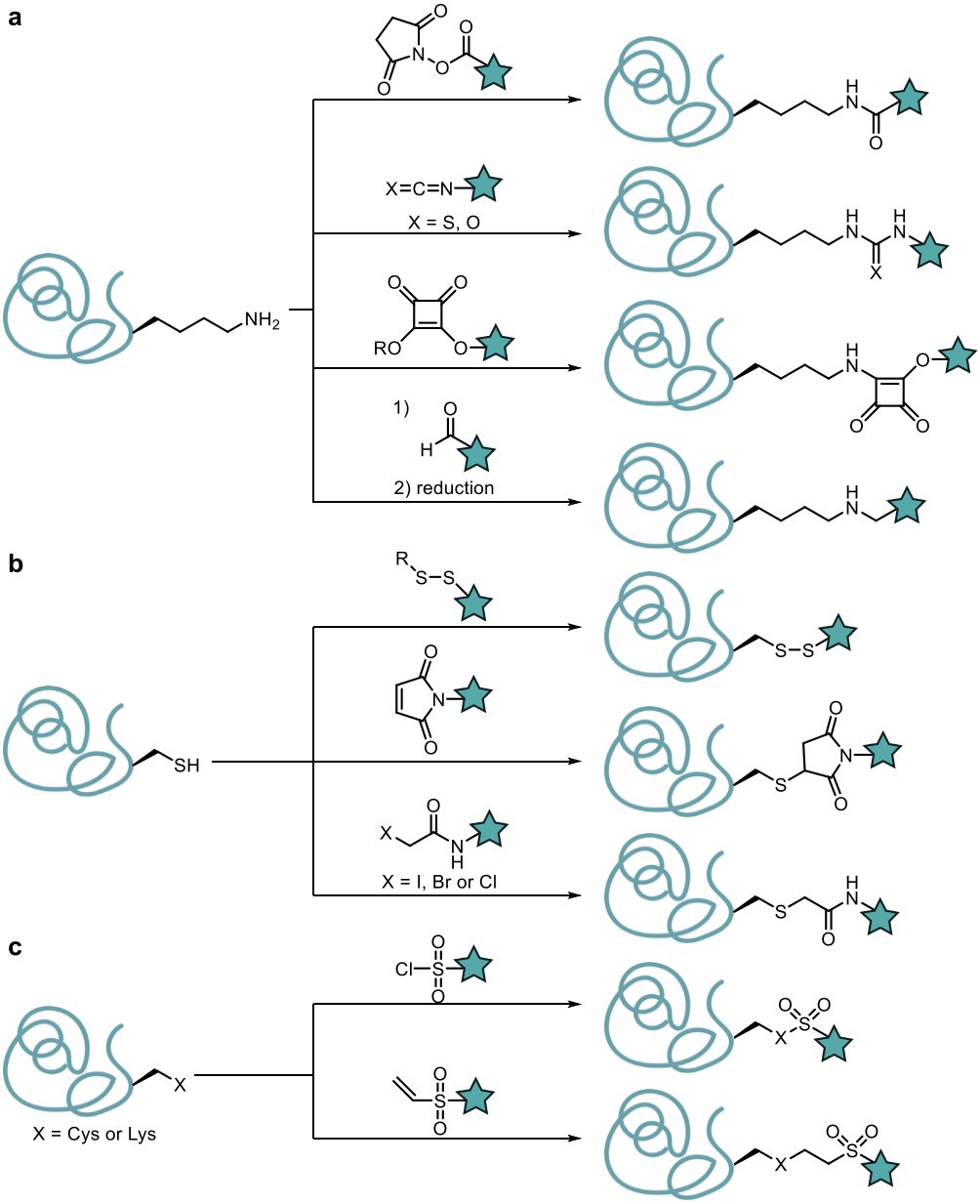


Figure 5-1: Classic reagents employed for chemoselective labeling of lysine or cysteine (the blue star represents a cargo of choice). a) Lysine-selective labeling. **b)** Cysteine-selective labeling. **c)** Lysine- or cysteine-selective labeling methods depending on the experimental conditions.

Despite being less avaiable, cysteine chemoselective labeling can be achieved if free nucleophilic cysteines are present in the protein surface. They can occur naturally (which is the case for certain proteins, such as albumin or glutathione), or they can be inserted through site-directed mutagenesis, normally exchanging a naturally occurring solvent-exposed natural amino acid with similar structural characteristics. In these cases, the cysteine selectivity over lysine can be achieved through variations in the pH values: in acidic media, lysines will be predominantly protonated and therefore not nucleophilic, which will increase the selectivity towards cysteines. By this means, a number known functionalities can react chemoselectively with cysteines. For instance, certain disulfides can be suitable reagents for a disulfide exchange;[14] Micheal acceptors, like maleimides,[15] can undergo 1,4-addition reactions; and α-halocarbonyls, such as haloacetamides,[16] are electrophilic reagents for nucleophilic substitution reactions (**Fig. 5-1, b**).

Finally, there are certain reagents that their selectivity towards lysine or cysteine is very much dependent on the labeling conditions, and are used for both kinds of chemoselective labeling. Examples of such functional groups are, for example, electrophilic reagents like sulfonyl chlorides[17] and Micheal acceptors like vinyl sulfones[18] (**Scheme 5-1, c**).

5.1.2. Results:

Encouraged by the excellent reactivity in S_NAr reactions for both **1** and **2** described in **Chapter 3** (**Section 3.1.**), we envisaged that 3-bromotetrazines could be suitable regents for chemical protein modification. Since we observed good reactivity under mild conditions with aromatic and aliphatic amines, alcohols and thiols, we performed and initial screen of reactivity with a mixture of the nucleophilic natural amino acids, with a Fmoc protecting group on their N^{α} to avoid side-reactions (**Fig. 5-2**). Unfortunately, we were unable to use Fmoc-cysteine as it generates disulfide bridges unless a reducing agent is added in the media, which would also reduce the tetrazine ring.

We followed the reaction of each Fmoc-amino acid through LC-MS, comparing their peak areas before and after the addition of **2.** Results show how over 1 equivalent of **2**, no significant labeling can be observed, maybe due to certain instability of **2** in

aqueous media. However, at 10 added equivalents, a preferential reactivity with Fmoc-lysine is observed, followed by Fmoc-tyrosine. Importantly, no cross-reactivity with other Fmoc-amino acids is detected, not even after the addition of 50 equivalents of **2**.

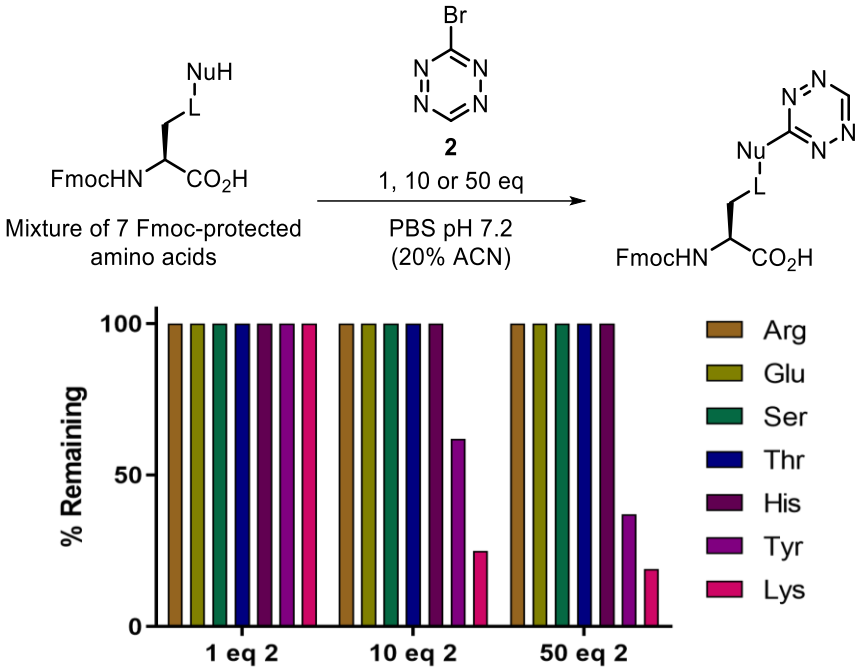


Figure 5-2: Amino acid reactive preference of 3-bromotetrazine 2. a) Reaction scheme of the mixture of the different protected amino acids with **2.** Reactions were performed in PBS pH 7.2 (with 20% acetonitrile to ensure Fmoc-protected amino acids solubilization). **b)** Percentage of remaining of each Fmoc-protected amino acid according to peak areas in the LC-MS spectra.

As expected, a preferential reactivity with Fmoc-lysine was observed. Consequently, we re-examined the kinetics in the iEDDA cycloaddition determined in **Chapter 3** (**Section 3.3** and **3.6.3.**) for tetrazinyl-amino acids with **BCN-OH**. Our observations that monosubstituted tetrazines resulting from the S_NAr reaction with bromotetrazine **2** react faster than the products with **1** correlates with previous literature.[19] Specifically for lysine, the second order rate constant of **11a** (k_2 = 1.33 $M^{-1}s^{-1}$, **Fig. 5-3, a**) is in the range of the relatively fast and quantitative bioorthogonal reactions (**Fig. 5-3, b**), and is ~35 times faster than that of **10a** (k_2 = 0.04 $M^{-1}s^{-1}$). As a result, we decided to continue the protein labeling experiments only with 3-bromo-1,2,4,5-tetrazine **2**.

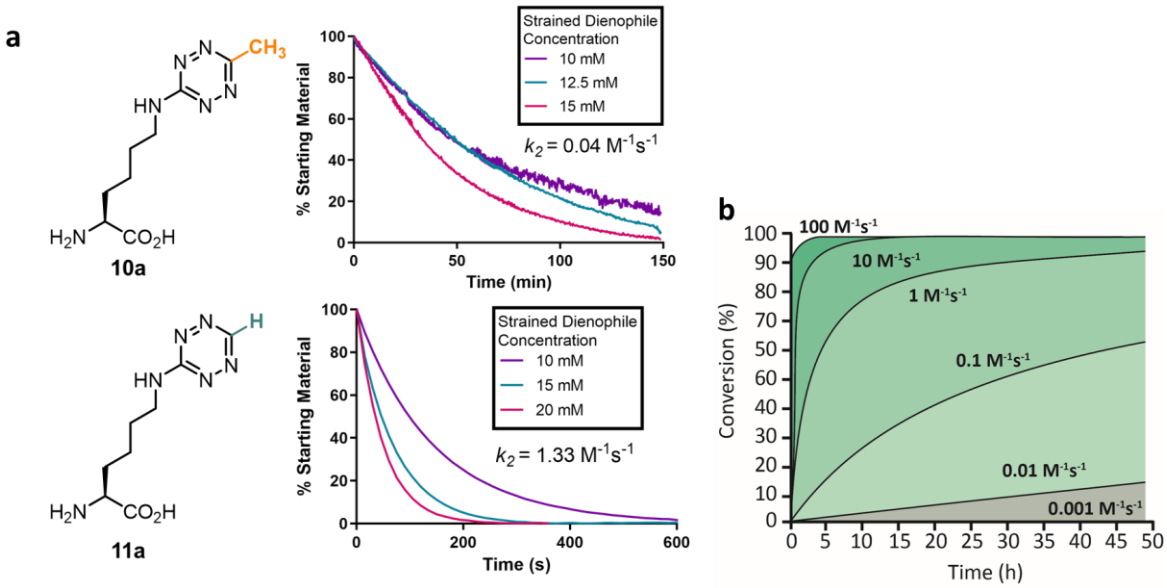


Figure 5-3: 2nd order rate constant of tetrazinyl-lysines 10a and 11a. a) Comparison between **10a** and **11a** in iEDDA reaction with different controlled excesses of **BCN-OH** as the strained dienophiles. **b)** Simulation of 2nd order reactions between reactants at 100 µM concentration. Adapted from Oliveira *et al*. (reference 10, **Chapter 3**).

To date, there is only a single article reporting chemoselective reagents to decorate proteins with tetrazines, specifically employing carboxylic acid derivatives suitable to specifically target lysines upon activation.[20] Interestingly, the functional group normally incorporated for iEDDA reactions is the dienophile. However, we anticipated that inserting the tetrazine instead would be a desirable approach, since the strained dienophiles commonly used for bioorthogonal reactions tend to be bulky, apolar molecules that can be masked by hydrophobic interactions with the labeled protein.[21] Furthermore, some of the most commonly used alkenes or alkynes suffer from limited stability in physiological media, which is the case for the broadly used **TCO**, that undergoes isomerization to the *cis* form.[22] Alternatively, tetrazines are polar molecules with tendency for being at the solvent-exposed region of the labelled proteins, and their stability can be tuned depending on their substituents, being the EDG-substituted tetrazines much more stable than those with EWG.[20]

Therefore, to validate **2** as a potential reagent for chemical protein modification, Ribonuclease A (RNAse A) was chosen as a model protein due to its relative small size (13.6 KDa), high stability and limited number of solvent-exposed lysine residues (10 in total). Also, it had been previously used in bioconjugation experiments, where the most reactive lysines had been determined.[23] To our delight, we observed that **2** was

soluble in aqueous media, which is an advantage in comparison with other labeling reagents, where the use of co-solvents like DMF or DMSO is required. We attempted the labeling using 10 or 50 equivalents of **2** at 3 different pH values: 6.0, 7.2 or 8.0. (**Fig. 5-4**).

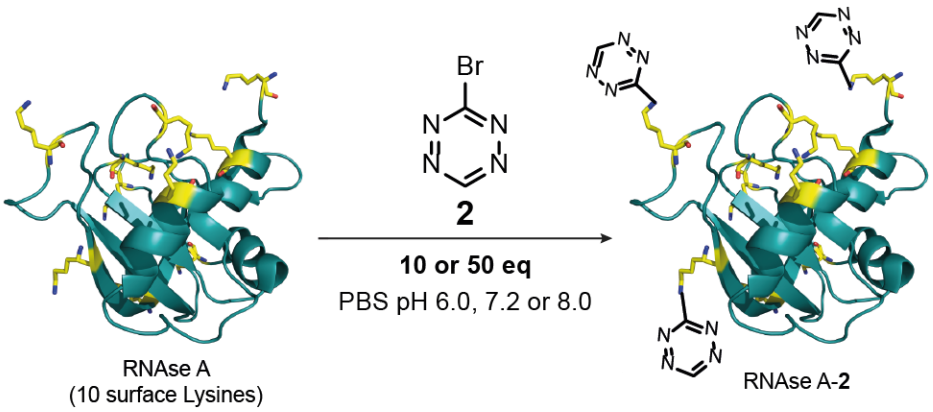


Figure 5-4: RNAse A labeling scheme with different equivalents of reagent 2 at different pH values.

After 30 minutes of reaction time, the different protein samples were analyzed intact by LC-MS (**Fig. 5-5**). As expected, less modified residues were observed at more acidic pH values, where a higher number of protonated amines impede their nucleophilic attack.[24] Likewise, more labeling was detected when a higher number of equivalents of **2** were added.

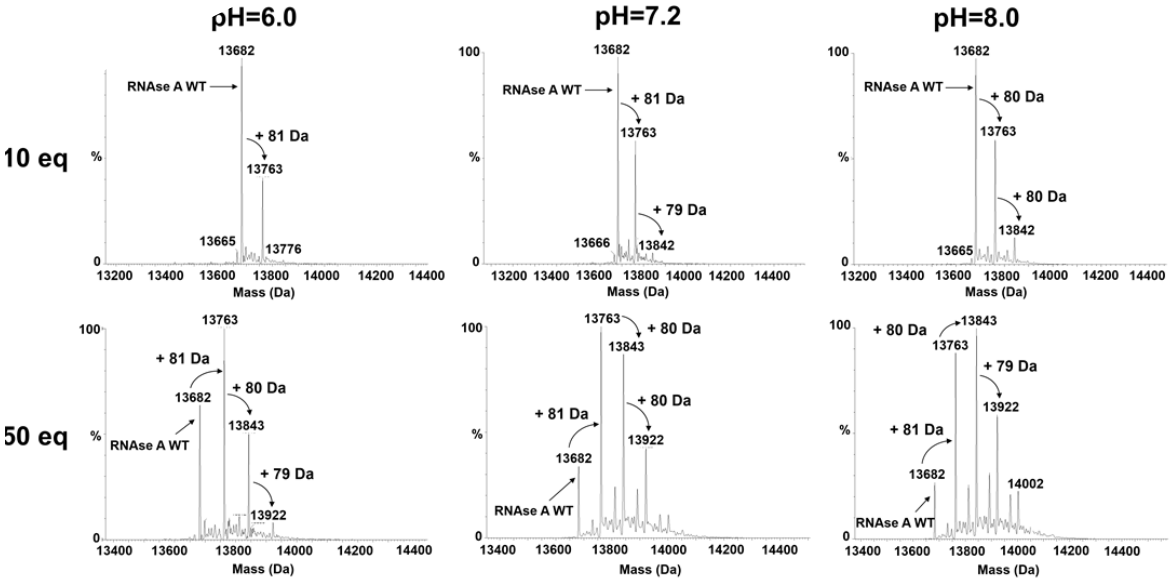


Figure 5-5: Intact protein LC-MS of the labeled RNAse A with 2 under different conditions. Deconvoluted MS1 spectra corresponding to the different labeling conditions. The 79-81 Da mass increase between each major peak correspond to the addition of a single molecule of **2.**

Quantification of each labeled species was approximated through the area under the peak in the obtained chromatograms, and its distribution plotted (**Fig. 5-6**).

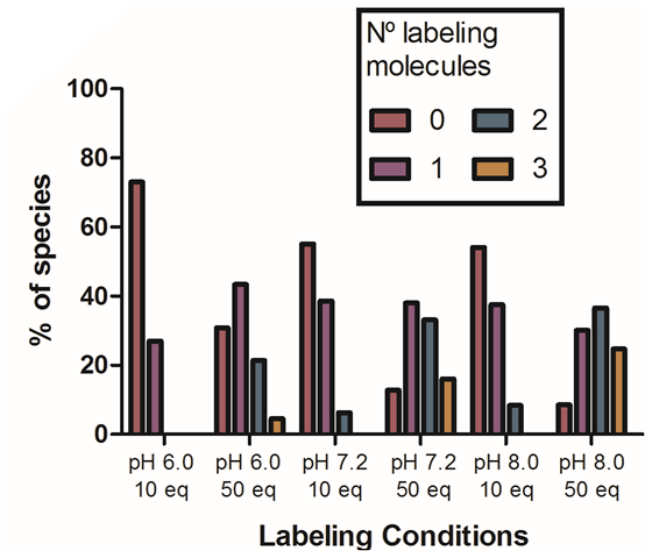


Figure 5-6: Quantification of the relative abundance of the different species of labeled RNAse A-**2**.

Additionally, in chemical protein modification it is of utmost importance that the labeling molecules do not disrupt the secondary structure of the protein. It is known that hydrophobic labeling reagents can be buried within the protein structure,[21] which may cause a disruption of its tertiary structure. Tetrazines, however, are relatively hydrophilic, which might be favorable for their localization at the solvent exposed regions of the proteins. For that reason, we decided to compare the structures of unlabeled RNAse A with the same protein when labeled with a different number of equivalents of **2** by circular dichroism. The obtained spectra revealed that the protein folding seemed to be completely unaffected by the labeling with 10 equivalents, and minimally perturbed when using 50 (**Fig. 5-7**).

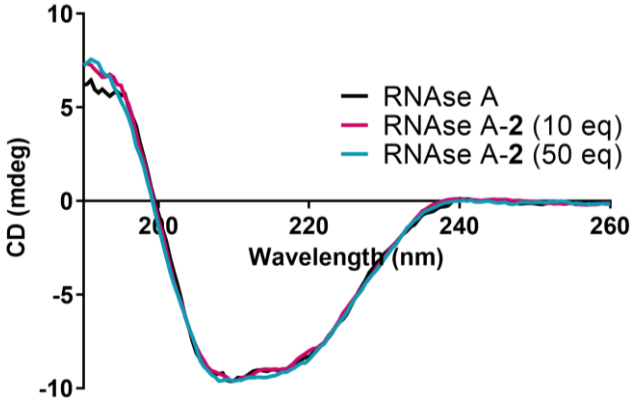


Figure 5-7: Circular dichroism spectra of RNAse A labeled with different equivalents of 2. Experiments with RNAse A, RNAse A-2 (10 eq) and RNAse A-1 (50 eq) were performed in PBS pH 7.2, in which NaCl was replaced by NaF to avoid absorbance interference.

5.2. Chemoselective lysine-labeling with **2**

5.2.1. Determination of reactive sites by LC-MS/MS:

To be able to confirm the preference for lysines in the experiment between **2** and the Fmoc-protected amino acid mixture, we decided to perform LC-MS/MS studies on the digested protein. Our aim was to determine accurately the labeling sites of bromotetrazine **2**, under our labeling conditions, in order to corroborate the chemoselective character of **2**. Again, RNAse A was used as the model protein and, since we wanted to assure enough labeling ratio, we chose 50 equivalents of **2** in PBS pH 7.2 as the standard conditions in all future labeling experiments.

Consequently, the tryptic digestion of the full-length protein showed that lysines were primarily targeted, as >95% of the modified peptides were lysine-labeled (**Fig. 5-8, a** and **b**). Only minimal reactions with tyrosine (Tyr76) and histidine (His12, His48 and His105) were observed, accounting for less than 5% of the peptides that showed modification. Therefore, this experiment demonstrates the chemoselectivity of **2** for lysines in a protein without free nucleophilic cysteines, in good agreement with the selectivity observed with the mixture of Fmoc-amino acids.

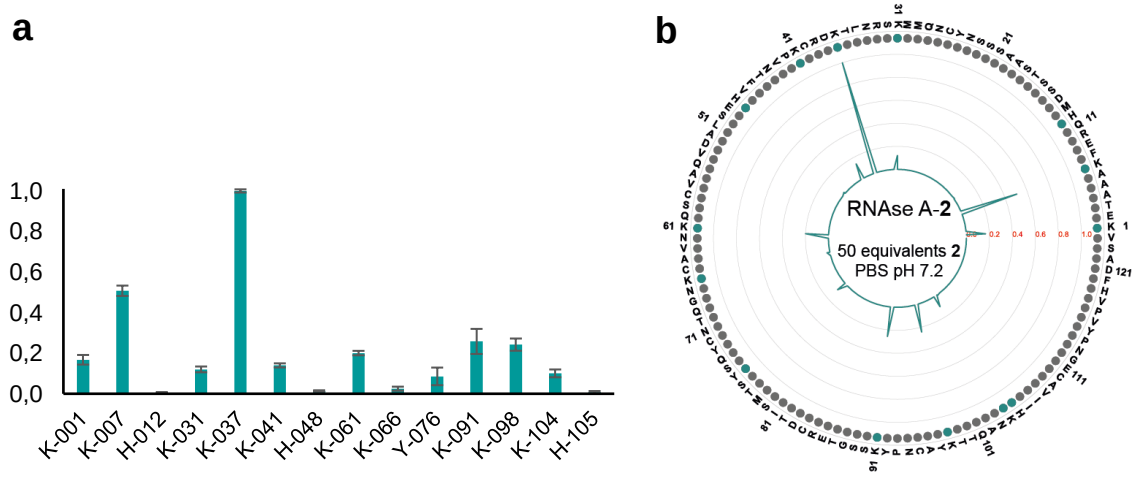


Figure 5-8: Labeling sites determination of RNAse A-2 by digested protein analysis through LC-MS/MS. a) Ratio of labeled sites with compound **2**, normalized to lysine 37. Labeling with 50 equivalents of **2** in PBS pH 7.2 was carried out in triplicates. **b)** Representation of labeled sites in the full sequence of RNAse A. The height of each peak corresponds with the ratio of labeling, normalized to lysine 37.

We then wanted to test the selectivity in a protein having available cysteine thiols. To this end, we reacted bovine serum albumin with **2** under the standard conditions (50 equivalents **2**, PBS pH 7.2). Somewhat surprisingly, we did not find any significant substitution on the free nucleophilic cysteine 58 whereas several lysine residues readily incorporated the tetrazine fragment. (**Fig. 5-9**). Therefore, it is confirmed that **2** is able to discriminate between lysines and cysteines, being the reaction highly chemoselective.†

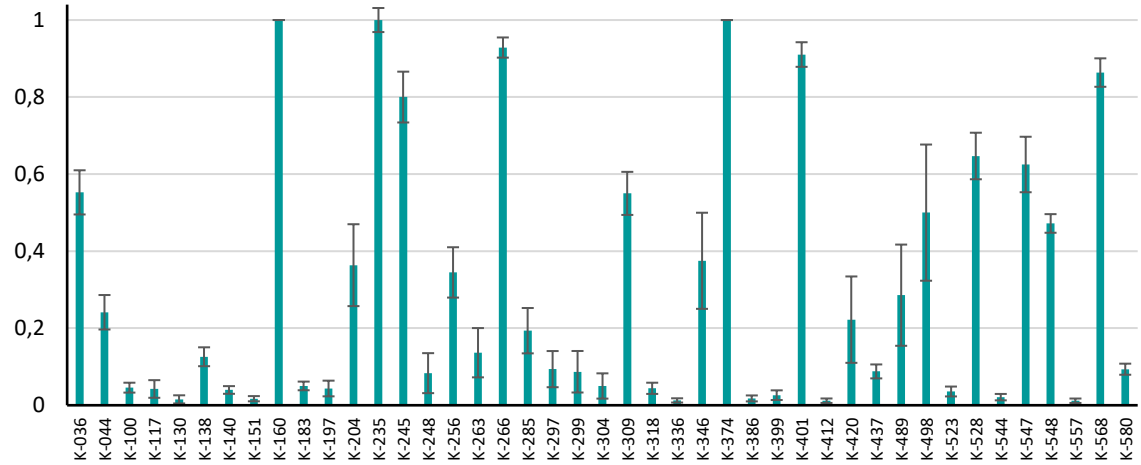


Figure 5-9: Labeling sites determination of Albumin-2 by digested protein analysis through LC-MS/MS. The labeling was performed under the same conditions as the ones used for RNAse A (50 equivalents of **2** in PBS pH 7.2).

5.3. Summary:

3-bromotetrazine **2** is a suitable reagent for protein labeling. LC-MS/MS studies on a model protein (RNAse A) in solution in which **2** was added revealed a selective reactive preference towards lysines (*chemoselectivity*), thereby resulting in a homogeneous population of 3-aminotetrazines at the protein surface. The amount of labeling can be tuned varying the number of equivalents and the pH of the solution.

†In a recent article,[25] a chemoselectivity of **2** towards cysteines is described. Probably, the divergence in the results arises from different labeling conditions: herein, pH 7.2 was the chosen condition, whereas pH 4.15 was used in Schnell *et al*.

5.4. Bibliography:

1 O. Boutureira and G. J. L. Bernardes. Advances in chemical protein modification. *Chem. Rev.* 2015, **115**, 2174–2195.

2 N. Krall, F. P. da Cruz, O. Boutureira and G. J. L. Bernardes. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* 2015, **8**, 1–11.

3 T. L. Schlick, Z. Ding, E. W. Kovacs and M. B. Francis. Dual-Surface Modification of the Tobacco Mosaic Virus. *J. Am. Chem. Soc.* 2005, **127**, 3718–3723.

4 J. Gavrilyuk, H. Ban, M. Nagano, W. Hakamata and C. F. Barbas. Formylbenzene Diazonium Hexafluorophosphate Reagent for Tyrosine-Selective Modification of Proteins and the Introduction of a Bioorthogonal Aldehyde. *Bioconjug. Chem.* 2012, **23**, 2321–2328.

5 M. Lorenzi, C. Puppo, R. Lebrun, S. Lignon, V. Roubaud, M. Martinho, E. Mileo, P. Tordo, S. R. A. Marque, B. Gontero, B. Guigliarelli and V. Belle. Tyrosine-Targeted Spin Labeling and EPR Spectroscopy: An Alternative Strategy for Studying Structural Transitions in Proteins. *Angew. Chem., Int. Ed.* 2011, **50**, 9108–9111.

6 S. Jia, D. He and C. J. Chang. Bioinspired thiophosphorodichloridate reagents for chemoselective histidine bioconjugation. *J. Am. Chem. Soc.* 2019, **141**, 7294–7301.

7 L. P. Kozlowski. Proteome-pI: proteome isoelectric point database. *Nucleic Acids Res.* 2017, **45**, D1112–D1116.

8 C. B. Rosen and M. B. Francis. Targeting the N terminus for site-selective protein modification. *Nat. Chem. Biol.* 2017, **13**, 697–705.

9 F. R. Wurm and H. A. Klok. Be squared: Expanding the horizon of squaric acid-mediated conjugations. *Chem. Soc. Rev.* 2013, **42**, 8220–8236.

10 J. S. Nanda and J. R. Lorsch. in *Methods in Enzymology*, 2014, pp. 87–94.

11 P. Rosa-Neto, B. Wängler, L. Iovkova, G. Boening, A. Reader, K. Jurkschat and E. Schirrmacher. [18 F]SiFA-isothiocyanate: A New Highly Effective Radioactive Labeling Agent for Lysine-Containing Proteins. *ChemBioChem* 2009, **10**, 1321–1324.

12 D. E. Mason and D. C. Liebler. Quantitative Analysis of Modified Proteins by LC–MS/MS of Peptides Labeled with Phenyl Isocyanate. *J. Proteome Res.* 2003, **2**, 265–272.

13 D. Chen, M. M. Disotuar, X. Xiong, Y. Wang and D. H.-C. Chou. Selective N-terminal functionalization of native peptides and proteins. *Chem. Sci.* 2017, **8**, 2717–2722.

14 H. P. Hemantha, S. N. Bavikar, Y. Herman-Bachinsky, N. Haj-Yahya, S. Bondalapati, A. Ciechanover and A. Brik. Nonenzymatic Polyubiquitination of Expressed Proteins. *J. Am. Chem. Soc.* 2014, **136**, 2665–2673.

15 J. S. Nanda and J. R. Lorsch. in *Methods in Enzymology*, 2014, pp. 79–86.

16 G. Schwarz, S. Beck, M. G. Weller and M. W. Linscheid. MeCAT—new iodoacetamide reagents for metal labeling of proteins and peptides. *Anal. Bioanal. Chem.* 2011, **401**, 1203–1209.

17 J.-K. Lin and J.-Y. Chang. Chromophoric labeling of amino acids with 4-dimethylaminoazobenzene-4'-sulfonyl chloride. *Anal. Chem.* 1975, **47**, 1634–1638.

18 J. Morales-Sanfrutos, J. Lopez-Jaramillo, M. Ortega-Muñoz, A. Megia-Fernandez, F. Perez-Balderas, F. Hernandez-Mateo and F. Santoyo-Gonzalez. Vinyl sulfone: A versatile function for simple bioconjugation and immobilization. *Org. Biomol. Chem.* 2010, **8**, 667–675.

19 J. A. Wagner, D. Mercadante, I. Nikic, E. A. Lemke and F. Gräter. Origin of Orthogonality

of Strain-Promoted Click Reactions. *Chem. Eur. J.* 2015, **21**, 12431–12435.

20 A. Maggi, E. Ruivo, J. Fissers, C. Vangestel, S. Chatterjee, J. Joossens, F. Sobott, S. Staelens, S. Stroobants, P. Van Der Veken, L. Wyffels and K. Augustyns. Development of a novel antibody-tetrazine conjugate for bioorthogonal pretargeting. *Org. Biomol. Chem.* 2016, **14**, 7544–7551.

21 M. K. Rahim, R. Kota and J. B. Haun. Enhancing reactivity for bioorthogonal pretargeting by unmasking antibody-conjugated trans -cyclooctenes. *Bioconjug. Chem.* 2015, **26**, 352–360.

22 R. Rossin, S. M. Van Den Bosch, W. Ten Hoeve, M. Carvelli, R. M. Versteegen, J. Lub and M. S. Robillard. Highly reactive trans-cyclooctene tags with improved stability for diels-alder chemistry in living systems. *Bioconjug. Chem.* 2013, **24**, 1210–1217.

23 X. Chen, K. Muthoosamy, A. Pfisterer, B. Neumann and T. Weil. Site-selective lysine modification of native proteins and peptides via kinetically controlled labeling. *Bioconjug. Chem.* 2012, **23**, 500–508.

24 M. J. Matos, B. L. Oliveira, N. Martínez-Sáez, A. Guerreiro, P. M. S. D. Cal, J. Bertoldo, M. Maneiro, E. Perkins, J. Howard, M. J. Deery, J. M. Chalker, F. Corzana, G. Jiménez-Osés and G. J. L. Bernardes. Chemo- and Regioselective Lysine Modification on Native Proteins. *J. Am. Chem. Soc.* 2018, **140**, 4004–4017.

25 S. D. Schnell, L. V. Hoff, A. Panchagnula, M. H. H. Wurzenberger, T. M. Klapötke, S. Sieber, A. Linden and K. Gademann. 3-Bromotetrazine: labelling of macromolecules via monosubstituted bifunctional s -tetrazines. *Chem. Sci.* 2020, **11**, 3042–3047.

5.5. Experimental section

5.5.1. Labeling procedures with **2**

5.5.1.1. Mixture of Fmoc-protected amino acids with **2**:

Each N_α-Fmoc-protected amino acid[††] (Fmoc-Arg-OH, Fmoc-Glu-OH, Fmoc-Ser-OH, Fmoc-Thr-OH, Fmoc-His-OH, Fmoc-Tyr-OH and Fmoc-Lys-OH) was dissolved in DMSO to a final concentration of 100 mM, and 3 µL of each added to 800 µL PBS pH 7.2 (20% ACN) for a final concentration of 0.3 mM each. Then, a stock solution of **2** 100 mM in water was prepared and the relevant volumes added to the previous solution (1 equivalent = 3 µL, 10 equivalents = 30 µL, 50 equivalents = 150 µL). Finally, the required water to reach total volume of 1000 µL was added. The experiment was carried out in triplicates.

Each reaction mixture was shaken 30 mins at room temperature, after which the samples were diluted 1:30 with 1% formic acid aqueous solution supplemented with 3% ACN and directly injected to the LC/MS. Each sample was loaded to a reversed phase C18 analytical column (BioBasic C18, 5µm, 2.1x150 mm, Thermo) with a 90 min run, using two consecutive steps with linear gradients from 5% to 80% of B in 60min, from 80% to 90% of B in 9 min, followed by isocratic elution at 90 % of B in 11 min and stabilization to initial conditions (A= 0.1% FA in water, B= 0.1% FA in ACN). The chromatographic system used was a Finnigan, Mod Surveyor MS (Thermo Electron Corporation) comprised by a quaternary pump and an autosampler. The column outlet was directly connected to an Advion TriVersa NanoMate (Advion) fitted on an LTQ-FT Ultra mass spectrometer (Thermo). Full MS1 scans were acquired in the FT with the resolution (defined at 400 m/z) set to 100,000. The ion count target value was 1,000,000. Spray voltage in the NanoMate source was set to 1.70 kV. Capillary voltage and tube lens on the LTQ-FT were tuned to 35 V and 100 V. The spectrometer was working in positive polarity mode. At least one blank run before of analysis were performed in order to ensure the absence of cross contamination from previous samples. Data processing was performed with Xcalibur software vs 2.0SR2 (Thermo

[††] Fmoc-cysteine was not included in the mixture due to technical limitations (need to add a reducing agent to avoid disulfide bond formation, leading to tetrazine reduction to dihydrotetrazine).

Scientific). Relative quantification of the starting material (Fmoc-aa-OH) was calculated from XIC (extracted ion chromatogram) peak areas using QuanBrowser from Xcalibur software.

5.5.1.2. Labeling of proteins with **2**:

0.9 mg **2** were dissolved in 1 mL H_2O (to reach a concentration of 5591 µM) and the relevant volumes were added to a solution of 3.42 µL RNAse A (Qiagen) 100 mg/mL in 996.6 µL PBS buffer at pH 6.0, 7.2 or 8.0, ([RNAse A]= 25 µM) to reach the desired number of equivalents (10 eq = 44.7 µL, 50 eq = 223.6 µL). Reactions were left at room temperature for 30 min with end-to-cap mixing, followed by buffer exchange using Amicon® Ultra-2 mL centrifugal filters 3000 MWCO to 150 mM Ammonium Acetate pH 7.4 and concentrated to half of the volume.

The same experimental procedure was followed for the labeling of bovine serum albumin (Merck), only with 50 equivalents of **2** in PBS pH 7.2.

5.5.2. Circular dichroism:

The circular dichroism spectra of 10 µM solutions of RNAse A and RNAse A-**2** (10 and 50 eq) in phosphate buffered saline (PBS, replacing NaCl for NaF) pH 7.2 were measured in a JASCO spectropolarimeter at 293 K using a 1 mm path length quartz cuvette acquired every 0.5 nm with 5 accumulations.

5.5.3. Intact protein analysis by LC-MS:

20 µL of sample were injected automatically to a BioSuite pPhenyl 1000 (Waters, 10 µm RPC 2.0x75 mm) column at a flow rate of 100 µL/min using an Acquity UPLC system (Waters Corporation) provided with a Binary Solvent Manager and an automatic Autosampler. Samples were eluted using a linear gradient from 2% to 5% B in 5min and from 5% to 80% B in 60min (A= 0.1% FA in water, B= 0.1% FA in CH_3CN). The column outlet was directly connected to an LCT-Premier XE mass spectrometer (Waters). Capillary voltage and cone voltage were set to 3000 V and 100 V respectively. Desolvation temperature and source temperature were set to 350°C and 120°C. Cone gas flow and desolvation gas flow were set to 50 and 600 L/h. The

mass spectrometer acquired full MS scans (400-4000 m/z) working in positive polarity mode.

Data was acquired with MassLynx software, V4.1.SCN704 (Waters Inc.). MS spectra corresponding to chromatographic peaks were summed. Charged species in the resulting spectrum were deconvoluted to their zero charged average masses using the integrated MaxEnt1 (maximum entropy) algorithm in MassLynx Software Vs. 4.1.SCN704 (Waters). The algorithm calculates deconvoluted masses peak intensities.

5.5.4. Digested protein analysis by LC-MS/MS:

The different RNAse A, RNAse A-**2**, Albumin and Albumin-**2** samples were tryptic digested directly in solution. Briefly, samples were reduced with 1,4-dithiotheitol (DTT) 2 mM for 15 min at 95°C and alkylated for 30 min in the dark with 3-indoleacetic acid (IAA) 5 mM. Then, the unreacted IAA was quenched adding DTT 2 mM. The solution digestion was performed with trypsin (0.1 µg/µL) in 50mM NH_4HCO_3 at 37°C overnight. The digestion was stopped by adding formic acid. Samples were desalted using polyLC C18 and reconstituted in 25 µL of 1% formic acid aqueous solution with 3% ACN for MS analysis. Each sample was injected twice in order to have technical replicates.

The nano-LC-MS/MS set up was as follows: Digested peptides were diluted in 1% FA with 3% ACN. Samples were loaded to a µ-precolumn 300 µm i.d. × 5 mm PepMap100, 5 µm, 100 Å, C18 (Thermo Scientific) at a flow rate of 15 µl/min using a Dionex Ultimate 3000 chromatographic system (Thermo Scientific). Peptides were separated using a C18 analytical column (Acclaim PepMap® RSLC 50 µm × 50 cm, nanoViper, C18, 2 µm, 100Å, Thermo Scientific) with a 90 min run, comprising three consecutive steps with linear gradients from 3 to 35% B in 63 min, from 35 to 50% B in 5 min, and from 50 % to 85 % B in 2 min, followed by isocratic elution at 85 % B in 5 min and stabilization to initial conditions (A= 0.1% FA in water, B= 0.1% FA in ACN). The column outlet was directly connected to an Advion TriVersa NanoMate (Advion) fitted on an Orbitrap Fusion Lumos™ Tribrid (Thermo Scientific). The mass spectrometer was operated in a data-dependent acquisition mode. Survey MS scans were acquired in the orbitrap with the resolution (defined at 200 m/z) set to 120,000. The lock mass was user-defined at 445.12 m/z in each Orbitrap scan. The top speed (most intense)

ions per scan were fragmented by HCD and detected in orbitrap. The ion count target value was 400,000 and 10,000 for the survey scan and for the MS/MS scan respectively. Target ions already selected for MS/MS were dynamically excluded for 15s. Spray voltage in the NanoMate source was set to 1.60 kV. RF Lens were tuned to 30%. Minimal signal required to trigger MS to MS/MS switch was set to 5,000. The spectrometer was working in positive polarity mode and singly charge state precursors were rejected for fragmentation. Two technical replicas were analyzed per sample.

A database search was performed with Proteome Discoverer (PD) software v2.1 (Thermo) using Sequest HT search engine and Amanda 2.0 algorithm. MaxQuant (MQ) 1.6.2.3 software with Andromeda search engine was also used. The database used was SwissProt *E. coli* release 2017 06 with user proteins and contaminants. Searches were run against targeted and decoy database to determine the false discovery rate (FDR). Search parameters included trypsin specificity, allowing for two missed cleavage sites, oxidation in methionine, **2** (C_2N_4; 80.0122 Da) in lysine, tyrosine, histidine and cysteine as dynamic modifications. Peptide mass tolerance was 10 ppm and the MS/MS tolerance was 0.02 Da. The ptmRS node in PD was used to provide confidence measure for the localization of the modification in the peptide sequences identified. Peptides with a FDR < 1% were considered as positive identifications with a high confidence level. Only pSites with a PTM probability threshold of 75% were considered unambiguous sites.

Site modification ratios (r) were calculated taking into account the PSMs (peptide spectral matches) of the corresponding modified with **2** and non-modified peptides identified comprising that site:

$$r = \frac{\text{Nmod.}}{\text{Nmod.} + \text{Nunmod.}}$$

Where $N_{mod.}$ / $N_{unmod.}$ denote the modified/unmodified PSMs count for a given site. The ratio was calculated from the results obtained with Sequest HT (PD), Amanda 2.0 (PD) and Andromeda (MQ) search engines. For the final output, all node results were combined by choosing the node with the maximum number of PSM for each Mod-site. The uncertainty of the ratio was calculated considering the $N_{mod.}$ and $N_{unmod.}$ values

pair for each site as a binomial distribution. This allows to obtain a variance for each $N_{mod.}$ and $N_{unmod.}$ values pair and to calculate their associated error.

5.5.4.1. LC-MS/MS results for RNAse A-**2**:

Each modified residue was quantified and normalized to the most detected one (in this case, Lysine 37) (**Table 5-1**).

Table 5-1: Determined labeled peptides in RNAse A-2 by LC-MS/MS. (Mod. = modified, No Mod. = unmodified).

Site	Mod.	No Mod.	Mod. + No Mod.	Ratio	Ratio error
K-001	30	149	179	0,167597765	0,023704815
K-007	98	95	193	0,507772021	0,025449329
H-012	2	401	403	0,004962779	0,003483167
K-031	49	359	408	0,120098039	0,014292146
K-037	7	0	7	1	0,0078125
K-041	126	775	901	0,139844617	0,010069109
H-048	12	889	901	0,013318535	0,00376852
K-061	181	721	902	0,200665188	0,010989977
K-066	6	232	238	0,025210084	0,009908557
Y-076	3	32	35	0,085714286	0,04345259
K-091	8	23	31	0,258064516	0,061735081
K-098	31	97	128	0,2421875	0,030125285
K-104	20	179	199	0,100502513	0,019291053
H-105	2	250	252	0,007936508	0,00554546

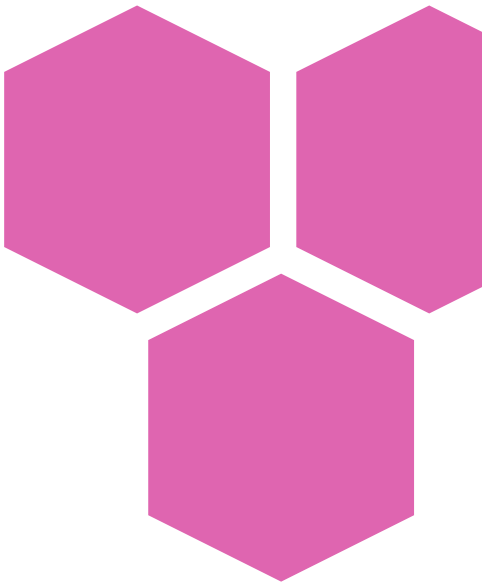
5.5.4.2. LC-MS/MS results for Albumin-**2**:

Each modified residue was quantified and normalized to the most detected one (in this case, Lysine 160) (**Table 5-2**).

Table 5-1: Determined labeled peptides in Albumin-2 by LC-MS/MS. (Mod. = modified, No Mod. = unmodified).

Site	Mod	No Mod.	Mod. + No Mod.	Ratio	Ratio error
K-036	21	17	38	0.552632	0.057350413
K-044	14	44	58	0.241379	0.044731553
K-100	11	230	241	0.045643	0.012845222
K-117	3	68	71	0.042254	0.022887606

K-130	2	129	131	0.015267	0.010550511
K-138	18	125	143	0.125874	0.024497282
K-140	14	339	353	0.03966	0.009983822
K-151	6	347	353	0.016997	0.006763913
K-160	15	0	15	1	3.05176E-05
K-183	17	322	339	0.050147	0.011274919
K-197	4	88	92	0.043478	0.020357883
K-204	4	7	11	0.363636	0.106305071
K-235	5	0	5	1	0.03125
K-245	20	5	25	0.8	0.06596969
K-248	2	22	24	0.083333	0.051928783
K-256	10	19	29	0.344828	0.065347995
K-263	3	19	22	0.136364	0.063970771
K-266	78	6	84	0.928571	0.026169801
K-285	6	25	31	0.193548	0.058849412
K-297	3	29	32	0.09375	0.046945519
K-299	2	21	23	0.086957	0.053887131
K-304	2	38	40	0.05	0.032782427
K-309	22	18	40	0.55	0.055898904
K-318	8	174	182	0.043956	0.01454283
K-336	5	394	399	0.012531	0.005499612
K-346	3	5	8	0.375	0.12475562
K-374	156	0	156	1	1.09476E-47
K-386	5	277	282	0.01773	0.007720628
K-399	4	149	153	0.026144	0.012567163
K-401	61	6	67	0.910448	0.031913478
K-412	4	335	339	0.011799	0.005796007
K-420	2	7	9	0.222222	0.112097436
K-437	18	187	205	0.087805	0.018114097
K-489	2	5	7	0.285714	0.131357205
K-498	2	2	4	0.5	0.176776695
K-523	7	189	196	0.035714	0.012790844
K-528	22	12	34	0.647059	0.060406614
K-544	6	279	285	0.021053	0.008326648
K-547	15	9	24	0.625	0.072027691
K-548	101	113	214	0.471963	0.024168293
K-557	4	338	342	0.011696	0.00574606
K-568	57	9	66	0.863636	0.036933542
K-580	32	310	342	0.093567	0.014350064



Chapter 6

Application of 1,2,4,5-tetrazinyl-labeled proteins in bioorthogonal drug release

6.1. Biorthogonal activation of prodrugs

6.1.1. Background:

The use of prodrugs as therapeutic agents dates back from ancient times, when medicines based on herbal extracts containing glycosides activated upon hydrolysis were employed for the treatment of common diseases. More recently, one of the first active principles ever synthesized and manufactured in modern drug discovery, acetylsalicylic acid, is an enzymatically deacetylated prodrug that yields the bioactive compound salicylic acid.[1]

Traditionally, the use of prodrugs aimed at increasing the bioavailability of the active principle, by improving their pharmacokinetic profile in terms of absorption, distribution, metabolization or excretion.[2] Some examples of commonly used prodrugs in therapy are the antihypertensive enalapril, the antihistaminic loratadine, or the antiviral valaciclovir.[3] Modern approaches rely on the activation of the prodrug specifically at the site of action, which becomes of particular relevance in anticancer treatments involving drugs with high cytotoxicity as a mean to lower their systemic side effects. The activation of these type of prodrugs can be triggered by different agents: by chemical reactions, such as hydrolysable esters,[4] reducible disulfide linkers[5] or compounds sensitive to oxygen levels;[6,7] or by enzymes, such as compounds labile to esterases, proteases or phosphatases.[8] In all cases, the prodrug should be ideally stable for systemic circulation, and the activator should be uniquely present in the site of release.

Recently, different bioorthogonal reactions have been employed for the activation of prodrugs *in vivo* (**Fig. 6-1**). The followed approach normally involves two participating molecules: the drug carrying a bioorthogonal functional group that masks its activity; and an activator molecule, carrying the bioorthogonal reactant counterpart. Upon bioorthogonal ligation between them, the formed product is designed so that it undergoes a rearrangement, finally leading to an elimination reaction of the free, active form of the drug. Using antibodies or other similar vectors, a targeted drug delivery and specific liberation at the site of action can be achieved.[9]

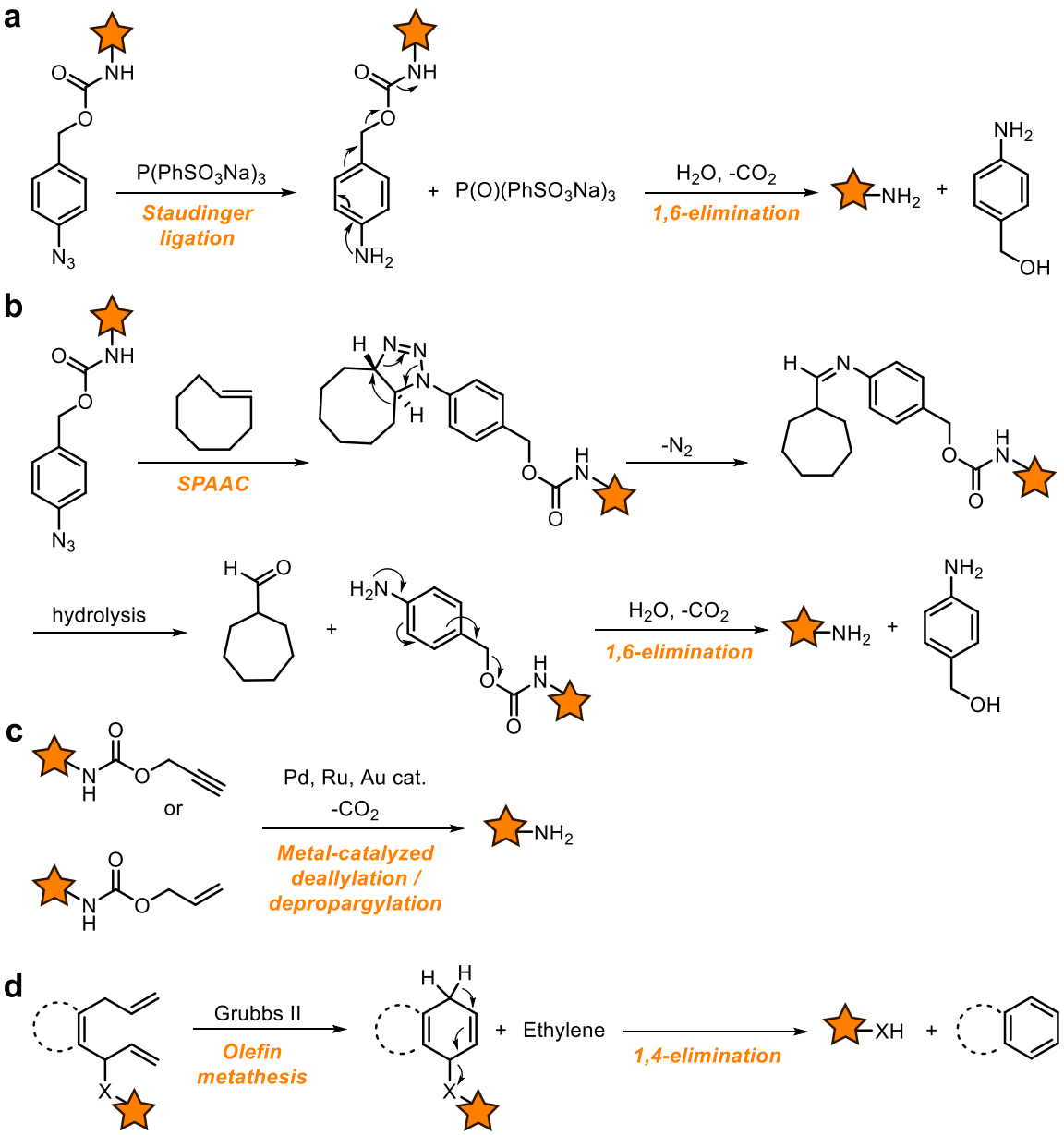


Figure 6-1: Bioorthogonal reactions employed for drug release (the orange star represents a cargo of choice). a) Based on the Staudinger ligation. **b)** Based on the strain-promoted azide-alkyne cycloaddition (SPAAC). **c)** Based on metal-catalyzed reactions. **d)** Based on olefin metathesis.

The first examples of bioorthogonal reactions applied for drug release employed the Staudinger ligation to liberate the anticancer drug doxorubicin (**Fig. 6-1, a**).[10,11] To generate a useful prodrug of doxorubicin with a lowered cytotoxicity, the primary amine in the aminoglycoside ring of the drug responsible for its DNA intercalating activity[12] is derivatized in the form of a 4-azidobenzyl carbamate. A reduction of the azide via Staudinger ligation leads to the formation of a self-immolative 4-aminobenzyl carbamate that undergoes an elimination to liberate free doxorubicin

and carbon dioxide. However, in addition to the very slow kinetics suffered by the Staudinger ligation ($k_2 \sim 10^{-3}$ $M^{-1}s^{-1}$), the undesired oxidation of the phosphine acting as the activator and its limited solubility in physiological media were critical limitations for the widespread applicability of this reaction.

In order to overcome such limitations, a modification of the SPAAC reaction between a similar 4-azidobenzyl carbamate intermediate and TCO was described (**Fig. 6-1, b**).[13] The reaction takes places through the formation of a 1,2,3-triazole, which rearranges to the corresponding aldimine, generating molecular nitrogen. Since the reaction is designed to take place under physiological (aqueous) media, the aldimine is then hydrolyzed, leading to the formation of 4-aminobenzyl carbamate that eliminates in the same abovementioned fashion. Although the kinetic rate constant was significantly higher (0.017-0.027 $M^{-1}s^{-1}$), it was not yet ideal to reach significant values in physiological concentrations.

Other modern cross-coupling reactions have also been employed for prodrug activation. Several reports describe the deallylation or depropargylation of caged substrates, catalyzed by Pd, Ru or Au (**Fig. 6-1, c**).[14–17] Similarly, a ring closing metathesis employing the Ru-based Hoveyda-Grubbs II catalyst was also shown to trigger prodrug, although with low overall liberation in biological media (<35%) (**Fig. 6-1, d**).[18] In both cases, the *in vivo* applicability of such strategies is limited due to the metal-related toxicity and the slow kinetics associated with such reactions.

6.2. 1,2,4,5-tetrazines for prodrug activation

6.2.1. Background:

In order to achieve kinetic rate constants suitable for significant release *in vivo*, different strategies were developed involving the iEDDA cycloaddition. The first example, named *click-to-release*, used the reactive strained TCO fused to doxorubicin through a carbamate linker in position 2, and a set of differently substituted tetrazines as the activators (**Fig. 6-2, a**).[19] Upon iEDDA ligation and subsequent N_2 loss, the 3,4-dihydropyridazine intermediate undergoes a 1,4-elimination reaction, releasing the free amine of the cargo and carbon dioxide, and eventually rearomatizing the pyridazine. Other strained dienophiles, such as benzonorbornadiene derivatives,

have also been employed reaching near-quantitative release yields (**Fig. 6-2, b**).[20] The proposed mechanism suggests the generation of an isoindole or isobenzofurane intermediate, ultimately responsible for the liberation of the cargo though elimination of the carbamate linker. In both examples, the elimination of the carbamate linker leads to the liberation of a free amine that is the responsible to exert the function in the active drug. In order to liberate other functional groups, a novel strategy relying on a vinyl ether/tetrazine pair was developed for the release of alcohols.[21] Interestingly, despite the slow 2nd order rate constant for the reaction (~$5 \cdot 10^{-4}$ $M^{-1}s^{-1}$), intracellular liberation of a fluorogenic coumarin was observed after 4 hours incubation at micromolar concentrations.[21]

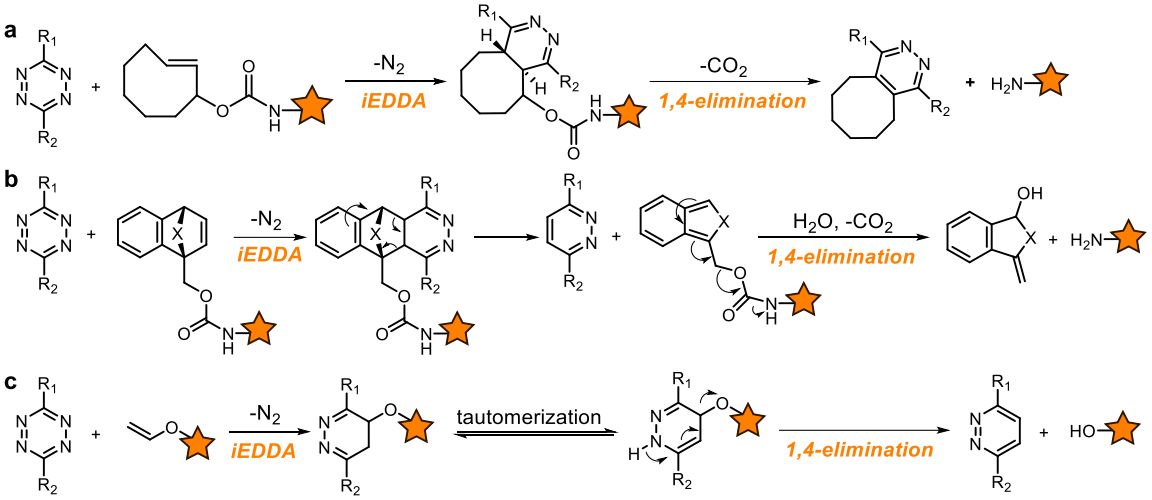


Figure 6-2: Release reactions based on the iEDDA cycloaddition between tetrazines and different dienophiles. a) With TCO. **b)** With benzonorbornadiene. **c)** With vinyl ethers.

Out of these strained dienophiles and tetrazines pairs, the most widely applied is the one involving the TCO/tetrazine pair, probably due to the exceptional kinetics it provides. Interestingly, is has not only been employed for *in vivo* prodrug liberation,[22,23] but also to restore protein activity through the unmasking of its catalytic residues.[24–26] The chemical parameters of the participating functional groups governing this reaction have been studied in depth, both computationally[27] and experimentally.[28] Interestingly, while the presence of EWG in the tetrazine ring increase the kinetics of the initial iEDDA cycloaddition, they are detrimental for the release of the cargo (**Fig. 6-3, a**). These substituents influence the different regioisomers formed upon iEDDA cycloaddition, ultimately responsible for the successful release of the cargo or the generation of a non-releasable undesired by-

product (**Fig. 6-3, b**). Therefore, a compromise between the substituents in the activator tetrazine needs to be met in order to achieve significant release in suitable kinetics values to take place under *in vivo* conditions.

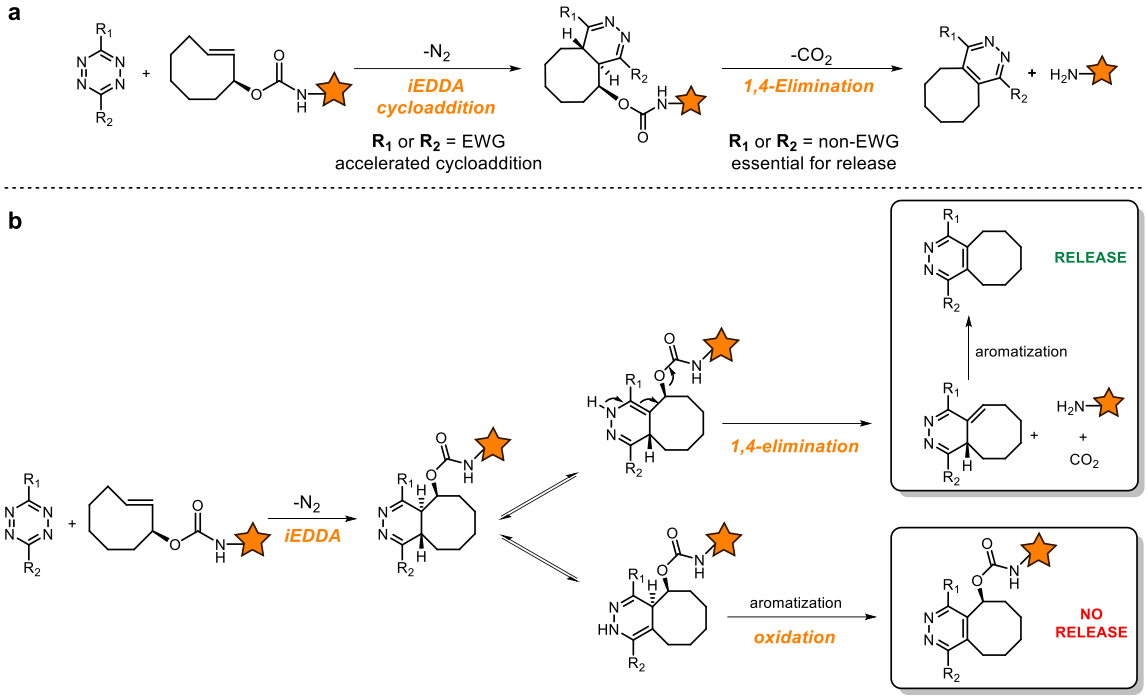
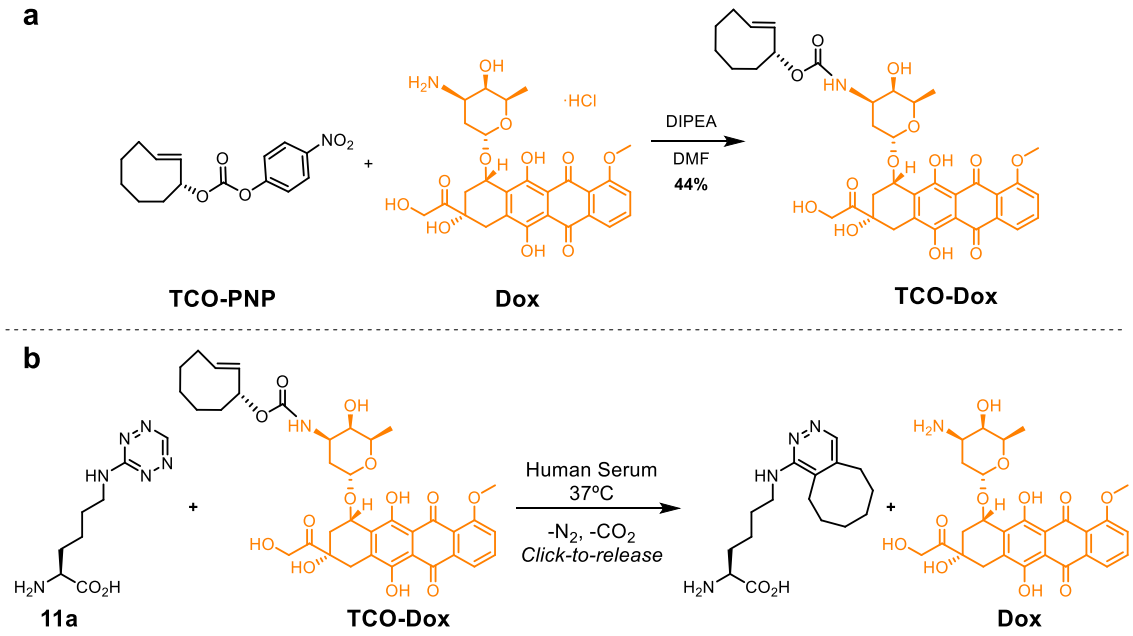


Figure 6-3: Mechanism of the *click-to-release* reaction between the TCO/tetrazine pair. **a)** The electronic behavior of the substituents on the activator tetrazine are critical for the release of the cargo**. b)** There are 2 possible reaction mechanisms, one leading to the 1,4-elimination and successful cargo release, and the other one with a premature aromatization that prevents the elimination reaction.

6.2.2. Results:

In light of the chemical features that the tetrazine fragment require to undergo *click-to-release* reactions with strained dienophiles, we envisaged that the resulting proteins labeled with **2** (seen in **Chapter 5**) could be suitable as biorthogonal activators of TCO carbamates. As a model to qualitatively and quantitatively determine the liberation of a given cargo, we employed lysine derivative **11a** in reaction with a TCO-doxorubicin carbamate (**TCO-Dox**; **Scheme 6-1, b**). **TCO-Dox** was synthesized following a previously reported method,[19] employing an activated derivative of the axial isomer of *trans*-cyclooct-2-en-1-ol in the form of a *p*-nitrophenyl carbonate (**TCO-PNP**; **Scheme 6-1, a**).



Scheme 6-1: Evaluation of the *click-to-release* reaction using lysine-derived tetrazines. a) Synthesis of **TCO-Dox** from an activated trans-cyclooct-2-en-1-ol precursor. **b)** *Click-to-release* reaction between **11a** and **TCO-Dox** *in vitro* under physiological conditions.

The *in vitro* release of doxorubicin under physiological conditions (in human serum at 37 ºC) was determined by Ultra-performance liquid chromatography (UPLC). The amount of liberated free doxorubicin was determined using the area under the curve (AUC) of its specific peak at different time-points, and corrected by a factor of stability (see **Experimental Section 6.3.2.** for more details). Under these conditions, we determined the maximum liberation of approximately 50% of the total amount of doxorubicin for the monosubstituted tetrazine **11a**, achieved after ca. 15 hours of reaction (**Figure 6-4, left**). Importantly, the slower 2nd order rate constant k_2 determined in **Chapter 3** for the cycloaddition reaction of methyltetrazinyl-lysine analogue **10a** (0.04 $M^{-1}s^{-1}$, versus the 1.33 $M^{-1}s^{-1}$ for **11a**), also affected the release percentage of doxorubicin, which was considerably slower (**Figure 6-4, right**).

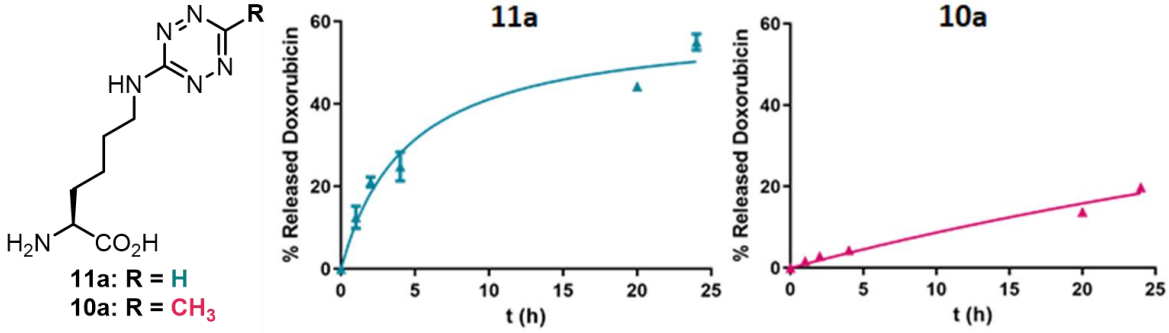


Figure 6-4: Release percentage of the reaction between TCO-Dox and 11a or 10a.

To showcase the potential of the labeled proteins with reagent **2** to function as targeted bioorthogonal activators for drug delivery, we employed the monoclonal antibody Trastuzumab, a frontline in HER2⁺ breast cancer treatment.[29] First, Trastuzumab was conveniently labeled using 50 equivalents of **2** at pH 7.2 to afford Trastuzumab-**2**, which was characterized by LC-MS (**Figure 6-5, a**; **Experimental Section 6.3.3.**). Then, BT474 HER2⁺ cells were incubated with the labeled antibody at 250 nM, followed by the addition of **TCO-Dox** to determine if the released doxorubicin could exert a cytotoxic effect at 72 h post-administration (**Figure 6-5, b**, **Experimental Section 6.3.4.**).

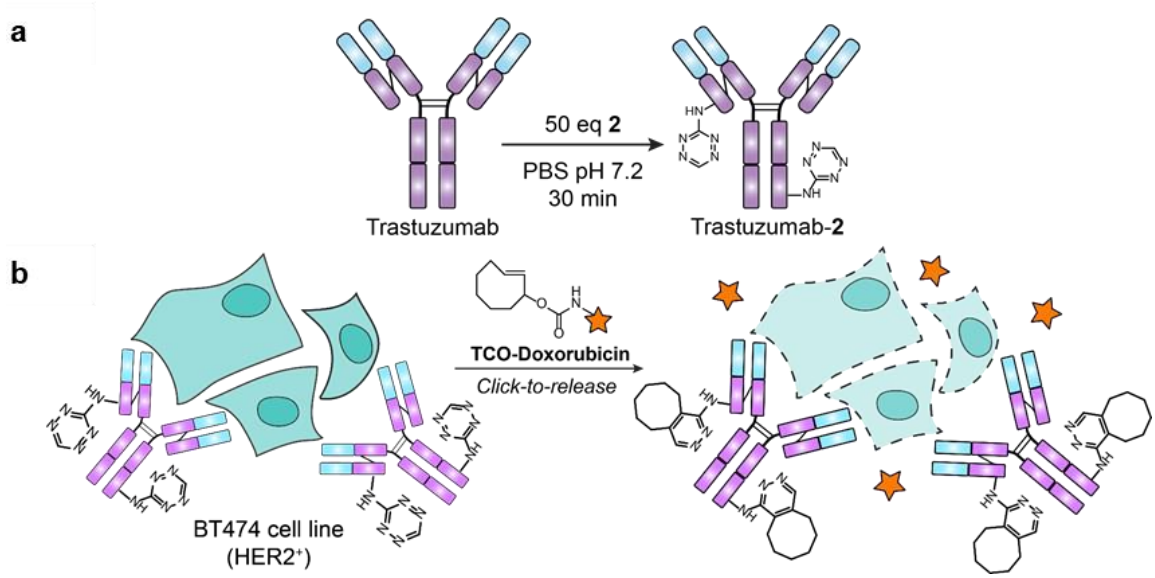


Figure 6-5: Application of Trastuzumab-2 in the activation of TCO-Dox in cell culture. a) Labeling protocol of Trastuzumab with 50 equivalents of **2** in PBS pH 7.2 to generate the modified Trastuzumab-**2. b)** Incubation of BT474 cells (HER2⁺) with Trastuzumab-**2** to bioorthogonally liberate doxorubicin from **TCO-Dox**.

First, under these experimental conditions, the cytotoxic concentration 50 (CC_{50}) of the active drug doxorubicin was determined to be 29.7±10.4 nM (**Fig. 6-6, a**). As a control, we demonstrated that at the used concentration (250 nM) there were no differences in cell growth for wild-type Trastuzumab, Trastuzumab-**2** or Trastuzumab with covalently linked doxorubicin (Trastuzumab-**Dox**; **Fig. 6-6, b**). Most importantly, when incubating at different concentrations of **TCO-Dox**, a significant decrease in the CC_{50} value when co-administered with Trastuzumab-**2** compared to the cells with unlabeled Trastuzumab was obtained, thus demonstrating the release of doxorubicin and its increased cytotoxic effect (**Figure 6-6, c**).

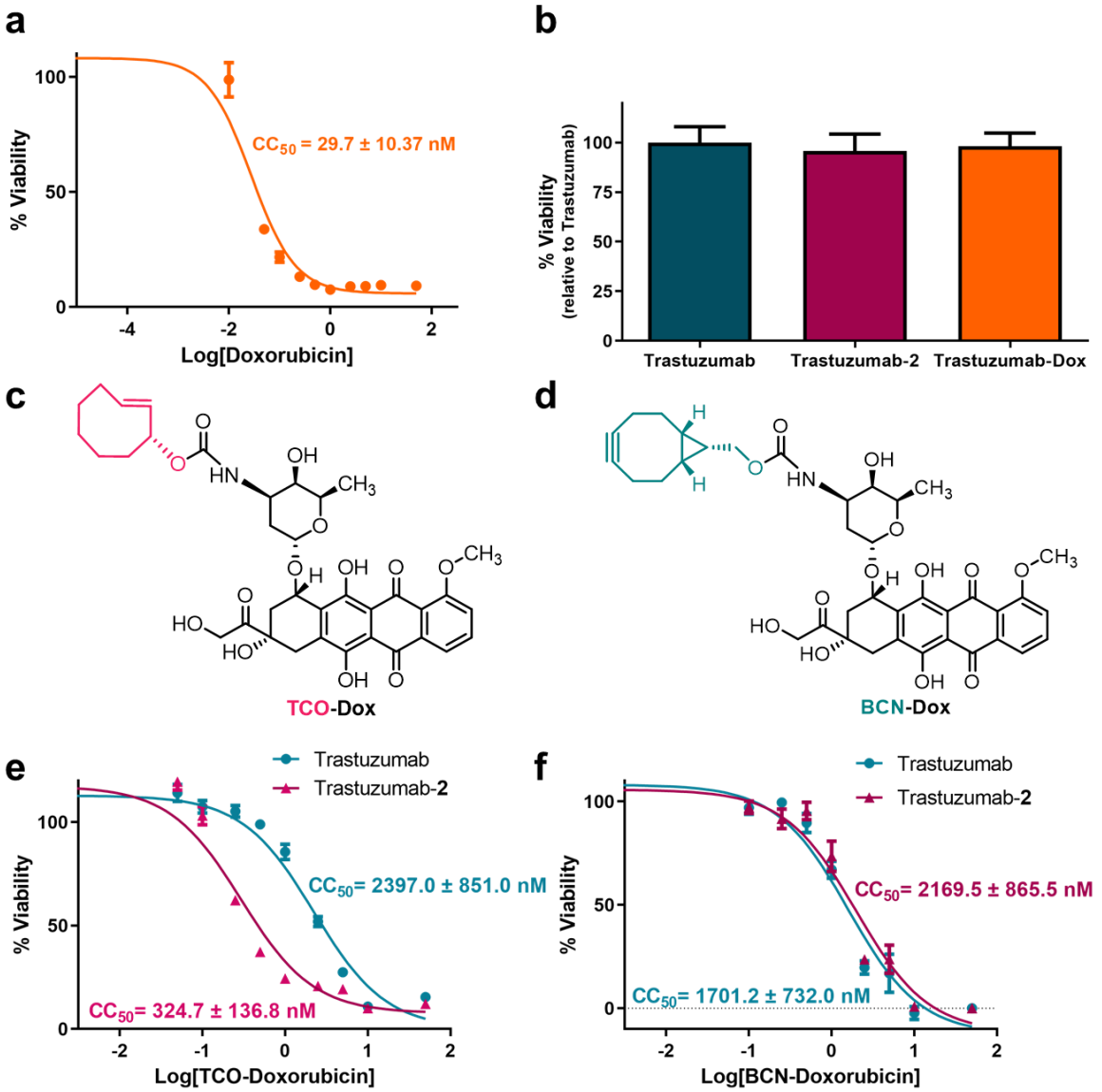


Figure 6-6: ***Click-to-release*** **between Trastuzumab labeled with 2 and TCO-Dox in BT474 HER2⁺ cell culture. a)** Cell viability curve at increasing concentrations of doxorubicin (with 250 nM Trastuzumab in the cell media). **b)** Comparison of the proliferative effects of Trastuzumab, Trastuzumab-**2** and Trastuzumab-**Dox** at 250 nM concentration (relative to incubation with wild-type Trastuzumab). **c)** Cell viability curve at increasing concentrations of **TCO-Dox** incubated with unlabeled Trastuzumab (blue line) or Trastuzumab-**2** (purple line). **d)** Cell viability curve at increasing concentrations of **BCN-Dox** incubated with unlabeled Trastuzumab (blue line) or Trastuzumab-**2** (purple line). **Note:** Cell viability was determined using WST-1 cell proliferation assay (Roche).

Moreover, to demonstrate that the *click-to-release* reaction was indeed taking place to activate the drug, a non-releasable version of **TCO-Dox** (using **BCN-Dox** instead, unable to undergo *click-to-release* reactions) was employed under the same experimental conditions. In this context, no significant differences between

incubating with unlabeled Trastuzumab or Trastuzumab-**2** were observed (**Figure 6-6, d**). In summary, the near 10-fold increase in toxicity obtained when incubating **TCO-Dox** with Trastuzumab-**2** compared to unlabeled, wild-type Trastuzumab clearly demonstrates the potential of this strategy to attain specific liberation of active drugs at the site of action.

6.3. Summary:

We have demonstrated that the formed tetrazinyl-lysines (seen in **Chapter 5**) can undergo *click-to-release* reactions with suitable *trans*-cyclooctene carbamates precursors to liberate cargos of choice. In particular, we have used tetrazinyl-lysine **11a** as a model to determine the reaction parameters with **TCO-Dox**, a doxorubicin derivative acting as a prodrug. In physiological conditions (human serum at 37 ºC), ca. 50% liberation was accomplished after 15 h post-reaction.

To showcase a potential application of these modified proteins with tetrazines to trigger *click-to-release* reactions *in vivo*, we have labeled the therapeutically relevant monoclonal antibody Trastuzumab and performed a bioorthogonal prodrug activation experiment in cell culture. We envisage that our approach can lead to the development of targeted drug delivery methods combining the antigen-specific behavior of certain proteins with the ability of tetrazines to trigger bioorthogonal activation of prodrugs.

6.4. Bibliography:

1 W. Sneader. The discovery of aspirin: a reappraisal. *BMJ*. 2000. **321**. 1591–1594.

2 V. J. Stella, W. N. A. Charman and V. H. Naringrekar. Prodrugs. Do they have advantages in clinical practice? *Drugs*. 1985. **29**. 455–473.

3 K.-M. Wu. A New Classification of Prodrugs: Regulatory Perspectives. *Pharmaceuticals*. 2009. **2**. 77–81.

4 B. Testa and J. M. Mayer. in *Hydrolysis in Drug and Prodrug Metabolism*, Verlag Helvetica Chimica Acta, Zürich, 2006, pp. 419–534.

5 M. H. Lee, J. L. Sessler and J. S. Kim. Disulfide-Based Multifunctional Conjugates for Targeted Theranostic Drug Delivery. *Acc. Chem. Res.* 2015. **48**. 2935–2946.

6 J. J. Yeh and W. Y. Kim. Targeting Tumor Hypoxia With Hypoxia-Activated Prodrugs. *J. Clin. Oncol.* 2015. **33**. 1505–1508.

7 X. Peng and V. Gandhi. ROS-activated anticancer prodrugs: a new strategy for tumor-specific damage. *Ther. Deliv.* 2012. **3**. 823–833.

8 M. Rooseboom, J. N. M. Commandeur and N. P. E. Vermeulen. Enzyme-Catalyzed Activation of Anticancer Prodrugs. *Pharmacol. Rev.* 2004. **56**. 53–102.

9 X. Ji, Z. Pan, B. Yu, L. K. De La Cruz, Y. Zheng, B. Ke and B. Wang. Click and release: Bioorthogonal approaches to 'on-demand' activation of prodrugs. *Chem. Soc. Rev.* 2019. **48**. 1077–1094.

10 M. Azoulay, G. Tuffin, W. Sallem and J.-C. Florent. A new drug-release method using the Staudinger ligation. *Bioorg. Med. Chem. Lett.* 2006. **16**. 3147–3149.

11 R. van Brakel, R. C. M. Vulders, R. J. Bokdam, H. Grüll and M. S. Robillard. A Doxorubicin Prodrug Activated by the Staudinger Reaction. *Bioconjug. Chem.* 2008. **19**. 714–718.

12 O. Tacar, P. Sriamornsak and C. R. Dass. Doxorubicin: an update on anticancer molecular action, toxicity and novel drug delivery systems. *J. Pharm. Pharmacol.* 2013. **65**. 157–170.

13 S. S. Matikonda, D. L. Orsi, V. Staudacher, I. A. Jenkins, F. Fiedler, J. Chen and A. B. Gamble. Bioorthogonal prodrug activation driven by a strain-promoted 1,3-dipolar cycloaddition. *Chem. Sci.* 2015. **6**. 1212–1218.

14 P. K. Sasmal, C. N. Streu and E. Meggers. Metal complex catalysis in living biological systems. *Chem. Commun.* 2013. **49**. 1581–1587.

15 T. Völker, F. Dempwolff, P. L. Graumann and E. Meggers. Progress towards bioorthogonal catalysis with organometallic compounds. *Angew. Chem. Int. Ed. Engl.* 2014. **53**. 10536–10540.

16 C. Streu and E. Meggers. Ruthenium-induced allylcarbamate cleavage in living cells. *Angew. Chem., Int. Ed.* 2006. **45**. 5645–5648.

17 R. M. Yusop, A. Unciti-Broceta, E. M. V. Johansson, R. M. Sánchez-Martín and M. Bradley. Palladium-mediated intracellular chemistry. *Nat. Chem.* 2011. **3**. 239–243.

18 V. Sabatino, J. G. Rebelein and T. R. Ward. "Close-to-Release": Spontaneous Bioorthogonal Uncaging Resulting from Ring-Closing Metathesis. *J. Am. Chem. Soc.* 2019. **141**. 17048–17052.

19 R. M. Versteegen, R. Rossin, W. Ten Hoeve, H. M. Janssen and M. S. Robillard. Click to release: Instantaneous doxorubicin elimination upon tetrazine ligation. *Angew. Chem., Int. Ed.* 2013. **52**. 14112–14116.

20 M. Xu, J. Tu and R. M. Franzini. Rapid and efficient tetrazine-induced drug release from

highly stable benzonorbornadiene derivatives. *Chem. Commun.* 2017. **53**. 6271–6274.

21 E. Jiménez-Moreno, Z. Guo, B. L. Oliveira, I. S. Albuquerque, A. Kitowski, A. Guerreiro, O. Boutureira, T. Rodrigues, G. Jiménez-Osés and G. J. L. Bernardes. Vinyl Ether/Tetrazine Pair for the Traceless Release of Alcohols in Cells. *Angew. Chem., Int. Ed.* 2017. **56**. 243–247.

22 R. Rossin, S. M. J. Van Duijnhoven, W. Ten Hoeve, H. M. Janssen, L. H. J. Kleijn, F. J. M. Hoeben, R. M. Versteegen and M. S. Robillard. Triggered Drug Release from an Antibody-Drug Conjugate Using Fast 'click-to-Release' Chemistry in Mice. *Bioconjug. Chem.* 2016. **27**. 1697–1706.

23 R. Rossin, S. M. Van Den Bosch, W. Ten Hoeve, M. Carvelli, R. M. Versteegen, J. Lub and M. S. Robillard. Highly reactive trans-cyclooctene tags with improved stability for diels-alder chemistry in living systems. *Bioconjug. Chem.* 2013. **24**. 1210–1217.

24 J. Li, S. Jia and P. R. Chen. Diels-Alder reaction–triggered bioorthogonal protein decaging in living cells. *Nat. Chem. Biol.* 2014. **10**. 1003–1005.

25 G. Zhang, J. Li, R. Xie, X. Fan, Y. Liu, S. Zheng, Y. Ge and P. R. Chen. Bioorthogonal Chemical Activation of Kinases in Living Systems. *ACS Cent. Sci.* 2016. **2**. 325–331.

26 L. Liu, Y. Liu, G. Zhang, Y. Ge, X. Fan, F. Lin, J. Wang, H. Zheng, X. Xie, X. Zeng and P. R. Chen. Genetically Encoded Chemical Decaging in Living Bacteria. *Biochemistry*. 2018. **57**. 446–450.

27 J. C. T. Carlson, H. Mikula and R. Weissleder. Unraveling Tetrazine-Triggered Bioorthogonal Elimination Enables Chemical Tools for Ultrafast Release and Universal Cleavage. *J. Am. Chem. Soc.* 2018. **140**. 3603–3612.

28 X. Fan, Y. Ge, F. Lin, Y. Yang, G. Zhang, W. S. C. Ngai, Z. Lin, S. Zheng, J. Wang, J. Zhao, J. Li and P. R. Chen. Optimized Tetrazine Derivatives for Rapid Bioorthogonal Decaging in Living Cells. *Angew. Chem., Int. Ed.* 2016. **55**. 14046–14050.

29 C. A. Hudis. Trastuzumab — Mechanism of Action and Use in Clinical Practice. *N. Engl. J. Med.* 2007. **357**. 39–51.

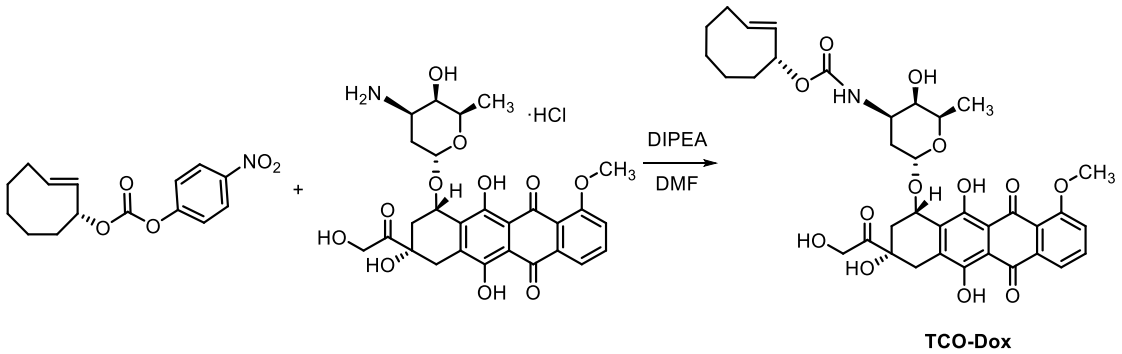
30 B. A. Badeau, M. P. Comerford, C. K. Arakawa, J. A. Shadish and C. A. Deforest. Engineered modular biomaterial logic gates for environmentally triggered therapeutic delivery. *Nat. Chem.* 2018. **10**. 251–258.

31 N. G. Yabbarov, G. A. Posypanova, E. A. Vorontsov, O. N. Popova and E. S. Severin. Targeted delivery of doxorubicin: Drug delivery system based on PAMAM dendrimers. *Biochem.* 2013. **78**. 884–894.

6.5. Experimental Section

6.5.1. Organic syntheses of doxorubicin analogues:

(*trans*-N-cyclooct-2-enyl)-doxorubicin carbamate (TCO-Dox)



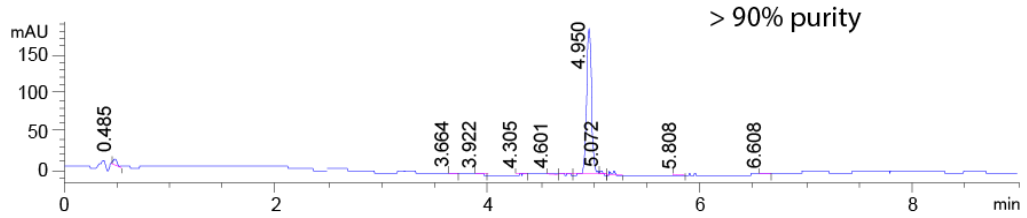
TCO-Dox

The previously reported synthetic procedure was followed.[19] Briefly, a solution of doxorubicin hydrochloride (Fluorochem; 87.6 mg, 0.151 mmol, 1.1 eq) and DIPEA (26 µL, 0.151 mmol, 1.1 eq) in DMF (3 mL) was added dropwise to the commercially available activated *trans*-cyclooctene (TCO), axial isomer, PNP ester (Sichem; 40 mg, 0.137 mmol, 1 eq), and the reaction mixture was stirred 24 h in the dark, at which point TLC showed completion. The reaction mixture was concentrated by rotatory evaporation and the solvent fully removed by high vacuum. The resulting residue was purified by flash chromatography, using 0-15% MeOH in DCM as the mobile phase. The relevant fractions were combined and the solvent evaporated under reduced pressure. The resulting residue was resuspended in Et_2O and the formed solid was filtered over a C-type crucible filter to afford 41.9 mg (44% yield) of the desired product as a red solid.

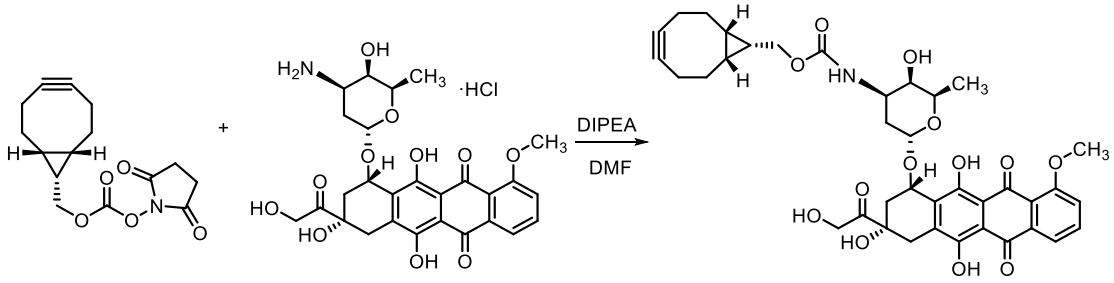
The obtained ^1H NMR spectrum is consistent with the axial isomer.

^1H NMR (400 MHz, $CDCl_3$) δ 13.99 (s, 1H), 13.27 (s, 1H), 8.04 (m, 1H), 7.79 (t, *J* = 8.1 Hz, 1H), 7.40 (d, *J* = 8.5 Hz, 1H), 5.75 (s, 1H), 5.50 (d, *J* = 13.2 Hz, 2H), 5.33 – 5.23 (m, 3H), 5.07 (s, 1H), 4.76 (d, *J* = 4.7 Hz, 3H), 4.53 (s, 1H), 4.15 (d, *J* = 6.6 Hz, 2H), 4.09 (s, 5H), 3.87 (s, 2H), 3.69 (d, *J* = 7.4 Hz, 2H), 3.29 (d, *J* = 18.9 Hz, 2H), 3.05 (d, *J* = 19.0 Hz, 2H), 2.98 (s, 0H), 2.43 (s, 1H), 2.34 (d, *J* = 14.7 Hz, 2H), 2.18 (d, *J* = 14.7 Hz, 2H), 2.09 – 1.83 (m, 4H), 1.61 (s, 4H), 1.42 (s, 0H), 1.33 – 1.24 (m, 10H), 1.03 (s, 1H), 0.84 (s, 4H)

UPLC:



N-((1R,8S,9S)-bicyclo[6.1.0]non-4-yn-9-yl)-doxorubicin carbamate (BCN-Dox)



BCN-Dox

Following a previously reported methodology,[30] a solution of doxorubicin hydrochloride (Fluorochem; 29.6 mg, 0.051 mmol, 1.5 eq) and DIPEA (12 µL, 0.068 mmol, 2 eq) in DMF (0.5 mL) was added dropwise to the commercially available BCN-succinimidyl carbonate (Merck; 10 mg, 0.034 mmol, 1 eq) in a degassed vial. The reaction mixture was stirred overnight, at which point TLC showed completion. The reaction mixture was then concentrated by rotatory evaporation and the solvent fully removed by high vacuum. The resulting residue was purified by flash chromatography, using 0-15% MeOH in DCM as the mobile phase. The relevant fractions were combined and the solvent evaporated under reduced pressure. The resulting residue was resuspended in Et_2O and the formed solid was filtered over a C-type crucible filter to afford 14.2 mg (59% yield) of the desired product as a red solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 13.99 (s, 1H), 13.25 (s, 1H), 8.05 (dd, *J* = 7.7, 1.1 Hz, 1H), 7.79 (dd, *J* = 8.5, 7.7 Hz, 1H), 7.40 (dd, *J* = 8.6, 1.1 Hz, 1H), 5.51 (d, *J* = 4.0 Hz, 1H), 5.30 (s, 1H), 5.03 (d, *J* = 8.6 Hz, 1H), 4.76 (s, 2H), 4.54 (s, 1H), 4.14 (d, *J* = 6.6 Hz, 1H), 4.10 (s, 2H), 4.09 (s, 3H), 3.85 (s, 1H), 3.67 (s, 1H), 3.35 – 3.20 (d, *J* = 18.9 Hz, 1H), 3.04 (d, *J* = 18.9 Hz, 1H), 3.00 (s, 1H), 2.39 – 2.31 (m, 1H), 2.27 – 2.13 (m, 4H), 1.88 (dd, *J* = 13.2, 5.3 Hz, 2H), 1.77 (td, *J* = 13.2, 4.2 Hz, 1H), 1.53 (d, *J* = 15.6 Hz, 2H), 1.30 (d, *J* = 6.5 Hz, 4H), 1.25 (s, 1H), 0.97 – 0.81 (m, 3H) ppm

^{13}C NMR (126 MHz, $CDCl_3$) δ 213.98, 187.34, 186.90, 161.24, 156.34, 155.96, 155.84, 135.97, 135.69, 133.75, 133.68, 121.06, 120.05, 118.61, 111.78, 111.61, 100.90, 98.94, 76.78, 69.82, 67.42, 65.69, 63.02, 56.85, 47.02, 35.82, 34.19, 30.42, 29.18, 29.14, 21.55, 20.24, 17.84, 17.01 ppm

HRMS (ESI$^+$) calculated for $C_{38}H_{41}NO_{13}Na$ $[M+Na]^+$: 742.24701 / found 742.24868

IR U_{max}: 3416 (broad band), 2914, 2847, 1716, 1616, 1577, 1508, 1409, 1282, 1204, 1116, 1069, 1009 cm^{-1}

6.5.2. Doxorubicin release from **TCO-Dox**:

In triplicates, 825 µL of Human Serum (Sigma-Aldrich) was mixed with 100 µL Hank's Balanced Salt Solution (10% final concentration), 50 µL **TCO-Dox** 10 mM stock solution in DMSO (final concentration of 500 µM) and the relevant volumes of **10a** or **11a** at 100 mM stock solution in water to add an excess of 5 equivalents. Water was added when necessary to reach 1000 µL final sample volume. Samples containing only **TCO-Dox** or free doxorubicin were prepared in the same fashion and used as controls.

At each desired time-point, 25 µL samples were taken, mixed with 100 µL cold methanol to precipitate serum proteins and centrifuged 20 mins at 14000x g. Then, 40 µL of the supernatant were taken and 2 µL of the solution injected to UPLC (Waters Acquity equipped with Acquity photodiode array detector, flux rate 0.610 ml/min, Acquity UPLC BEH C18 Column, 130 Å, 1.7 µm, 2.1 mm x 100 mm), using solvent A=0.045% trifluoroacetic acid in water, and B=0.036% trifluoroacetic acid in acetonitrile. The amount of **TCO-Dox** and liberated doxorubicin present in each sample was quantified by the AUC of each characteristic peak at 233 nm (absorbance maximum) using a calibration curve of doxorubicin at different concentrations. Values were corrected by a factor dependent on stability of doxorubicin at each time-point.

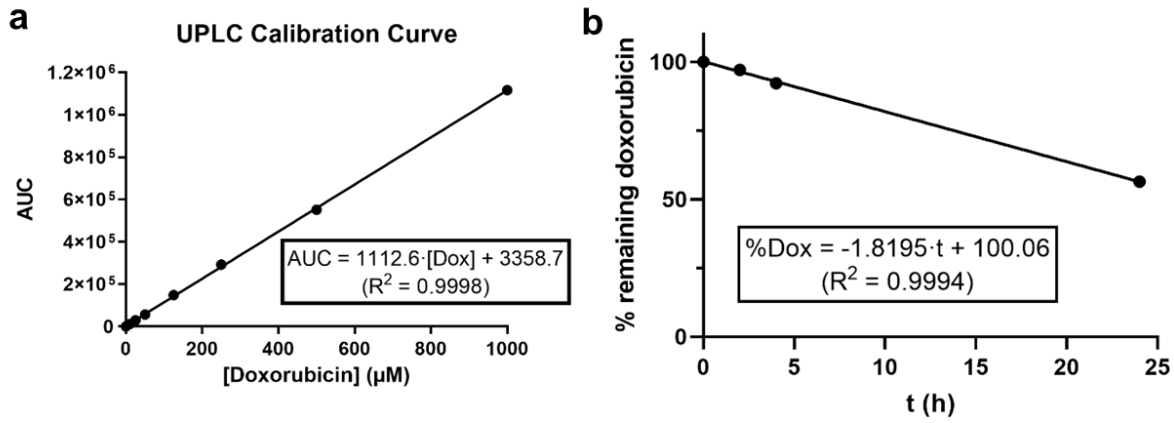


Figure 6-7: UPLC quantification of doxorubicin. a) Calibration curve of free doxorubicin at the experimental conditions, correlating the AUC with specific concentration values. **b)** Stability of doxorubicin in serum at 37 ºC obtained by the AUC values.

The percentage of release of doxorubicin from **TCO-Dox** with 5 equivalents of compounds **10a** or **11a** in physiological conditions was quantified and plotted in GraphPad Prism 8. The data for each compound was fitted to a non-linear regression curve, from which the liberation half-life ($t_{1/2}$), maximum percentage of doxorubicin liberated and error values were extracted.

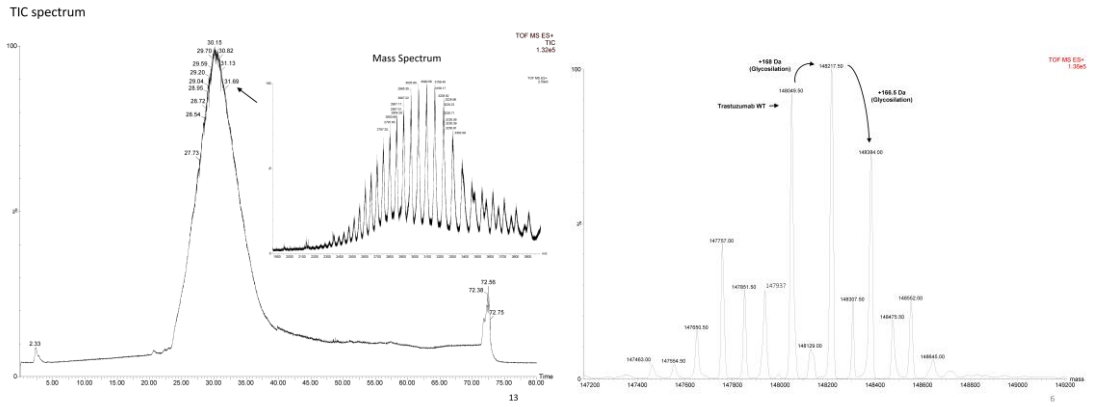
6.5.3. Trastuzumab labeling

6.5.3.1. Trastuzumab labeling with **2** (Trastuzumab-**2**):

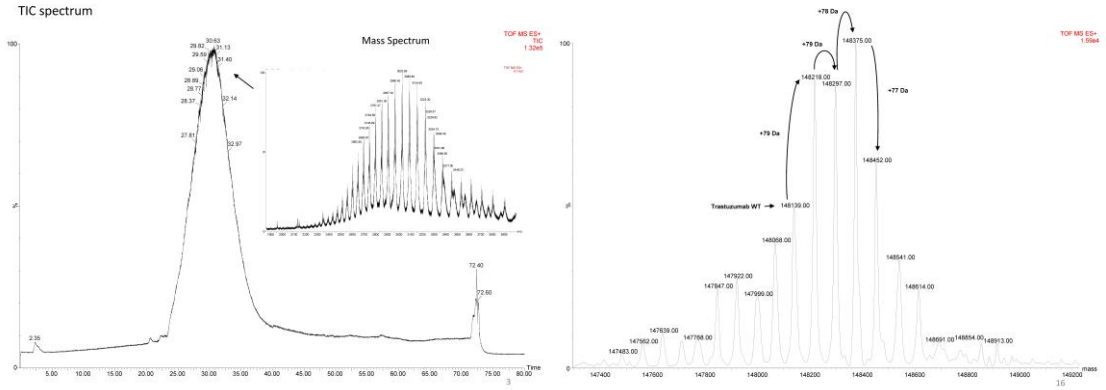
3.4 mg **2** were dissolved in 1 mL H_2O (to reach a concentration of 21.12 mM), and 3.2 µL of the resulting solution were added to a solution of 10 µL Trastuzumab (Carbosynth) 20 mg/mL in 190 µL PBS buffer pH 7.2 (50 eq). Reactions were left at room temperature for 30 min with end-to-cap mixing, followed by buffer exchange using Amicon® Ultra-0.5 mL centrifugal filters 100000 MWCO (Millipore) diluting in an Ammonium Acetate 150 mM solution to a final volume of 20 µL ([Trastuzumab-**2**]$_f$= 100 µM). Unlabeled Trastuzumab was treated the same way and used as control.

Intact protein analysis of samples before and after reaction with 50 equivalents of **2** at pH 7.2 were performed (following protocol from **Chapter 5, Section 5.2.3.**). LC-MS chromatograms with inset of the full MS spectra of the major chromatographic peak and deconvoluted MS spectra are provided from left to right.

Trastuzumab:



Trastuzumab-2:



6.5.3.2. Trastuzumab labeling with doxorubicin (Trastuzumab-**Dox**):

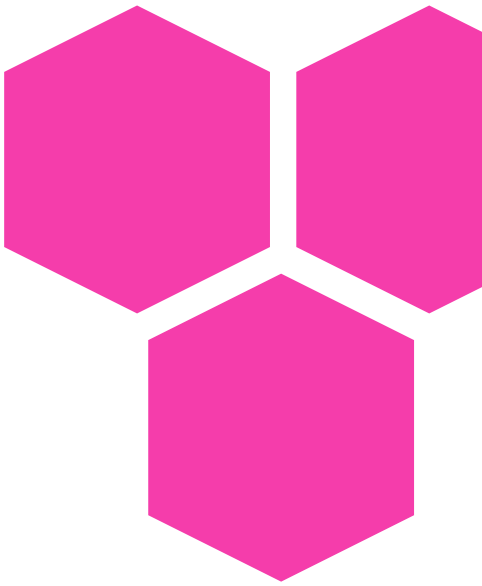
As a control, Trastuzumab was covalently labeled with doxorubicin (Trastuzumab-**Dox**). Briefly, 2.1 mg doxorubicin hydrochloride (Fluorochem) were dissolved in DMSO to reach a concentration of 50 mM, and 4.5 µL of the stock solution were added to 150 µL Trastuzumab at 50 µM, followed by the addition of EDC hydrochloride (Fluorochem) at a final concentration of 3 mM and Sulfo-NHS (Merck Millipore) at a final concentration of 5 mM. The reaction mixture was gently shaken for 2 hours at room temperature, followed by the purification of the product using PD Minitrap G-25 columns (GE Healthcare). After concentration using Amicon® Ultra-0.5 mL centrifugal filters 100000 MWCO, doxorubicin content in Trastuzumab-**Dox** was calculated by the ratio of protein absorption at 230 nm ($\varepsilon_{Doxorubicin}$ = 215380 $M^{-1}cm^{-1}$

at 280 nm) and doxorubicin absorption at 481 nm ($\varepsilon_{Doxorubicin}$ = 10410 $M^{-1}cm^{-1}$ at 481 nm), and determined of 0.82[31].

6.5.4. CC_{50} determination in HER2$^+$ breast cancer cell culture:

BT474 HER2$^+$ cells were maintained in a humidified cell culture incubator at 37 ºC with 5% CO_2 using Dulbecco's Modified Eagle Medium (DMEM, ThermoFisher Scientific) supplemented with 10% heat-inactivated Fetal Bovine Serum in the presence of Penicillin and Streptomycin. Cells were plated at a concentration of 10000 cells/well in a flat-bottomed 96-well plate (SPL Life Sciences), and after 24 h incubation the media was removed from each well and 100 µL of Trastuzumab or Trastuzumab-**2** 0.5 µM in fresh media were added. In triplicates, after 1 hour, serial dilutions (0, 0.05, 0.1, 0.25, 0.5, 1, 2.5, 5, 10 µM) of free doxorubicin, **TCO-Dox** or **BCN-Dox** were added and incubated for 72 hours. Then, the media was removed, each well washed with PBS once, and 100 µL of fresh media with 10% WST-1 Cell Proliferation Reagent (Roche) were added. After 75 min incubation time, absorbances at 450 nm were measured, using absorbances at 630 nm as reference wavelengths following manufacturer's protocol. As a control, Trastuzumab, Trastuzumab-**2** and Trastuzumab-**Dox** were incubated at the used experimental concentration (250 nM of the specified antibody in each case).

The absorbance values were transformed into percentages of viability (% viability), considering the incubation with 0 nM doxorubicin, **TCO-Dox** or **BCN-Dox** as the 100% viability value. The % viability values were plotted against the $\log_{10}$[**Dox**], $\log_{10}$[**TCO-Dox**] or $\log_{10}$[**BCN-Dox**] in GraphPad Prism 8. The data for each compound was fitted to a non-linear dose-response regression curve, from which the liberation cytotoxic concentrations 50 (CC_{50}) and the 95% confidence intervals were extracted.



Chapter 7

Introduction to genetic code reassignment

7.1. The manual of life:

Instead of exploiting the reactivity of certain amino acid side chains, an alternative strategy for protein labeling is the co-translational incorporation of unnatural amino acids directly in the protein sequence. This can be achieved engineering the genetic code of certain organisms through the manipulation of their translation machinery to biosynthesize proteins with an increased chemical content.

The genetic code is the set of rules followed by all organisms to transform the genetic information encoded in nucleic acids into amino acids that, when linked together through peptide bonds, generate proteins (**Fig. 7-1**). In this process, called translation, the information previously transcribed from DNA to RNA is finally "translated" into proteins, conforming what is commonly known as the central dogma of molecular biology.[1]

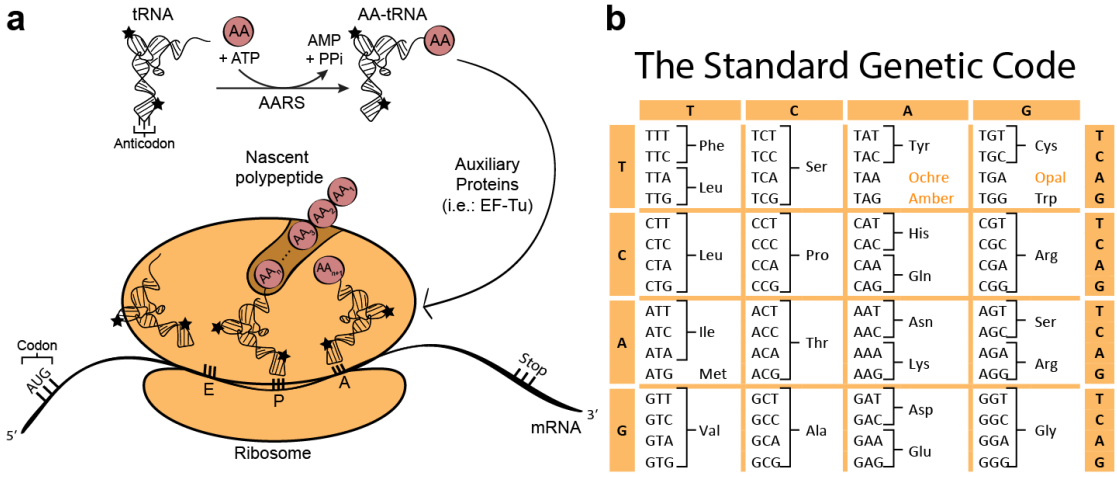


Figure 7-1: Translation process. a) The information encoded in the messenger RNA (mRNA) is translated into proteins in the ribosome, through the recognition of nucleotide triplets (codons) by the anticodons of the aminoacylated transfer RNAs (tRNAs). Aminoacylated tRNAs (those being the covalently linked tRNAs with their cognate amino acids by ester bonding) are formed by specific aminoacyl-tRNA synthetase (AARS) in an ATP-dependent process. **b)** The genetic code determines which amino acid will be translated by each codon. Out of the possible 64 combinations of nucleotide triplets, 61 encode for the 20 proteinogenic amino acids, and 3 determine translation termination. Since there are more codons than amino acids, the code is redundant (or degenerate).

Fully unveiled in the mid-1960s,[2] the genetic code was initially deemed as a "frozen accident" (term coined by Francis Crick): *frozen* because, once every organism evolved from a common ancestor with that particular set of instructions, any variations in the code would have critical outcomes in the specific organism and lead

to its death; and *accident* because it seemed to have evolved in a particular way by serendipity more than by any other compelling reason, with no obvious correlations between each amino acid and its cognate codon(s).

However, in the following decades, certain codon-amino acid associations became apparent.[3,4] For instance, the first letter of the codon can be generally linked to the biosynthetic precursor of the amino acid. That is the case for the amino acids formed from pyruvate, which share the first letter in the codon: thymine, T. The second letter is connected to the polarity of the amino acid. And the third letter is related to the redundancy of the code, as in most cases it is information-free (up to eight amino acids have a fourfold degeneracy).[3,5]

These physicochemical correlations, and the degeneracy itself, makes the genetic code outstandingly resistant to point mutations: when they occur, the amino acid identity is sometimes preserved, but if changed, a similar amino acid is incorporated in place. In fact, compared to at least one million of other randomly generated genetic codes, the existing one is the best in mitigating potential mutations.[6,7]

Therefore, the genetic code is universal and robust, and provides the set of rules for the translation of genes, which is an equally robust and tightly regulated process. Therefore, the manipulation of the translation apparatus is challenging and displays several technical constraints that limit the modulations that can be imposed in the system. Indeed, the *frozen accident* is a demonstration that evolution itself could not expand the amino acid identities beyond the existing ones, thereby stalling the growth of the genetic code. This was probably due to the saturation of recognition signals in the tRNA structure, which effectively blocked the emergence of new tRNA identities.[8]

Nevertheless, an artificial expansion can be achieved taking advantage of some of the properties of the genetic code. In particular, the degeneracy offers the possibility to reassign some redundant codons with new identities, different from the ones determined in the genetic code, without compromising the viability of the organism. When these non-canonical identities are generated, the incorporation of non-canonical amino acids (ncAAs) is achieved and the genetic code is successfully expanded.

7.2. Expanding Nature's chemical repertoire:

Genetic code expansion (GCE) can be achieved by changing the identity of a codon that already encodes for a given amino acid (*sense codon reassignment*), using one of the three naturally occurring stop codons as an "open" codon to encode for the chosen ncAA (*stop codon suppression*), or generating new previously non-existent codons (*codon creation*),[9,10] for instance through quadruplet codons or with new nucleobases.[11–14] In this Ph.D. Thesis we focalized in methods for stop codon suppression, as they offer the possibility of site-selective incorporation of ncAA into synthetic transcripts to expand protein function.

There are 3 stop codons: amber (TAG), ochre (TAA) and opal or umber (TGA). Out of them, the more frequently reassigned is the amber for two main reasons: firstly, it is the rarest stop codon;[15] and secondly, in *E. coli* the amber codon is only recognized by the release factor 1 (RF1) to terminate translation. For stop codon suppression (or, in this particular case, *amber codon suppression*) the artificial ncAA incorporation machinery needs to outcompete the termination machinery. Hence, recognition by one release factor only, compared for instance with the ochre stop codon, which is recognized by RF1 and RF2, becomes a positive factor. For this reason, certain *E. coli* strains have been engineered with a depleted RF1, or even with all amber codons mutated to the other stop codons.[10,16]

In order to reassign the amber stop codon to encode for a chosen ncAA, a dedicated translational machinery needs to be expressed in the host organism. This orthogonal translation system (OTS) is normally composed by a tRNA with a modified anticodon to decode the reassigned codon (for amber suppression, a $tRNA_{CUA}$) and an engineered AARS that specifically aminoacylates this tRNA with the ncAA of interest (**Fig. 7-2, a**).[17] Ideally, the orthogonal AARS should be unreactive towards natural amino acids and endogenous tRNAs, the tRNA should only be a substrate for the orthologous AARS, and the ncAA should not be recognized by the host's AARS. If these requirements are fulfilled, the OTS can be considered truly orthogonal (**Fig. 7-2, b**).[17]

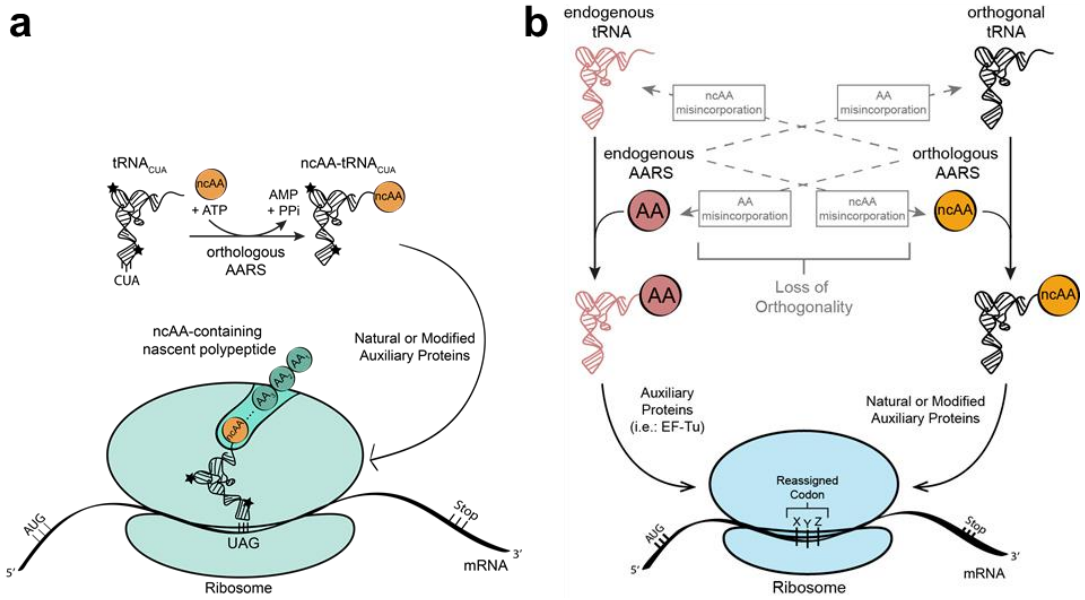


Figure 7-2: Function of an OTS. a) The chosen ncAA is aminoacylated by the orthologous AARS to its cognate $tRNA_{CUA}$, and the aminoacylated tRNA translocates to the ribosome to participate in translation and incorporate the ncAA is response to the in-frame amber codon in the protein sequence of the mRNA. **b)** The orthogonality of the system is determined by the lack of cross-reactivity between the natural components of the translation machinery of the host organism and the artificially expressed OTS. Deliberate misincorporations could have a critical effect in the organism.

Most often, orthogonal AARS/tRNA pairs are derived from species phylogenetically distant to the host, and can be engineered to accept a variety of ncAAs.[18] To date, there are several AARS/tRNA pairs developed for GCE purposes. However, by far the most commonly used are the tyrosyl-tRNA synthetase from *Methanocaldoccocus jannaschii* and its cognate tRNA with a modified anticodon (*Mj*TyrRS/$tRNA^{Tyr}_{CUA}$) and the pyrrolysyl-tRNA synthetase from *Methanosarcina barkeri* and its natural suppressor tRNA (*Mb*PylRS/$tRNA^{Pyl}_{CUA}$).[18] During this Ph.D. Thesis, both systems have been used aimed at different host organisms. **Chapter 7** will be dedicated to the *Mj*TyrRS/tRNA pair, while **Chapter 8** will be focused on the *Mb*PylRS/tRNA pair.

7.3. Bibliography:

1 F. Crick. Central Dogma of Molecular Biology. *Nature* 1970, **227**, 561–563.

2 F. H. C. Crick. The origin of the genetic code. *J. Mol. Biol.* 1968, **38**, 367–379.

3 F. J. R. Taylor and D. Coates. The code within the codons. *Biosystems* 1989, **22**, 177–187.

4 E. V. Koonin and A. S. Novozhilov. Origin and evolution of the genetic code: The universal enigma. 2008, 45–62.

5 S. D. Copley, E. Smith and H. J. Morowitz. A mechanism for the association of amino acids with their codons and the origin of the genetic code. *Proc. Natl. Acad. Sci.* 2005, **102**, 4442–4447.

6 S. J. Freeland and L. D. Hurst. The genetic code is one in a million. *J. Mol. Evol.* 1998, **47**, 238–248.

7 D. Haig and L. D. Hurst. A quantitative measure of error minimization in the genetic code. *J. Mol. Evol.* 1991, **33**, 412–417.

8 A. Saint-Léger, C. Bello, P. D. Dans, A. G. Torres, E. M. Novoa, N. Camacho, M. Orozco, F. A. Kondrashov and L. Ribas de Pouplana. Saturation of recognition elements blocks evolution of new tRNA identities. *Sci. Adv.* 2016, **2**, e1501860.

9 J. W. Chin. Expanding and reprogramming the genetic code. *Nature* 2017, **550**, 53–60.

10 M. J. Lajoie, D. Söll and G. M. Church. Overcoming Challenges in Engineering the Genetic Code. *J. Mol. Biol.* 2016, **428**, 1004–1021.

11 D. A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J. M. Foster, I. R. Corrêa and F. E. Romesberg. A semi-synthetic organism with an expanded genetic alphabet. *Nature* 2014, **509**, 385–388.

12 Y. Zhang, J. L. Ptacin, E. C. Fischer, H. R. Aerni, C. E. Caffaro, K. San Jose, A. W. Feldman, C. R. Turner and F. E. Romesberg. A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* 2017, **551**, 644–647.

13 H. Neumann, K. Wang, L. Davis, M. Garcia-Alai and J. W. Chin. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* 2010, **464**, 441–444.

14 S. Hoshika, N. A. Leal, M.-J. Kim, M.-S. Kim, N. B. Karalkar, H.-J. Kim, A. M. Bates, N. E. Watkins, H. A. SantaLucia, A. J. Meyer, S. DasGupta, J. A. Piccirilli, A. D. Ellington, J. SantaLucia, M. M. Georgiadis and S. A. Benner. Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* 2019, **363**, 884–887.

15 M. J. Lajoie, A. J. Rovner, D. B. Goodman, H.-R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church and F. J. Isaacs. Genomically Recoded Organisms Expand Biological Functions. *Science* 2013, **342**, 357–360.

16 T. M. Wannier, A. M. Kunjapur, D. P. Rice, M. J. McDonald, M. M. Desai and G. M. Church. Adaptive evolution of genomically recoded Escherichia coli. *Proc. Natl. Acad. Sci.* 2018, **115**, 3090–3095.

17 P. O'Donoghue, J. Ling, Y.-S. Wang and D. Söll. Upgrading protein synthesis for synthetic biology. *Nat. Chem. Biol.* 2013, **9**, 594–598.

18 A. Dumas, L. Lercher, C. D. Spicer and B. G. Davis. Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci.* 2015, **6**, 50–69.

Chapter 8

Genetic code expansion with *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase / $tRNA^{Tyr}_{CUA}$ pair

8.1. Incorporation of ncAAs in prokaryotic organisms:

Methanocaldococcus jannaschii is a thermophilic methanogenic archaeon initially isolated from an hydrothermal vent in the East Pacific Rise in 1983.[1] Basic studies on the structural features of its translation apparatus hinted that its tyrosyl-tRNA synthetase (TyrRS) could be potentially used as an OTS in *E. coli*, particularly for two main reasons:[2,3] on the one hand, just like all TyrRSs,[4] it does not present an editing domain in its structure. Editing domains are responsible for ensuring a proper AA-tRNA aminoacylation by breaking the incorrect AA-tRNA linkage when misacylation events take place. Lacking an editing domain enables the efficient aminoacylation of ncAAs without its subsequent removal from the charged $tRNA^{Tyr}$.[2] Although this is not a unique feature of *Mj*TyrRS -as much as two-thirds of the AARS species have no editing domain-,[4] it is a useful characteristic for GCE purposes. On the other hand, *Mj*TyrRS does not recognize the sequence on the anticodon loop of its cognate *Mj*$tRNA^{Tyr}$ because it also lacks an anticodon binding domain.[3] As a result, the anticodon sequence of *Mj*$tRNA^{Tyr}$ can be mutated freely to decode any nucleotide triplet, enabling a successful codon reassignment.

Additionally, the cognate *Mj*$tRNA^{Tyr}$ also presents some structural particularities that allow its orthogonal use in *E. coli*. Specifically, unlike bacterial $tRNA^{Tyr}$, that presents a long variable arm and the G1:C72 pair, archaeal $tRNA^{Tyr}$ contains a short variable arm and a C1:G72 pair.[5] These features are sufficient to avoid cross-reactivity between *Mj*$tRNA^{Tyr}$ and bacterial TyrRS. However, the identity elements of archaeal $tRNA^{Tyr}$ are highly similar to the eukaryotic ones, which prevents the use of the *Mj*TyrRS/$tRNA^{Tyr}_{CUA}$ pair as an OTS in eukaryotic organisms.

It was for all these reasons that a mutated *Mj*TyrRS*/$tRNA^{Tyr}_{CUA}$ pair was eventually used as the first OTS for site-specific GCE via amber codon suppression in *E. coli* in 2001.[5] Since then, several mutations in both the *Mj*TyrRS and the cognate *Mj*$tRNA^{Tyr}$ were aimed at optimizing the OTS for several ncAAs with different structures.

Mutations on the *Mj*TyrRS were directed towards decreasing the recognition of its natural substrate (Tyr) while increasing it for the desired ncAA. To that end, the crystal structure of *Bacillus stearothermophilus* TyrRS (PDB: 1TYD)[2] was key to

identify several mutation points near the substrate binding domain, which is well conserved between archaeal, bacterial and eukaryotic TyrRSs (**Fig. 8-1**).[5] Positions Tyr32, and Asp158, which are essential for Tyr specificity, together with Gln155, Gln173 and Tyr151, have been either randomized or mutated rationally and subjected to selection screenings to evolve the enzyme towards the desired ncAA.[6,7] Other common mutation sites, such as Glu107, Ile159 or Leu162, are important to enlarge the catalytic pocket for bulkier ncAAs.[6] From the *Mj*tRNA^{Tyr} side, a total of 11 nucleotides that do not interact directly with the enzyme have been mutated to increase the orthogonality in *E. coli*.[5] Using several rounds of positive and negative selections in libraries with these mutated residues, more than 60 ncAAs structurally similar to Tyr or Phe have been so far incorporated using the *Mj*TyrRS*/tRNA$^{Tyr}_{CUA}$ pair, with this number continuously growing.[8]

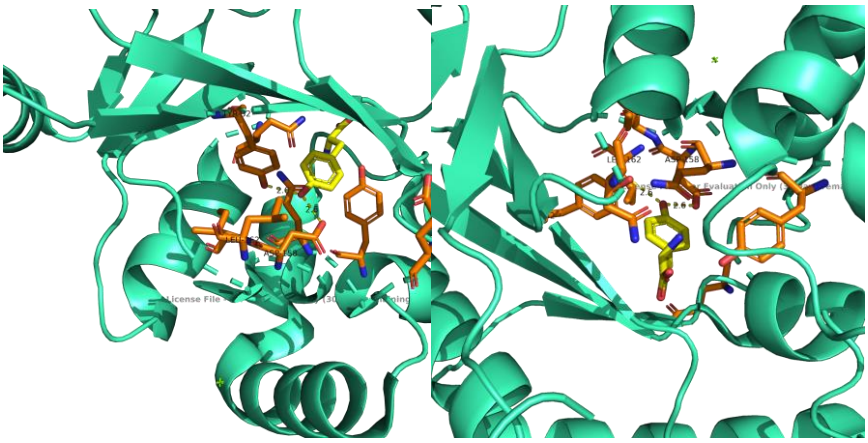


Figure 8-1: Crystal structure of wild type *Mj*TyrRS (PDB: 1J1U). *Mj*TyrRS shown in cyan ribbon, relevant residues for substrate recognition in orange sticks and the natural substrate (Tyr) in yellow sticks. Two perspectives are represented.

8.2. GCE with 1,2,4,5-tetrazines in *E. coli*

8.2.1. Background:

At the beginning of this Ph.D. thesis,† the only reported ncAAs containing the 1,2,4,5-tetrazinyl moiety had been incorporated in *E. coli* employing the *Mj*TyrRS*/tRNA$^{Tyr}_{CUA}$ pair (**Fig. 8-2**).[9,10] In both cases, the mutated residues that permitted the incorporation of such ncAA were Tyr32, Leu65, Phe108, Gln109, Asp158 and Leu162.

† Recently, the incorporation of a tetrazinyl-ncAA employing the PylRS*/tRNA$^{Pyl}_{CUA}$ in eukaryotes has been reported.[29]

Interestingly, these positions were equally mutated in another commonly used mutant *Mj*TyrRS*, named pCNF, that is considered one of the most polyspecific *Mj*TyrRS variant because it can recognize a large variety of structurally different ncAAs (**Table 8-1**).[11]

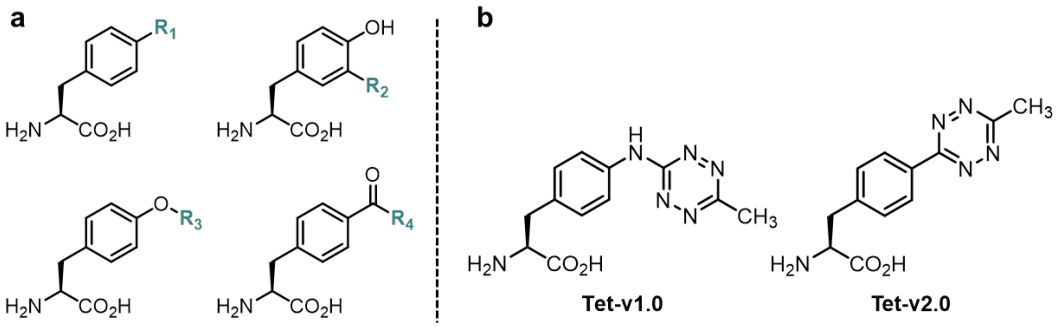


Figure 8-2: ncAAs incorporated by *Mj*TyrRS/tRNA pair. a) General structure of Phe or Tyr analogues recognized by the *Mj*TyrRS*/tRNA pair. **R1** = H, CH_3, N_3, NH_2, Ph, CN, I, Br, NO_2, OCF_3, OSO_3H. **R2** = OH, NH_2, I, NO_2. **R3** = CH_3, CF_3, $C(CH_3)_3$, benzyl, propargyl. **R4** = CH_3, phenyl, 4-F-phenyl, 4-I-phenyl, 4-NO_2-phenyl.[8] **b)** Structures of **Tet-v1.0**[9] and **Tet-v2.0**,[10] the only tetrazinyl-containing ncAAs incorporated in *E. coli*.

Table 8-1: List of the wild-type sequence and different reported *Mj*TyrRS mutants for the incorporation of Tet-v1.0, Tet-v2.0 and a polyspecific variant (pCNF).

Residue	32	65	108	109	158	162
*Mj*TyrRS-wt	Tyr	Leu	Phe	Gln	Asp	Leu
*Mj*TyrRS_Tet-v1.0	Glu	Ala	Pro	Ser	Gly	Gly
*Mj*TyrRS_Tet-v2.0	Gly	Gln	Ser	Asp	Ser	Asn
*Mj*TyrRS_pCNF	Leu	Val	Trp	Met	Gly	Pro

8.2.2. Results:

We envisaged that, given the structural similarities between the reported tetrazinyl-ncAAs and some of the amino acids previously synthesized (seen in **Chapter 3**), we could identify a previously existing mutant or generate a new *Mj*TyrRS*/$tRNA^{Tyr}_{CUA}$ pair suitable to recognize and incorporate our tetrazinyl-containing unnatural amino acids. We centered our efforts on the Phe or Tyr analogues, namely compounds **11b**, **10d** and **11d**. Importantly, **11a** is structurally identical to **Tet-v1.0**, thus we used it as a positive control for incorporation in all subsequent experiments. In terms of the *Mj*TyrRS* variants, we rationalized that *Mj*TyrRS_Tet-v1.0, *Mj*TyrRS_Tet-v2.0 or *Mj*TyrRS_pCNF could be suitable mutants to recognize **11b**, **10d** or **11d** as potential

ncAAs for amber codon suppression. Therefore, we requested all three mutants from Dr. Ryan Mehl's laboratory at Oregon State University, which we received in three different transformation plasmids, each one encoding for one of the synthetase variants under a lpp constitutive promoter, together with its cognate *Mj*tRNA$^{Tyr}_{CUA}$. The plasmid itself, named pDule, is a low copy number plasmid with tetracycline resistance and p15A origin of replication, which facilitates the co-transformation with commonly used expression plasmids because they do not share the same selection and replication features (**Fig. 8-3**).

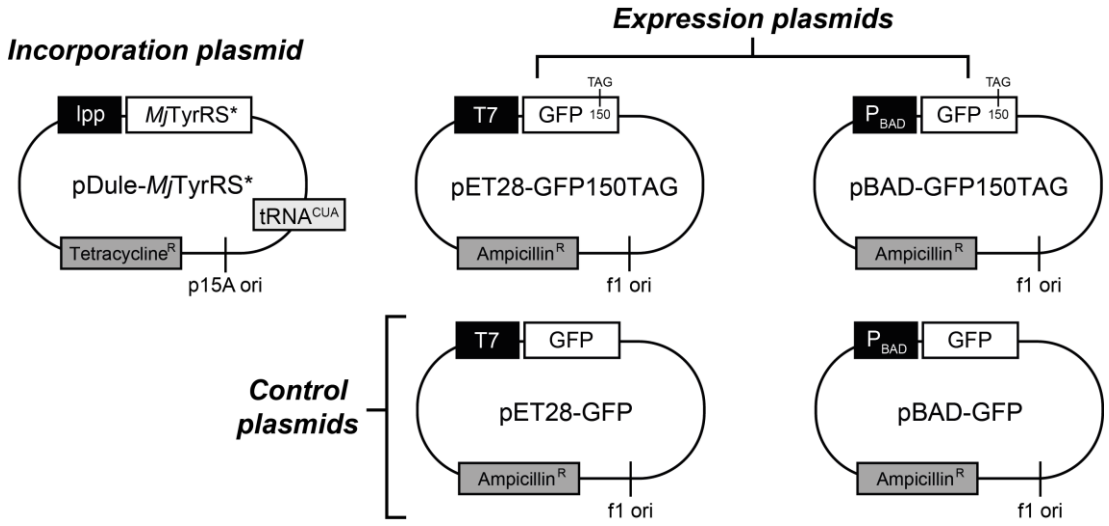


Figure 8-3: Maps of the different vectors used for ncAA incorporation, depending on their function. Left: vector with the required components of the incorporation machinery (*Mj*TyrRS*/tRNA$^{Tyr}_{CUA}$ pair). **Right:** gene of the protein of interest (in this case, a C-terminal His-tagged GFP is used as a model) with an in-frame amber stop codon at the desired site of incorporation (Asn150TAG). His-tagged GFP without the in-frame amber codon is used a control of expression.

As the expression plasmid to check for ncAA incorporation we used either pET28 or pBAD, that differ in the systems to induce protein expression: while pET28 is a T7 RNA polymerase-based vector that requires galactose or analogous molecules such as isopropyl β-D-1-thiogalactopyranoside (IPTG) for induction, pBAD employs the arabinose promoter. This is of importance because it limits the *E. coli* strain that can be used for protein expression, as pET expression systems need a bacterial strain modified with DE3 lysogens, whereas the pBAD vector does not require any special features as the arabinose operon is inherent in *E. coli*. Also, both of them confer ampicillin resistance and employ a f1 origin of replication. In both cases, a His-tagged Green Fluorescent Protein (GFP) with an amber stop codon at a permissive position (Asn150) was used to check for ncAA incorporation (GFP150TAG) (**Fig. 8-3**).

Initially, different strains of *E. coli* were transformed with pET28-GFP or pBAD-GFP (positive control of expression) and pET28-GFP150TAG or pBAD-GFP150TAG without the addition of any potential ncAA in the incubation media (negative control of expression). The tested strains were BL21 (DE3), which is a commonly used protein expression strain; and C321, a recoded strain in which all 321 amber stop codons had been replaced genome-wide with ochre stop codons (that is, all TAG codons to TAA), specifically engineered for GCE purposes.[12] In the beginning, typical expression conditions were used: bacteria transformed with both vectors (one incorporation and one expression plasmids) were grown in lysogeny broth (LB) at 37 °C until the optical density (OD) reached ≈ 0.4, moment at which the expression inducer (IPTG at 0.5 mM or arabinose at 0.1% final concentration) were added. At 3 hours post-induction, the cultures were centrifuged, and the pellet of bacteria lysed by sonication. For expression tests, the fraction of soluble proteins after centrifugation was loaded directly on a gel. When purification was required, a dedicated protocol was followed (**Fig. 8-4**).

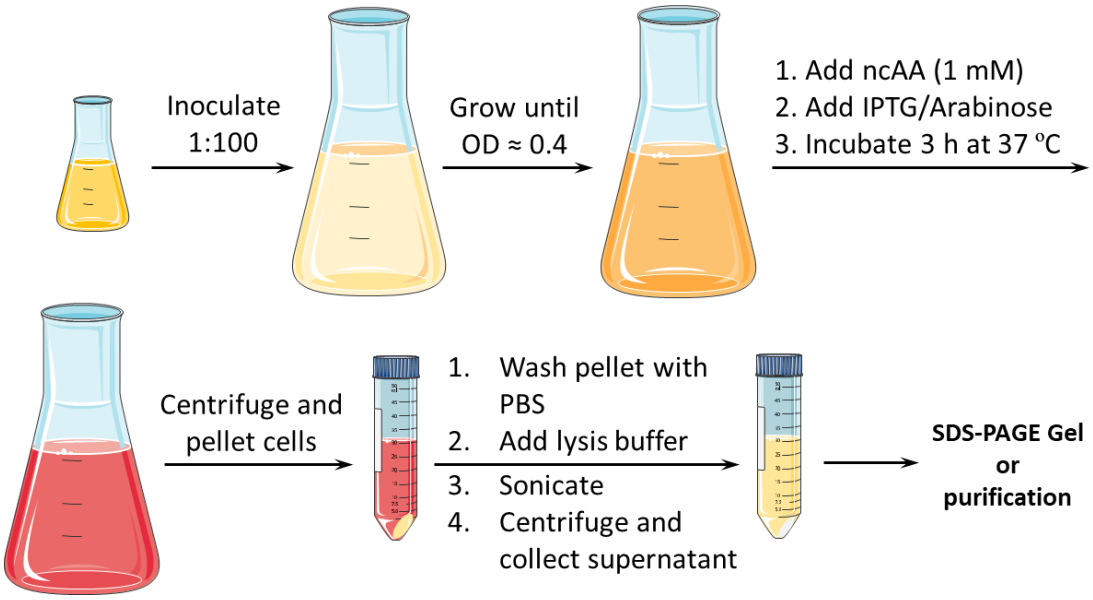


Figure 8-4: Purification protocol used initially for ncAA incorporation. Culture media represented in pink because of the typical coloration of compounds bearing the tetrazinyl moiety.

As expected, expression of GFP was detected in both strains when transformed with pBAD-GFP, but only in BL21 (DE3) for pET28-GFP because the C321 strain does not express DE3 lysogens for T7 promoter recognition. Importantly, the negative control of expression with GFP150TAG showed that only a faint band appeared at the expected molecular weight, probably consistent with some basal levels of stop codon

readthrough (**Fig. 8-5, a**). Next, as a control for the genetic incorporation protocol, the incubation of both strains transformed with pDule-*Mj*TyrRS_Tet-v1.0 and pBAD-GFP150TAG in the presence of 1 mM of **Tet-v1.0** was performed. pBAD was chosen as the vector because it enabled the test with both strains, while pET28 would be limited to BL21 (DE3). After 3 hours expression at 37 ºC, the culture was centrifuged, the pellet of cells collected and a purification protocol using His SpinTrap TALON columns was performed. These columns are prepacked with cobalt-IMAC (immobilized metal affinity chromatography) resin, which is more selective than nickel-charged resins, and suitable for purification using a benchtop centrifuge. In both cases, the desired protein (GFP-**Tetv1.0**) could be purified and the expected molecular weight verified by intact mass spectrometry (**Fig. 8-5, b**).

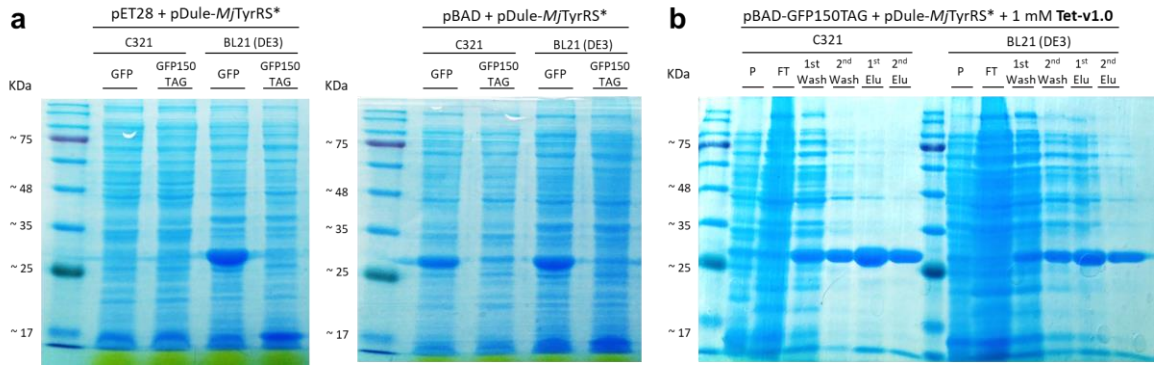


Figure 8-5: SDS-PAGE gels of the expression of GFP or GFP150TAG using either pET28/pBAD + pDule-*Mj*TyrRS* in BL21 (DE3) or C321. a) Expression test in the absence of the ncAA for the incorporation system (**Tet-v1.0**). The expected expression band for GFP appears at ≈ 27.8 KDa. **b)** Expression of GFP150TAG with the addition of **Tet-v1.0.** Each lane is a different step during the purification process. The expected expression band for GFP**-Tet-v1.0** appears at ≈ 27.9 KDa. **Note:** in both cases, pDule-*Mj*TyrRS* refers to the mutant for **Tet-v1.0** incorporation. (KDa = Kilodalton, P = Pellet, FT = Flow-through, Elu = Elution fraction).

In light of the positive outcome of the different expression tests, the different Tyr or Phe analogues (**11b**, **10d** and **11d**) were used to check for incorporation (**Fig. 8-6, a**). C321 was the selected strain to conduct further incorporation studies because it led to a cleaner purification, hence the vectors of choice were the pBAD class of plasmids. In this case, a positive control of expression (pBAD-GFP), a negative control of expression (pBAD-GFP150TAG in the absence of ncAA), and a positive control of incorporation (pBAD-GFP150TAG with the addition of 1 mM **Tet-v1.0** in the medium) were used to ensure the quality of results. Due to ncAA availability, the only vector

tested was pDule-*Mj*TyrRS_Tet-v1.0. Results after expression and purification showed that none of the tested amino acids seemed to be incorporated in the protein sequence, as the recovered faint band after elution did not significantly differ from the negative control of expression (**Fig. 8-6, b**). Nevertheless, intact mass spectrometry of the recovered elution fractions was performed, which did not show the expected molecular weights, indicating that the purified protein was indeed a product of amber stop codon readthrough. This is supported by the anti-His tag Western Blot performed, which indicated that all bands correspond to the expected construct containing the purification tag (**Fig. 8-6, c**).

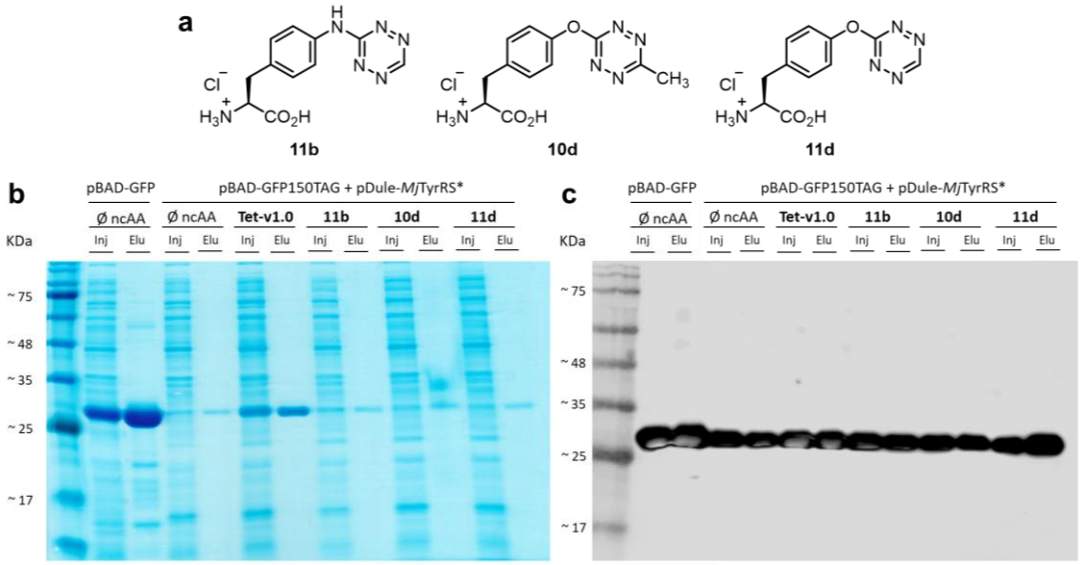
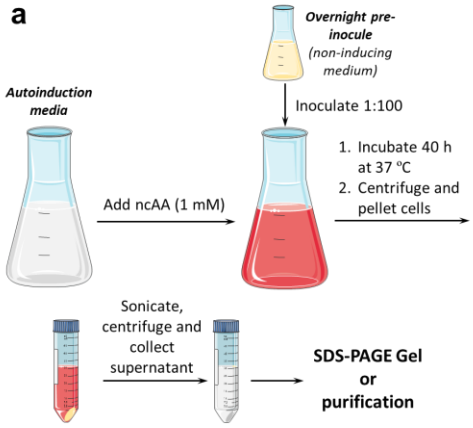


Figure 8-6: Protein expression and purification to check for 11b, 10d and 11d incorporation. a) Structures of **11b**, **10d** and **11d. b)** SDS-PAGE gel of the incorporation test of each amino acid, with the relevant controls, in double-transformed *E. coli* C321. **c)** Anti-His tag Western Blot of the same incorporation test. **Note:** in both cases, pDule-*Mj*TyrRS* refers to the mutant for **Tet-v1.0** incorporation. (KDa = Kilodalton, Inj = Injection, Elu = Elution fraction).

The failed incorporation for all three tested amino acids prompted us to rethink the experimental conditions used. In general, failure in the incorporation of a potential ncAA can arise for different reasons: limited stability of the amino acid in the incubation media, low cellular permeability, incompatibility with auxiliary translation factors or the ribosome itself, degradation of the translated protein due to a suboptimal folding, or poor aminoacylation activity by the mutant AARS.[13] The latter can be modulated to a certain extent in two different ways: either by selecting AARS mutants with an improved recognition for the given amino acid, or by decreasing competitive interactions for the enzyme, which can be achieved by depleting certain

natural amino acids from the culture media. Hence, we decided to employ a tailored autoinduction media[9,10] lacking tyrosine (the original substrate of *Mj*TyrRS*) and cysteine, as its reduction potential could limit the stability of the tetrazine ring in its reactive, oxidized state (**Fig. 8-7**). The autoinduction media triggered a milder induction for expression, which is sometimes beneficial to ensure proper protein folding, and the expression time was extended to maximize the protein production up to the metabolic saturation of the bacterial culture. The *E. coli* strain was also switched to DH10β, which is also a commonly used strain for GCE purposes, due to its higher translation fidelity and resilience to grow in depleted media compared to C321. Since DH10β does not express T7 polymerase, pBAD was also the vector of choice.

a



b

	Non-inducing medium (*pre-inocule*)	Autoinduction medium (**Arabinose induction**)	Autoinduction medium (**Lactose induction**)
Aspartate (5%, pH 7.5)	25 mL	25 mL	25 mL
Glycerol (10% m/v)	---	25 mL	25 mL
18 amino acid mix (25x)	20 mL	20 mL	20 mL
Arabinose (20% m/v)	---	1.25 mL	1.25 mL
Lactose (10% m/v)	---	---	10 mL
Glucose (40% m/v)	6.25 mL	625 µL	625 µL
M salts (25x)	20 mL	20 mL	20 mL
$MgSO_4$ (1 M)	1 mL	1 mL	1 mL
Trace metals (1000x)	500 µL	500 µL	500 µL
Sterile water	to 500 mL	to 500 mL	to 500 mL

Figure 8-7: Modified protein expression protocol for ncAA incorporation. a) Expression protocol based on a tailored autoinduction medium. **b)** Components of the different tailored media: a non-inducing medium lacking arabinose and lactose as the sugar source; an autoinducing media for pBAD vectors including arabinose; and the version for vectors with T7 promoters (i.e.: pET vectors), adding lactose. The *18 amino acid mix (25x)* includes all proteinogenic amino acids except for Tyr and Cys at a concentration of 5 mg/mL (to reach a final concentration in the culture of 200 µg/mL). *M salts (25x)* is a type of phosphate buffered solution (details in **Experimental Section 8.3.1.**). *Trace metals (1000x)* is a commercially available stock solution (Fisher Scientific).

With the modified expression protocol, an expression test with **11b**, **10d** and **11d** and the three *Mj*TyrRS* variants was performed. Whilst **10d** and **11d** did not yield any protein expression, **11b** successfully expressed GFP-**11b**. Strikingly, all three mutants showed various degrees of incorporation, demonstrating the tolerability of the different mutated enzymes (**Fig. 8-8, a**). Expression of GFP-**11b** was verified by intact mass spectrometry. The obtained protein mass corresponded to GFP in which the residue Asn150 (Mw = 132.1 g/mol) was substituted by **11b** (Mw = 260.3 g/mol),

thus a 128 Da difference between wild-type GFP and GFP-**11b** was detected (**Fig. 8-8, b**). Out of the three different variants, *Mj*TyrRS_Tet-v2.0 was ultimately chosen because it displayed the highest yield for GFP-**11b** (36.6 mg protein / L culture, 8% of the total yield obtained for wild-type GFP under the same expression conditions).

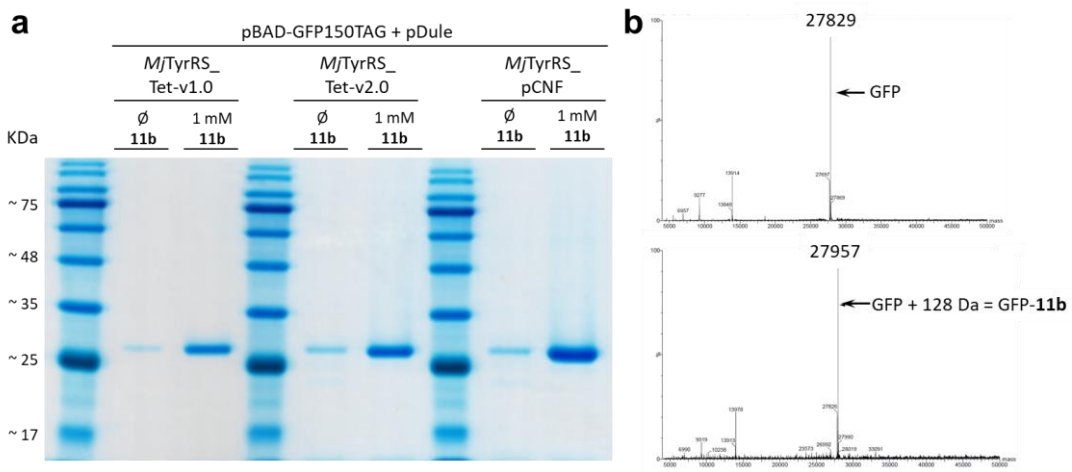


Figure 8-8: Expression test for the genetic incorporation of 11b by different *Mj*TyrRS* mutants. a) Elution fractions in the purification process of the different expressions with or without the addition of 1 mM **11b** in the autoinduction medium. The faint band in the expressions lacking **11b** are probably resultant of an amber stop codon readthrough.

To verify that the ncAA **11b** was reactive in the protein surface for bioorthogonal conjugation, GFP-**11b** was reacted with **BCN-OH** (10 equivalents) (**Fig. 8-9, a, left**).

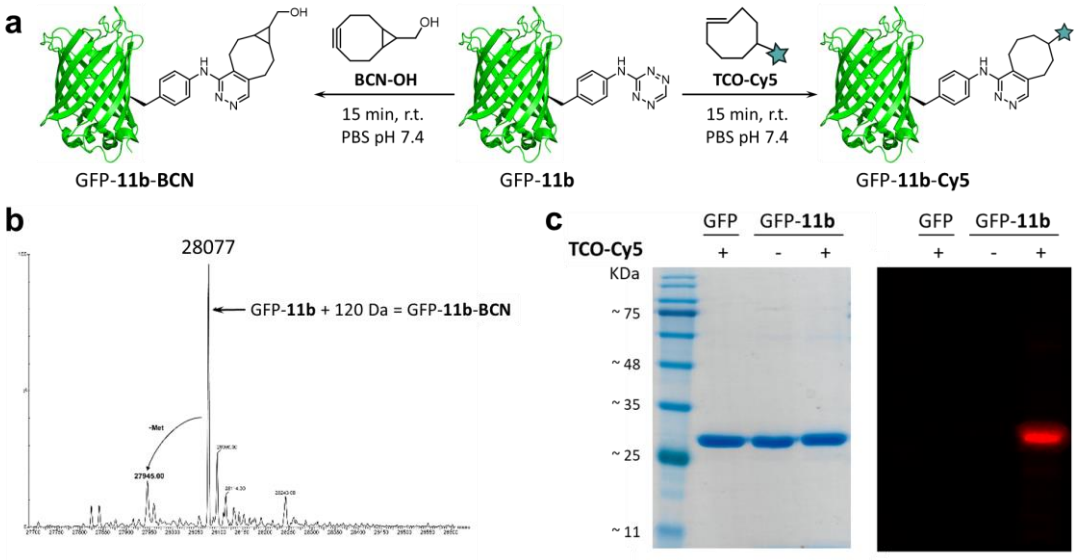


Figure 8-9: iEDDA reactivity of GFP-11b with strained alkenes and alkynes. a) Reaction between GFP-**11b** and the strained cyclooctyne **BCN-OH** or the strained *trans*-cyclooctene fused to cyanine 5 (**TCO-Cy5**). **b)** The completion of the reaction with **BCN-OH** was verified by intact mass spectrometry. **c)** Specificity of the reaction with **TCO-Cy5** was verified by Coomassie staining and fluorescence imaging of the SDS-PAGE gel.

Intact mass spectrometry of the obtained product demonstrated the reaction was complete, with no GFP-**11b** remaining (**Fig. 8-9, b**). Alternatively, TCO fused to a fluorophore (**TCO-Cy5**) was also reacted with GFP-**11b** (**Fig. 8-9, a, right**), and fluorescence imaging of the SDS-PAGE gel verified that only GFP-**11b** reacted with the fluorophore, showing that the reaction takes places uniquely due to the presence of the tetrazine ring in the ncAA (**Fig. 8-9, b**).

To showcase a potential application of the proteins with **11b** site-selectively incorporated, we envisaged that the unsubstituted 1,2,4,5-tetrazin-3-amino functionality in **11b** could potentially undergo bioorthogonal *click-to-release* reactions, similarly to the tetrazinyl-labeled lysine residues seen in **Chapter 6**. To demonstrate that **11b** can also undergo liberation reactions to release doxorubicin from the prodrug **TCO-Dox**, we performed the same *in vitro* reaction in human serum at 37 ºC (**Chapter 6, Section 6.3.2.**), monitoring the liberation of free doxorubicin over time by UPLC (**Fig. 8-10, a**). We verified that **11b** was able to liberate doxorubicin to a ca. 50% total liberation until reaching a plateau after 10 hours of reaction, demonstrating the potential application of the genetic incorporation of **11b** to generate proteins capable of activating prodrugs *in vivo* (**Fig. 8-10, b**).

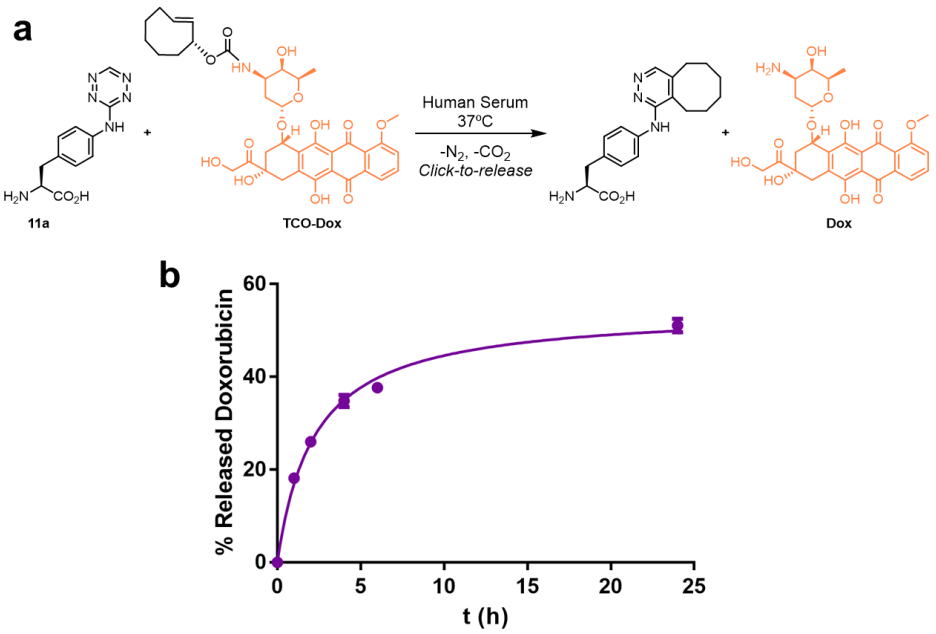


Figure 8-10: *Click-to-release* reaction of 11b to liberate doxorubicin from TCO-Dox. a) Reaction scheme of the *click-to-release* between **11b** and **TCO-Dox**. **b)** Liberation of doxorubicin from TCO-Dox at different time-points.

The advantage of genetically incorporating **11b** into proteins over targeting the reactive lysine residues is evident when considering the obtained products in each case. With lysine labeling, a heterogeneous population of labeled proteins in terms of substituted positions and total tetrazine loading is obtained, which might be detrimental for the pharmacokinetic profile of the conjugated antibody and can lead to batch-to-batch variability. However, the site-selective incorporation that can be achieved with **11b** enables the production of a homogeneous population of labeled protein. Therefore, we aimed at the production of an antigen-specific protein that could be expressed and purified in *E. coli* to genetically incorporate **11b**.

8.3. Genetic incorporation of ncAAs in antibody fragments

8.3.1. Background:

Our aim was to express and purify a therapeutically relevant protein that we could engineer in order to site-selectively incorporate **11b**. Ideally, the protein should be antigen-specific to enable a spatiotemporal control of drug activation, similarly to our results in **Chapter 6** with the tetrazinyl-labeled Trastuzumab and the targeted drug delivery to BT474 HER2$^+$ cells. However, the expression and purification of full-length antibodies in *E. coli* or other prokaryotic organisms is extremely challenging.[14,15] Not only is the presence of disulfide bonds critical for the oligomeric antibody structure a limitation because they cannot be selectively oxidized in *E. coli*, but also aggregation in inclusion bodies is very common. Dedicated protocols for protein denaturation and subsequent refolding exist, but they tend to suffer from low recovery yields.[16–18] Furthermore, the presence of post-translational modifications (PTMs), like glycosylation of specific residues, are essential to exert their function.[19] Therefore, in the specific case of GCE with ncAAs, where protein production is greatly decreased, this becomes a critical issue. For all these reasons, we anticipated that the production of antibody-like structures or antibody fragments, much more suited to prokaryotic expression, could be a suitable alternative.

We centered out attention specifically on nanobodies, because their monomeric structure and absence of PTMs allow for their expression in microbial systems such

as *E. coli*.[14] Nanobodies are the fragments of a particular class of heavy-chain-only antibodies found naturally in the serum of some camelids.[20] Their variable antigen binding domain, which retains full recognition ability, is commonly known as nanobody due to its small size and molecular weight (12-14 kDa compared to ≈ 150 KDa of a common antibody).[14] Nanobodies can be used in a similar manner to antibodies, but they display an improved pharmacokinetic profile and route of administration, hence they attract the interest of the pharmaceutical industry for their therapeutic applicability. Indeed, in February 2019 the first nanobody (Caplacizumab, Ablynx) was approved for the treatment of acquired thrombotic thrombocytopenic purpura, and several more are currently under development.[21] Of relevance, a recent report shows the genetic incorporation of a photoactivable ncAA into a nanobody using *E. coli* as the expression host,[22] demonstrating the viability of the strategy.

At the IRB Barcelona, the research group of Prof. Ernest Giralt have recently published the development of a nanobody for the specific blocking of the epidermal growth factor receptor (EGFR) activation.[23] EGFR is a transmembrane protein, member of the family of receptor tyrosine kinases, which is activated by epidermal growth factor (EGF).[24] In the context of cancer, overactivation of EFGR is a known driver of oncogenesis, thus it has been the target for many anticarcinogen strategies.[25] We postulated that, besides blocking EGFR activation, site-selective incorporation of **11b** in the nanobody designed by Prof. Giralt's laboratory could provide targeted drug delivery at the site of action, since EGFR is overexpressed in the tumor environment. Consequently, we started a collaboration with Dr. Salvador Guardiola, leader of the research project, to test the genetic incorporation of **11b** in the sequence of the specified nanobody, named Nb6.

8.3.2. Results:

We kindly received the expression plasmids and the expression protocol for Nb6 from Dr. Guardiola. Briefly, the C-terminal His-tagged Nb6 sequence is inserted in the pMECS under the T7 promoter (**Fig. 8-11, b**). Importantly, a secretion signal sequence upstream of the gene, namely a 22 N-terminal leader sequence of pectatelyase B (pelB) from *Erwinia carotovora*, directs the secretion of the protein to the bacterial

periplasmic space, where it is purified using a protocol relying on the bacterial lysis by osmotic stress.[26] *E. coli* WK6 strain is the preferable expression host due to its remodeled bacterial wall, which facilitates lysis by osmotic shock.

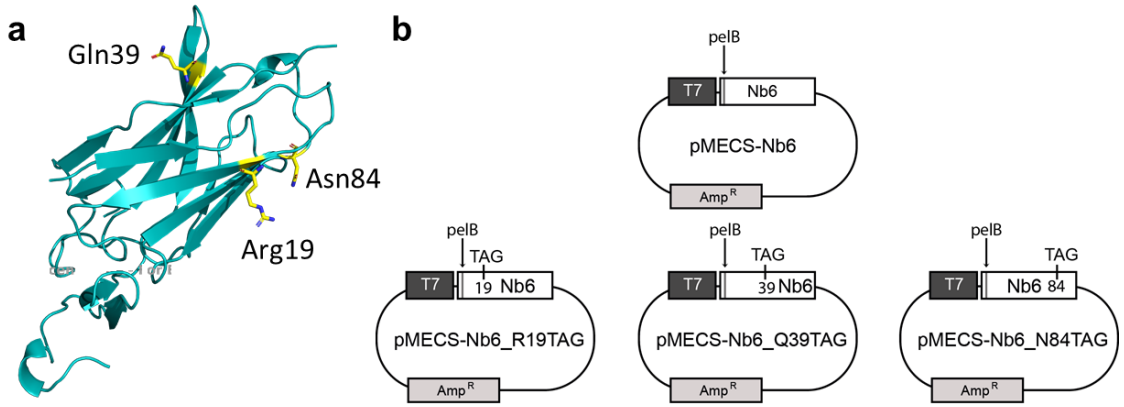


Figure 8-11: Nb6 incorporation sites and expression vectors. a) X-ray crystal structure of Nb6 represented as a cyan cartoon, with the chosen incorporation positions highlighted as yellow sticks. **b)** Vector maps used for Nb6 expression.

Initially, we decided three potential incorporation sites in the Nb6 structure based on three parameters: i) that they were not part of the antigen recognition domain, ii) that they were located at the protein surface and facing to the solvent exposed region, and iii) that they were not part of any of the β-sheets critical for the structure of the nanobody. Hence, based on the X-ray crystallographic structure of the nanobody provided by Dr. Guardiola,[23] we chose positions Arg19, Gln39 and Asn84 and mutated the relevant codons to the amber stop codons by site-directed mutagenesis directly in the pMECS-Nb6 vector (leading to -R19TAG, -Q39TAG and -N84TAG mutations; **Fig. 8-11, a**).

Therefore, we started with an expression test using *E. coli* WK6 cells double transformed with pDule-*Mj*TyrRS_Tet-v2.0 and pMECS-Nb6 (positive control) or pMECS-Nb6_Q39TAG (negative control). We used the autoinduction media specific for lactose induction, as WK6 is a suitable strain for T7 polymerase activation, and we incubated for 40 hours at 37 ºC. Then, the specific cell lysis in osmotic shock conditions, followed by purification in His Spintrap TALON columns (GE Healthcare). SDS-PAGE gel showed that the expected band at ≈ 16.8 KDa for Nb6 appeared in the positive control, whereas no band appeared in the elution lane of the negative control of the truncated gene (**Fig. 8-12, a**).

Next, the expression of Nb6 in the presence of 1 mM of **11b** was attempted in all 3 mutated constructs with amber codon insertion. Unfortunately, under our experimental conditions, no expression of Nb6-**11b** was observed for any of the constructs (**Fig. 8-12, b**), indicating that experimental variations are required in order to achieve the successful incorporation of **11b** in the nanobody.

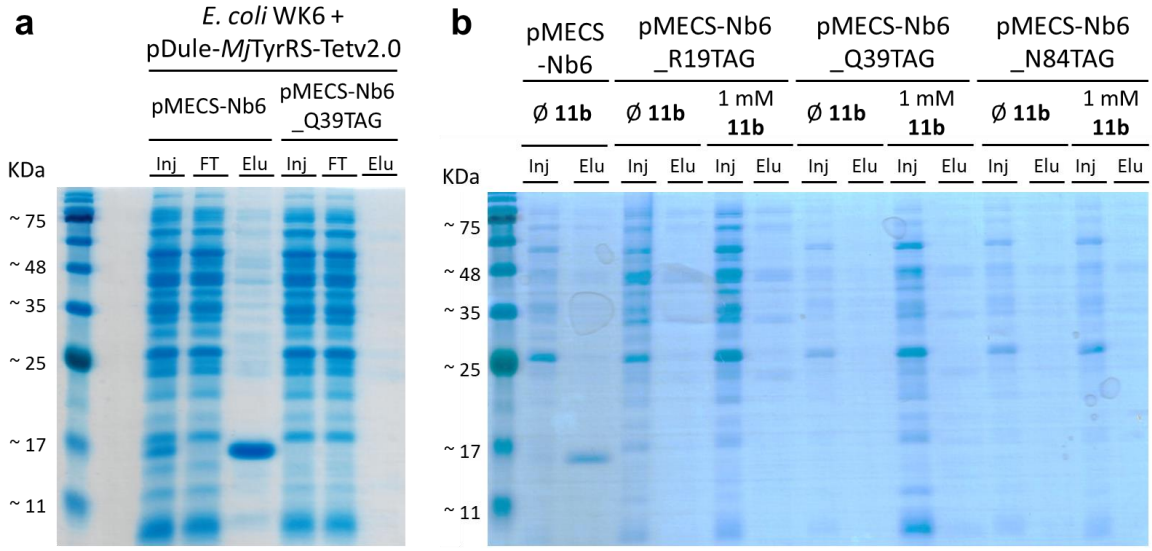


Figure 8-12: Expression tests of Nb6 nanobody. a) SDS-PAGE gel of the positive control of expression (wild-type Nb6) and negative control (truncated Nb6 with an in-frame amber stop codon at Gln39). **b)** SDS-PAGE gel of the expression test of the different constructs with an in-frame amber stop codon in the presence of 1 mM of **11b**, together with positive and negative control of expression (Inj = Injection, FT = Flow-through, Elu = Elution).

At the end of this Ph.D. Thesis, no further experiments were performed in order to achieve expression of Nb6-**11b**. We identified several potential reasons for the unsuccessful incorporation of **11b**, which should be explored beyond the current point. For instance, alternative *E. coli* strains should be explored beyond WK6, as well as other purification methods that do not rely on cell lysis by osmotic shock. As an example, in the recent report where GCE was performed to generate a ncAA-containing nanobody,[22] *E. coli* K12 (DE3) UT5600 cells were used as the expression host, and bacterial lysis was performed using a cell disruptor instead of osmotic shock. This provides a complementary protocol that could be tested in the future.

8.4. Summary:

We have genetically encoded the Phenylalanine derivative **11b** as a ncAA in *E. coli* using a previously existing TyrRS*/tRNA pair from the archaeon *Methanocaldococcus jannaschii*. The use of a tailormade autoinduction expression media lacking Tyrosine and Cysteine was critical to reduce competing ligands for the enzyme and achieve expression of a model protein with **11b** incorporated at a specific position (GFP-**11b**).

The ncAA **11b** can also trigger the bioorthogonal *click-to-release* reaction with **TCO-Dox**, which can be a useful attribute to produce nanobodies bearing a genetically encoded **11b** able to activate prodrugs specifically at the site of action.

However, the main limitation of the current incorporation system with the *Mj*TyrRS*/tRNA pair is the suboptimal expression yields obtained. To date, we have only been able to incorporate **11b** into a model protein (GFP) with very high expression yields, where the limited incorporation efficiency of **11b** by **Tet-v2.0** is not a critical factor.

The incorporation of **11b** into proteins with lower yields remains a challenge. For that reason, a research placement in Prof. Dieter Söll's laboratory at Yale University was planned from June to August 2020, which was aimed at providing the necessary tools to optimize the incorporation system by different directed evolution techniques.[27,28] Unfortunately, the scheduled placement had to be cancelled due to the COVID19 pandemic. However, the collaboration could be restarted in the future if deemed convenient, as it could help to increase the applicability of the genetic incorporation of **11b** for drug delivery purposes.

8.5. Bibliography:

1 W. J. Jones, J. A. Leigh, F. Mayer, C. R. Woese and R. S. Wolfe. Methanococcus jannaschii sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. *Arch. Microbiol.* 1983, **136**, 254–261.

2 P. Brick, T. N. Bhat and D. M. Blow. Structure of tyrosyl-tRNA synthetase refined at 2.3 Å resolution. *J. Mol. Biol.* 1989, **208**, 83–98.

3 B. A. Steer and P. Schimmel. Major anticodon-binding region missing from an archaebacterial tRNA synthetase. *J. Biol. Chem.* 1999, **274**, 35601–35606.

4 K. Oki, K. Sakamoto, T. Kobayashi, H. M. Sasaki and S. Yokoyama. Transplantation of a tyrosine editing domain into a tyrosyl-tRNA synthetase variant enhances its specificity for a tyrosine analog. *Proc. Natl. Acad. Sci.* 2008, **105**, 13298–13303.

5 L. Wang, A. Brock, B. Herberich and P. G. Schultz. Expanding the genetic code of Escherichia coli. *Science* 2001, **292**, 498–500.

6 Y. Zhang, L. Wang, P. G. Schultz and I. A. Wilson. Crystal structures of apo wild-type M. jannaschii tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O -methyl-L-tyrosine. *Protein Sci.* 2005, **14**, 1340–1349.

7 J. M. Turner, J. Graziano, G. Spraggon and P. G. Schultz. Structural Characterization of a p -Acetylphenylalanyl Aminoacyl-tRNA Synthetase. *J. Am. Chem. Soc.* 2005, **127**, 14976–14977.

8 A. Dumas, L. Lercher, C. D. Spicer and B. G. Davis. Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci.* 2015, **6**, 50–69.

9 J. L. Seitchik, J. C. Peeler, M. T. Taylor, M. L. Blackman, T. W. Rhoads, R. B. Cooley, C. Refakis, J. M. Fox and R. A. Mehl. Genetically Encoded Tetrazine Amino Acid Directs Rapid Site-Specific in Vivo Bioorthogonal Ligation with trans-Cyclooctenes. *J. Am. Chem. Soc.* 2012, **134**, 2898–2901.

10 R. J. Blizzard, D. R. Backus, W. Brown, C. G. Bazewicz, Y. Li and R. A. Mehl. Ideal Bioorthogonal Reactions Using A Site-Specifically Encoded Tetrazine Amino Acid. *J. Am. Chem. Soc.* 2015, **137**, 10044–10047.

11 D. D. Young, T. S. Young, M. Jahnz, I. Ahmad, G. Spraggon and P. G. Schultz. An Evolved Aminoacyl-tRNA Synthetase with Atypical Polysubstrate Specificity. *Biochemistry* 2011, **50**, 1894–1900.

12 M. J. Lajoie, A. J. Rovner, D. B. Goodman, H.-R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church and F. J. Isaacs. Genomically Recoded Organisms Expand Biological Functions. *Science* 2013, **342**, 357–360.

13 P. O'Donoghue, J. Ling, Y.-S. Wang and D. Söll. Upgrading protein synthesis for synthetic biology. *Nat. Chem. Biol.* 2013, **9**, 594–598.

14 I. Jovčevska and S. Muyldermans. The Therapeutic Potential of Nanobodies. *BioDrugs* 2020, **34**, 11–26.

15 O. Spadiut, S. Capone, F. Krainer, A. Glieder and C. Herwig. Microbials for the production of monoclonal antibodies and antibody fragments. *Trends Biotechnol.* 2014, **32**, 54–60.

16 M. A. Boss, J. H. Kenten, C. R. Wood and J. S. Emtage. Assembly of functional antibodies from immunoglobulin heavy and light chains synthesised in E. coli. *Nucleic Acids Res.* 1984, **12**, 3791–3806.

17 S. Cabilly, A. D. Riggs, H. Pande, J. E. Shively, W. E. Holmes, M. Rey, L. J. Perry, R. Wetzel and H. L. Heyneker. Generation of antibody activity from immunoglobulin polypeptide chains produced in Escherichia coli. *Proc. Natl. Acad. Sci.* 1984, **81**, 3273–3277.

18 T. Arakawa and D. Ejima. Refolding Technologies for Antibody Fragments. *Antibodies* 2014, **3**, 232–241.

19 R. Jefferis. in *Biosimilars of Monoclonal Antibodies*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2016, pp. 155–199.

20 S. Muyldermans. Nanobodies: Natural Single-Domain Antibodies. *Annu. Rev. Biochem.* 2013, **82**, 775–797.

21 C. Morrison. Nanobody approval gives domain antibodies a boost. *Nat. Rev. Drug Discov.* 2019, **18**, 485–487.

22 B. Jedlitzke, Z. Yilmaz, W. Dörner and H. D. Mootz. Photobodies: Light-Activatable Single-Domain Antibody Fragments. *Angew. Chem., Int. Ed.* 2020, **59**, 1506–1510.

23 S. Guardiola, M. Varese, M. Sánchez-Navarro, C. Vincke, M. Teixidó, J. García, S. Muyldermans and E. Giralt. Blocking EGFR Activation with Anti-EGF Nanobodies via Two Distinct Molecular Recognition Mechanisms. *Angew. Chem., Int. Ed.* 2018, **57**, 13843–13847.

24 R. S. Herbst. Review of epidermal growth factor receptor biology. *Int. J. Radiat. Oncol.* 2004, **59**, S21–S26.

25 S. Sigismund, D. Avanzato and L. Lanzetti. Emerging functions of the EGFR in cancer. *Mol. Oncol.* 2018, **12**, 3–20.

26 P. Singh, L. Sharma, S. R. Kulothungan, B. V. Adkar, R. S. Prajapati, P. S. S. Ali, B. Krishnan and R. Varadarajan. Effect of Signal Peptide on Stability and Folding of Escherichia coli Thioredoxin. *PLoS One* 2013, **8**, e63442.

27 D. I. Bryson, C. Fan, L. T. Guo, C. Miller, D. Söll and D. R. Liu. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* 2017, **13**, 1253–1260.

28 T. Suzuki, C. Miller, L. T. Guo, J. M. L. Ho, D. I. Bryson, Y. S. Wang, D. R. Liu and D. Söll. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* 2017, **13**, 1261–1266.

29 H. S. Jang, S. Jana, R. J. Blizzard, J. C. Meeuwsen and R. A. Mehl. Access to Faster Eukaryotic Cell Labeling with Encoded Tetrazine Amino Acids. *J. Am. Chem. Soc.* 2020, **142**, 7245–7249.

8.6. Experimental section:

8.6.1. Protein expression and purification with ncAAs:

Expression method A: *Escherichia coli* BL21 (DE3) or C321 cells were co-transformed with the incorporation plasmid (pDule-*Mj*TyrRS*) and the expression plasmid (pET28- or pBAD-GFP150TAG, or pET28- or pBAD-GFP for expression control). Each corresponding mutant was grown overnight shaking at 37 ºC, and 0.25 mL of the saturated culture was used to inoculate 25 mL LB medium supplemented with ampicillin (50 µg/mL) and tetracycline (15 µg/mL). The culture was shaken at 37 ºC until OD ≈ 0.4, moment which IPTG (0.5 mM final concentration) or arabinose (0.1% final concentration) was added, and 1 mM of ncAA when required. The culture was shaken at 37 ºC for 3 hours, upon which the cells were pelleted at 4000x g, the supernatant discarded, the remaining pellet washed with PBS once.

Expression method B: *Escherichia coli* DH10β cells were co-transformed with the incorporation plasmid (pDule-*Mj*TyrRS*) and the expression plasmid (pBAD-GFP150TAG, or pBAD-GFP for expression control). Each corresponding mutant was grown overnight shaking at 37 ºC in a tailormade non-inducing medium, and 0.25 mL of the saturated culture was used to inoculate 25 mL of the tailormade autoinduction medium for pBAD vectors (**Fig. 8-7** and **Table 8-2**) supplemented with ampicillin (50 µg/mL) and tetracycline (15 µg/mL), and 1 mM of ncAA when required. The culture was shaken at 37 ºC for 40 hours, upon which the cells were pelleted at 4000x g, the supernatant discarded, the remaining pellet washed with PBS once.

Table 7-2: Preparation of 25x M salts buffer for the tailormade autoinduction media.

	Amount of reagent	***Concentration***
Na_2HPO_4	88.7 g	0.625 M
KH_2PO_4	85.1 g	0.625 M
NH_4Cl	66.9 g	1.250 M
$Na_2SO_4 \cdot 10H_2O$	40.3 g	0.125 M
H_2O	to 1000 mL	

All used pET28, pBAD and pDule plasmids were provided by Ryan Mehl's laboratory (Oregon State University), except from pDule-MjTyrRS_pCNF, which was obtained from Addgene.

Purification method: cell pellets were resuspended in *Lysis buffer* and lysed in a Branson Sonifier SFX550 for 5 minutes (2.5 s ON, 2.5 s OFF, 5 min total time, 20% Amplitude). The resulting suspension was centrifuged at 15000x g for 20 at 4 ºC, and for expression tests, the supernatant was directly loaded onto an SDS-PAGE and stained with BlueSafe (NZY Tech). When protein purification was required, the supernatant was taken and purified on His Spintrap TALON columns (GE Healthcare) following manufacturer's protocol. Briefly, the supernatant (injection, Inj) is loaded onto the purification columns in fractions of 500 µL each, followed by centrifugation of the loaded sample (flow-through, FT). The resin is then washed twice with *Wash buffer*, and the His-tagged protein ultimately eluted in two fractions with *Elution buffer.* Each fraction was then loaded onto an SDS-PAGE gel stained and stained with BlueSafe (NZY Tech).

Expression method C: *Escherichia coli* WK6 cells were co-transformed with the incorporation plasmid (pDule-*Mj*TyrRS*) and the expression plasmid (three different pMECS-Nb6_TAG, or pBAD-GFP for expression control). Each corresponding mutant was grown overnight shaking at 37 ºC in a tailormade non-inducing medium, and 0.25 mL of the saturated culture was used to inoculate 25 mL of the tailormade autoinduction medium for pET28 vectors (with lactose induction) supplemented with ampicillin (50 µg/mL) and tetracycline (15 µg/mL), and 1 mM of ncAA when required. The culture was shaken at 37 ºC for 40 hours, upon which the cells were pelleted at 4000x g, the supernatant discarded, the remaining pellet washed with PBS once and the supernatant discarded after centrifugation. Lysis was performed in osmotic stress conditions, resuspending the pellets in 3 mL *Osmotic lysis buffer* and shaking vigorously at 4 ºC for 30 min, moment in which 6 mL of the *Osmotic lysis buffer* diluted 4 times was added and shaken for 30 min more. The resulting suspension was centrifuged at 15000x g for 20 at 4 ºC, and then the purification method proceeded in His Spintrap TALON columns (GE Healthcare) as specified in the previous section.

The pMECS-Nb6 vector was provided by Dr. Salvador Guardiola (Design, Synthesis and Structure of Peptides and Proteins group, Institute for Research in Biomedicine), and the

corresponding mutations to incorporate in-frame amber stop codons were performed with QuikChange II Site-Directed Mutagenesis kit following manufacturer's protocol.

Table 8-3: Specification of the used buffers.

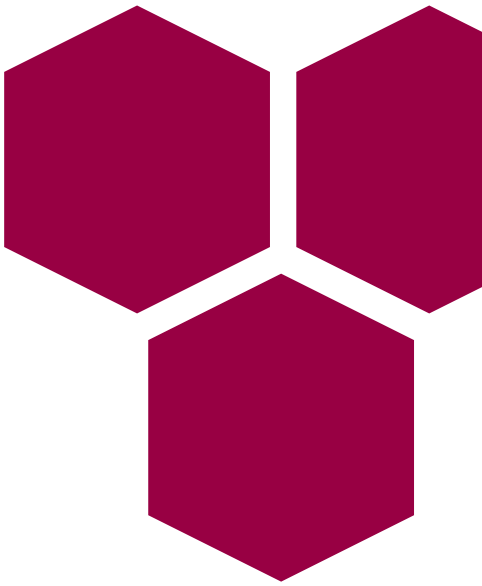
Lysis Buffer	***Wash Buffer***	***Elution Buffer***	***Osmotic Lysis Buffer***
20 mM NaPi pH 8.0	20 mM NaPi pH 8.0	20 mM NaPi pH 8.0	200 mM Tris pH 8.0
200 mM NaCl	200 mM NaCl	200 mM NaCl	0.5 mM EDTA
50 mM imidazole	50 mM imidazole	500 mM imidazole	500 mM sucrose
DNAse I (ThermoFisher) cOmplete Protease Inhibitor Cocktail (Roche)			

8.6.2. Intact mass spectrometry of expressed proteins:

Same experimental protocol as in **Chapter 5, Section 5.4.3.** was followed.

8.6.3. GFP-**11b** bioconjugation:

GFP-**11b** in PBS pH 7.4 buffer was at 10 µM concentration was mixed with 10 equivalents of either **BCN-OH** (Merck) or **TCO-Cy5** (Carbosynth) dissolved in DMSO at 1 mM solution. After addition of the corresponding volume of reactant, the reaction was mixed briefly and left for 15 minutes. Then, the excess of reagent was removed using Amicons 0.5 mL 10000 MWCO (Merck). For **TCO-Cy5** labeling, the protein sample was loaded directly onto and SDS-PAGE gel, which was read in an Amersham Typhoon biomolecular imager (GE Healthcare) with Cy5 fluorescence parameters (λ_{Ex} = 678 nm, λ_{Em} = 694 nm) and subsequently stained with BlueSafe (NZY Tech). In the case of **BCN-OH** labeling, the solvent was exchanged to ammonium acetate 10 mM for intact mass spectrometry.



Chapter 9

Genetic code expansion with *Methanosarcina barkerii* pyrrolysyl-tRNA synthetase / $tRNA^{Pyl}_{CUA}$ pair

9.1. Incorporation of ncAAs in eukaryotic organisms:

Methanosarcina barkeri is a methanogenic anaerobic archaeon that was described by Schnellen in 1947,[1] although the fully characterized type strain was found in 1966.[2] Initially isolated from mud samples, it is normally found in the rumen of cattle, where it contributes to the digestion of polysaccharides such as cellulose to produce carbon dioxide, methane and organic acids.[2] In 1998 and subsequent years, genome sequencing of *M. barkeri* revealed that several genes encoding for methylamine methyltransferases displayed an in-frame amber stop codon that did not result in translation termination.[3,4] When the crystal structure of one of these enzymes, MtmB, was solved in 2002,[5] it was discovered that the amber codon was suppressed by a unique residue, a lysine derivative that was named Pyrrolysine (Pyl, **Fig. 9-1, a**). Together with Selenocysteine (SeC), which is known as the 21st amino acid in the genetic code, Pyl was dubbed the 22nd proteinogenic amino acid.

Simultaneously, a dedicated translation system for its genetic encoding was discovered, formed by a novel AARS, named pyrrolysyl-tRNA synthetase (PylRS), and an amber suppressor tRNA ($tRNA^{Pyl}_{CUA}$).[6] Over the years, several homologs of this enzyme have been discovered in nearly 100 different archaea and bacteria beyond *M. barkeri.* However, for GCE purposes we focused specifically in the *Mb*PylRS/tRNA pair, as it is the mostly used and well-studied PylRS.

Similar to *Mj*TyrRS, *Mb*PylRS has neither an editing domain nor an anticodon binding loop. The former tolerates the aminoacylation of the $tRNA^{Pyl}$ with ncAAs without the subsequent cleavage of the formed bond. The latter allows mutations in the anticodon of the tRNA for further codon reassignments, even though wild-type $tRNA^{Pyl}_{CUA}$ is already capable of amber codon suppression with no modifications required.[7,8] Importantly, the *Mb*PylRS/tRNA pair is orthogonal in bacteria as well as eukaryotic organisms. Unique features of both the enzyme and the tRNA account for such specificity: $tRNA^{Pyl}$ presents a shortened variable loop specifically recognized by the enzyme, which rejects tRNAs with longer variable loops; the tRNA also displays a short D-loop, and a reduced spacing between the acceptor and D-stems, which

altogether form a compacted L-shaped tRNA recognized by the binding surface of PylRS, that does not interact with bulkier tRNAs.[9–11]

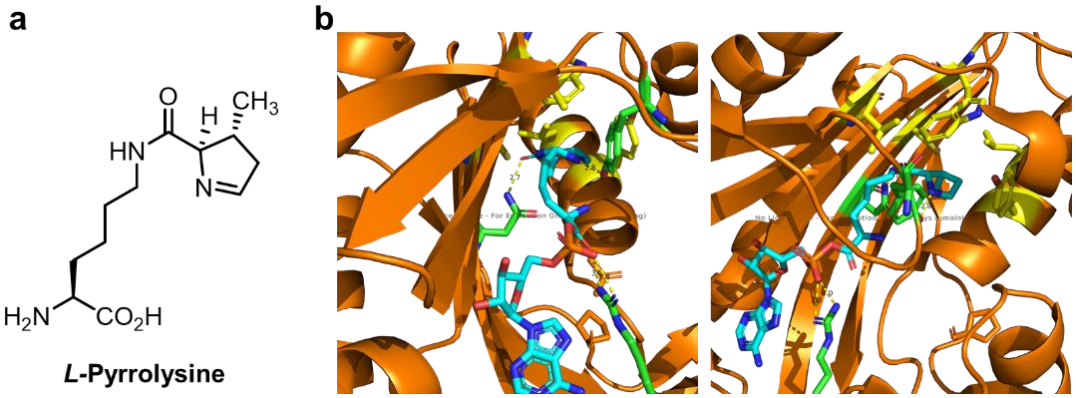


Figure 9-1: Pyl and PylRS. a) Chemical structure of N^6-{[(2*R*,3*R*)-3-methyl-3,4-dihydro-2*H*-pyrrol-2-yl]carbonyl}-L-lysine, also known as *L*-Pyrrolysine (Pyl). **b)** X-Ray crystal structure (PDB: 2Q7H) of *Mm*PylRS (orange ribbon), bound to its natural substrate (Pyl, cyan sticks). The interacting residues through hydrogen bonding (Arg330, Asn346 and Tyr384) are represented in green sticks, with the corresponding distance measurements annotated; suitable points of mutation to increase substrate recognition (Tyr306, Leu309, Cys348, Val401 and Trp417) are represented in yellow sticks. Two perspectives are shown.

PylRS is relatively promiscuous, and is already capable of accepting an array of ncAAs analogous to Pyl in its wild-type form without the need for specific mutations.[12] Importantly, unlike *Mj*TyrRS, that requires mutations in the substrate binding pocket to prevent recognition of Tyr, PylRS does not need further engineering because Pyl is not found naturally in most organisms. However, engineering of the binding pocket is required for the recognition of certain ncAA structures. The identification of key residues involved in substrate recognition was facilitated when the crystal structure of the catalytic domain of PylRS from another closely related methanogenic archaea (*Methanosarcina mazei*; alignment of *Mb*PylRS and *Mm*PylRS shown in **Experimental Section 9.3.4.**) was solved.[13] Three residues were identified to be forming hydrogen bonds with Pyl: Arg330 interacts with the oxygens in the α-carboxylic acid group; Asn346 bonds with the oxygen in the N^{ε}-amide group, creating an important interaction that needs to be maintained when designing ncAAs (with a notable exception, see below);[14] and Tyr384 forms a hydrogen bond with the nitrogen in the pyrrole ring.[15–17] While the first two normally need to be conserved, Tyr384 is suitable for mutagenesis to open de substrate binding site for ncAA with a functional group different to the pyrrole stemming from the N^{ε}-amide in Pyl. Other residues

surrounding the substrate binding domain and generally mutated to accommodate different ncAA structures are Tyr306, Leu309, Cys348, Val401 and Trp417† (**Fig. 9-1, b**).[7,12,15,18]

9.2. GCE with 1,2,4,5-tetrazines in mammalian cells

9.2.1. Background:

Generally, these set of mutations enable the genetic incorporation of ncAAs analogous to Pyl, that retain the lysine backbone functionalized in the N^{ε} with an amide or, more commonly, a carbamate (**Fig. 9-2, a**). The functionalization arising from the carbamate is variable, with bulky hydrophobic groups normally well tolerated. However, Phe analogues can also be recognized mutating two residues: Asn346, responsible for the hydrogen bond with the carbonyl of the amide, and Cys348 (**Fig. 9-2, a**).[19] Recently, a set of Phe derivatives, namely **Tet-v3.0**, have become the first (and only) ncAAs bearing a tetrazine functional group incorporated in eukaryotic cells.[20] Previously, two other tetrazinyl-containing ncAAs (**TetK** and **mTetK**) had been incorporated by the *Mb*PylRS*/$tRNA^{Pyl}_{CUA}$ pair in bacteria,[21] but they were not progressed into eukaryotic organisms (**Fig. 9-2, b; Table 9-1**).

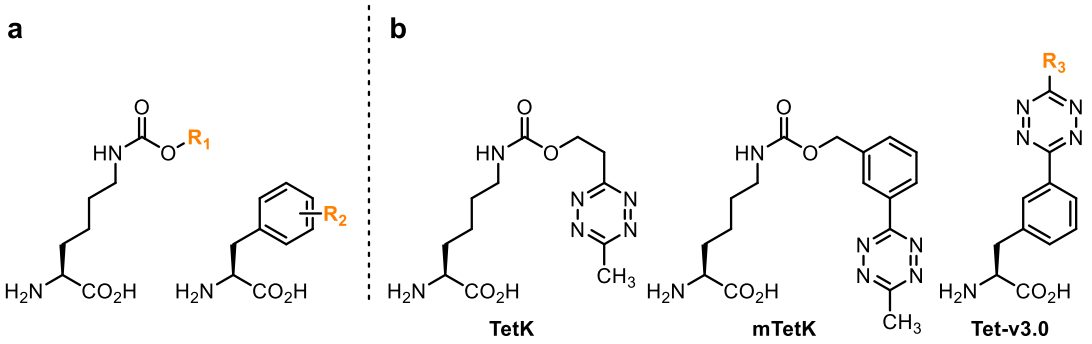


Figure 9-2: ncAAs incorporated by *Mb*PylRS/tRNA pair. a) General structure of Lys or Phe analogues recognized by the *Mj*TyrRS*/tRNA pair. **R_1**: $C(CH_3)_3$, cyclopentyl, propenyl, butenyl, propargyl, BCN, TCO, norbonenyl, 4-I-benzyl, 4-NO_2-benzyl. **R_2**: (*ortho*) H, CH_3, Cl, Br, I, OCH_3, CN, NO_2; (*meta*) CH_3, CF_3, OCH_3, NO_2, CN, N_3, F, Cl, Br, I; (*para*) 2-Cl-ethoxy, 2-Br-ethoxy, 2-I-ethoxy, 2-N_3-ethoxy.

† Residue numbers annotated for *Mm*PylRS; for *Mb*PylRS, the same residues are found with a -35 difference. Therefore, residues for hydrogen bonding with Pyl in *Mb*PylRS are Arg295, Asn311 and Tyr349; while typically mutated residues to accommodate different ncAAs are Tyr271, Leu274, Cys313, Val366 and Trp382.

b) Structures of **TetK**, **mTetK**[21] and **Tet-v3.0**,[20] the only tetrazinyl-containing ncAAs incorporated by the PylRS/tRNA^{Pyl} pair. **R$_3$:** CH_3, CH_2CH_3, $(CH_2)_3CH_3$, isopropyl.

Table 9-1: List of the wild-type sequence and different reported *Mb*PylRS mutants for the incorporation of TetK, mTetK and Tet-v3.0. *The notation for the *Mb*PylRS enzyme is annotated. The same conserved residues in the *Mm*PylRS enzyme, which is the one that has been referred in the text, are Leu305, Tyr306, Asn346 and Cys348. **Two different equally active PylRS variants are reported (I and II). **Note:** hyphen (-) means that the residue is not mutated from the wild-type enzyme.

Residue*	270	271	311	313
*Mb*PylRS-wt	Leu	Tyr	Asn	Cys
*Mb*PylRS-TetK/mTetK	-	Gly	-	Val
*Mb*PylRS-Tetv3.0 (I)**	Gly	-	Gly	Ala
*Mb*PylRS-Tetv3.0 (II)**	Gly	-	Gly	Ser

9.2.2. Directed evolution of PylRS*/tRNA pairs towards ncAAs:

Directed evolution of PylRS*/tRNA pairs to incorporate a given ncAA is normally required to achieve high incorporation efficiencies, especially when there are no OTSs reported for ncAAs with similar structures. Screening assays of active AARS mutants for GCE tend to be performed in bacteria, even when the end-goal is the application of the OTS in a different host organism.[22,23] The reason behind the use of bacteria, most commonly *E. coli*, is the number of cells that can be transformed and cultured, and the relatively simple selection conditions. By contrast, if selection was directly performed, for instance, in mammalian cells, the transformation efficiency and the number of cells per culture would limit vastly the number of different enzyme variants to be screened. As a result, selection assays are typically designed for *E. coli*, and rely on serial positive and negative survival methods aimed at screening OTSs capable of incorporating the ncAA while rejecting other natural amino acids, thereby ensuring orthogonality.[22–24]

A classically employed method relies on an initial *positive selection* assay based on antibiotic resistance in the presence of the ncAA.[25] In short, bacteria are initially transformed both with a given antibiotic resistance gene with an in-frame amber stop codon at a permissive position and the library of AARS*/tRNA pairs. By adding the ncAA in the media, only those AARS* variants capable of ncAA incorporation will translate the full-length protein and survive, whilst those that do not recognize the

ncAA will produce a truncated, inactive antibiotic resistance protein. Sequencing of the AARS* gene in the surviving clones allows identification of the active AARS* mutants. In a second step, a *negative selection* is performed by transforming the bacterial colonies isolated from the previous step with a lethal gene (for instance barnase,[26] CcdB toxin[27] or TolC[28]) truncated with a stop codon, and cultured in the absence of the ncAA. Those undesired AARS* that recognize natural amino acids will translate the full-length protein and not survive.[25]

After several rounds of selection with more stringent conditions, such as higher antibiotic and/or lower ncAA concentration in the positive selection, the enzyme is evolved further to accept favorable mutations, until the desired efficiency is achieved.[25] More modern directed evolution techniques, such as phage-assisted methods, have also been developed to improve catalytic activities, but they normally rely on AARS*/tRNA pairs already capable of incorporating ncAAs with moderate efficiency as a starting point, which can be subsequently fine-tuned to increase performance.[8,29,30]

9.2.3. Results:

Our aim was to generate the first[††] OTS for the genetic incorporation of a ncAA bearing a 1,2,4,5-tetrazine moiety in mammalian cells. Given the orthogonality of the different AARS/tRNA pairs, we aimed at evolving PylRS towards the recognition of a suitable ncAA. To that end, our first goal was to develop a suitable assay that allowed the screening and selection of multiple PylRS* mutants.

We initially designed the positive selection assay (**Fig. 9-3**) by inserting the kanamycin resistance gene into the previously used pBAD-GFP150TAG (in **Chapter 8**) to generate an Amp^R and Kan^R version of the plasmid. Next, we truncated the ampicillin resistance gene with an amber stop codon at the permissive position 183 (Ala183TAG) to use ampicillin resistance as the selective condition for active PylRS* variants. We hypothesized that the presence of the GFP150TAG gene in the vector would aid in the selection of the most active mutants by fluorescence quantification after arabinose induction, and reduce the number of positive rounds of selection

[††] At the time of the beginning of this project, there were no reports of tetrazinyl-ncAA incorporated in eukaryotic organisms.

required. Therefore, we used pBAD-GFP150TAG_Amp183TAG (Kan^R) as our selection plasmid for the positive selection assay.

As the library plasmid, we employed pDule-PylRS*, a vector encoding for 20 different variants of *Mb*PylRS* enzyme with rationally defined mutations at 4 positions near the substrate binding pocket: Tyr271, Leu274, Cys313 and Tyr349. We selected these 4 positions after an extensive literature review of the reported PylRS* for ncAA incorporation, and mutated each position to the usually described changes: Tyr271 was mutated to Ala; Leu274 to Met; Cys313 to Ala and Val; and Tyr349 to Phe, with all the possible combinations (the complete list of mutants is detailed in **Experimental Section 9.3.1.**). This library was constructed using multiple site-directed mutagenesis.

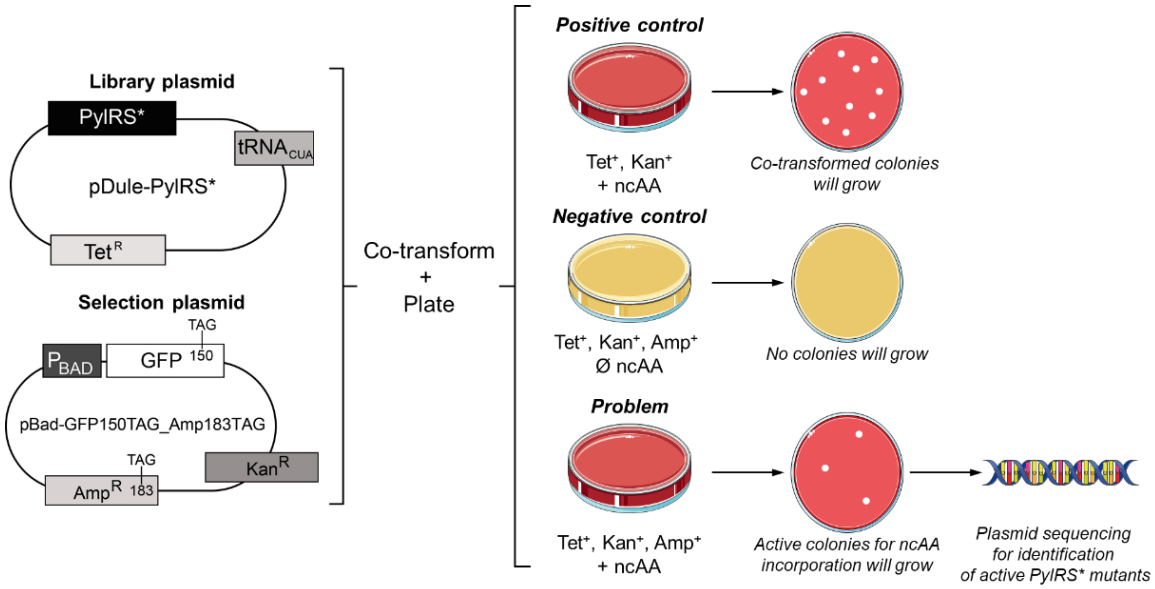


Figure 9-3: Positive selection assay for active PylRS* mutants towards a given ncAA. Both the plasmid with the PylRS* library and the selection plasmid with the truncated ampicillin resistance gene are co-transformed into *E. coli* and plated in three types of LB agar plates. ***Positive control:*** Plates without ampicillin will allow growth of co-transformed bacteria independently from PylRS* activity (the ncAA is added to ensure that is does not display any bacteriostatic or bactericide activity). ***Negative control:*** Plates with ampicillin and without ncAA will not resist antibiotic selective pressure. ***Problem****:* plates with ampicillin and ncAA will only allow growth of those PylRS* mutants that incorporate the ncAA and translate the full-length antibiotic resistance gene.

To demonstrate the correct functionality of the selection assay, a test using the commercially available N$^\varepsilon$-(*tert*-butoxycarbonyl)-*L*-lysine (**N$^\varepsilon$-Boc-Lys**) was performed (**Fig. 9-4, a**). **N$^\varepsilon$-Boc-Lys** is a well-known substrate for different

PylRS*,[10,22,31,32] and we anticipated that several mutants could be identified with our selection screen. Therefore, we conducted the experiment mixing equivalent quantities of each pDule-PylRS* mutant with the selection plasmid and transforming chemically competent *E. coli* BL21 (DE3) cells by heat shock. To check for the effect exerted by the amount of selection antibiotic and ncAA, 2 different ampicillin and **Nε-Boc-Lys** concentrations were tested.

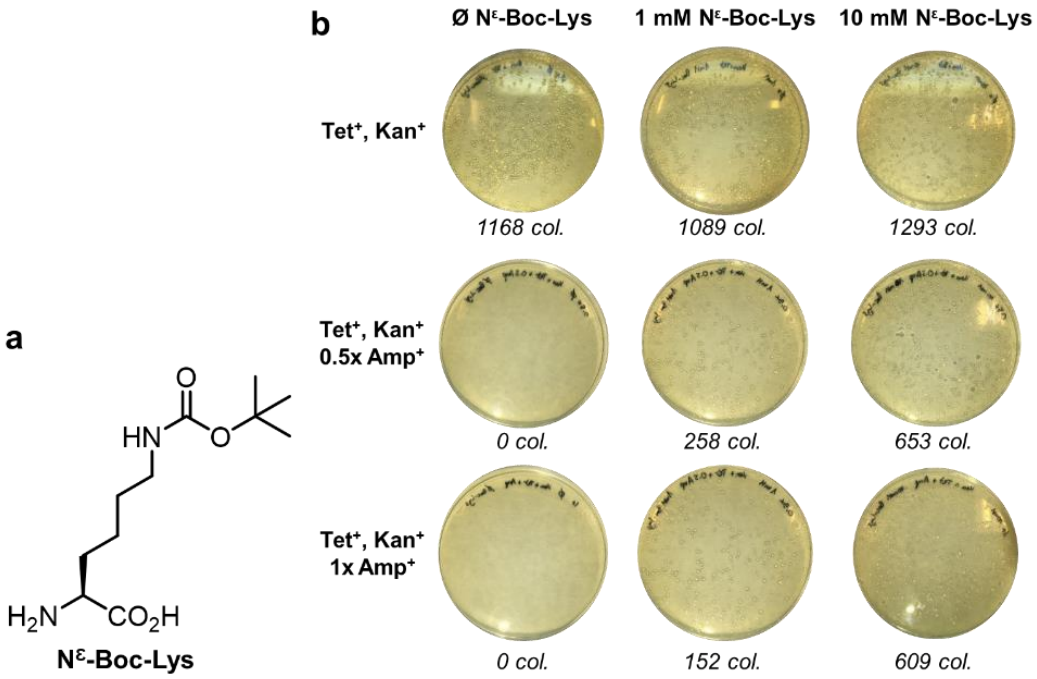


Figure 9-4: Positive selection screening assay using Nε-Boc-Lys as the model ncAA. a) Chemical structure of **Nε-Boc-Lys. b)** Nine different conditions varying **Nε-Boc-Lys** concentration and amount of ampicillin were tested. The colony count is annotated underneath each plate. Ampicillin concentration is a factor of the recommended for LB agar plates (1x Amp = 50 μg/mL; 0.5x Amp = 25 μg/mL).

Results based on number of colonies (**Fig. 9-4, b**) seemed to indicate that the assay was indeed selecting active PylRS* variants: first, all plates without ampicillin (*positive control*) showed a comparable amount of colony count; secondly, plates with ampicillin and absence of **Nε-Boc-Lys** (*negative control*) were lacking viable colonies. On a complementary experiment, we demonstrated that lowering antibiotic concentration below the threshold of 25 μg/mL led to growth of non-resistant colonies, and therefore we kept 0.5x ampicillin concentration in subsequent experiments. Thirdly, plates with **Nε-Boc-Lys** conferred viability in the presence of ampicillin, indicating full-length protein translation due to ncAA incorporation. As

expected, increasing **N$^{\varepsilon}$-Boc-Lys** increased the number of viable colonies, while more stringent selective conditions (higher ampicillin concentration) led to a decreased colony count.

Based on the obtained results, the maximum number of positive colonies obtained (in the Tet^{+}, Kan^{+}, 0.5x Amp^{+} and 10 mM **N$^{\varepsilon}$-Boc-Lys** plate) is more than 50% of the total colonies obtained in the positive control. This could lead to anticipate that half of the PylRS* variants in the library successfully incorporate **N$^{\varepsilon}$-Boc-Lys** as a ncAA. To demonstrate it, a total of 10 colonies were randomly selected, their PylRS gene was PCR-amplified and sequenced to identify the selected mutations (**Fig. 9-5, a**). Unexpectedly, only 5 different mutants were identified, with mutant #13 (bearing the Tyr349Phe mutation) being the most represented mutant (**Fig. 9-5, b**). This can be either as a result of the random selection of colonies, or it might suggest an enrichment of certain mutants over others in the library.

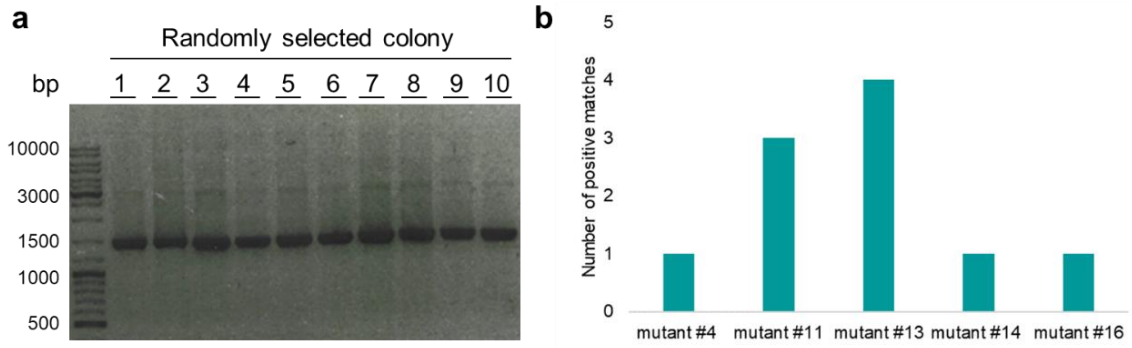


Figure 9-5: Identification of the selected PylRS* variants and their mutations. a) The PylRS* gene of each isolated colony is PCR-amplified, resulting in a single band at the expected 1478 bp. **b)** Number of times a given mutant was positively matched to a sequencing of the 10 randomly selected colonies (complete list of mutants with their mutations in **Experimental Section 9.3.1.**).

To test the activity of the library towards our set of tetrazinyl-containing amino acids (**10a-e**, **11a-e**), we decided to use the Lys and Phe analogues **10a** and **10d** (**Fig. 9-6, a**). Therefore, the positive selection assay was performed with the library of 20 PylRS* variants, using the least stringent conditions employed in the experiment with **N$^{\varepsilon}$-Boc-Lys** (25 μg/mL ampicillin concentration, 10 mM **10a** or **10d**). Unfortunately, no positive colonies were observed in the plates with either **10a** or **10d**, demonstrating that none of the PylRS* mutants were capable of their recognition (**Fig. 9-6, b**). In retrospective, the use of a Lys analogue functionalized with a carbamate in the N$^{\varepsilon}$, able to interact with Asn311 through hydrogen bonding,[14] would have been an interesting

Lys analogue to synthesize and test in addition to **10a**. Similarly, it is now clear that Phe analogues, such as **10d**, require mutagenesis of Asn311 in order to be accommodated by the enzyme.[19,20]

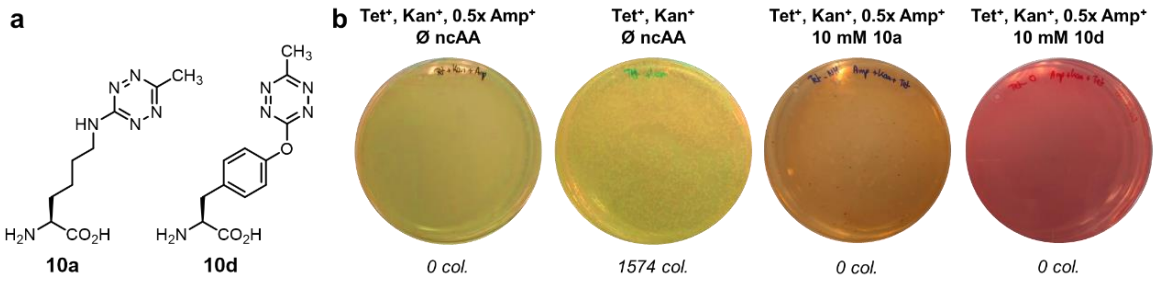


Figure 9-6: Positive selection assay with 10a and 10d. a) Chemical structures of **10a** and **10d**. **b)** Obtained plates after incubation at 37 ºC for 40 hours. The negative control, the positive control and both problem plates are shown from left to right. The colony count is annotated underneath each plate. (0.5x Amp = 25 µg/mL)

Considering the obtained results, we argued that the limited library of 20 PylRS* mutants was not covering enough enzyme variability, posing a constraint in the identification of active PylRS* variants towards tetrazinyl-containing amino acids. Therefore, we designed an enlarged enzyme library through site-saturation mutagenesis[33] in the same four residues (Tyr271, Leu274, Cys313 and Tyr349). We built this library through the synthesis of an artificial PylRS gene bearing degenerate NNK codons (where N = any nucleotide and K = G or T) at the defined positions. The use of NNK codons for site-saturation mutagenesis is very common, and more desirable than NNN codons because it minimizes de presence of undesired in-frame stop codons.[34,35] By flanking the synthetic gene with two different restriction sites, we inserted it into pDule (**Fig. 9-7, a**) and tested the transformed vector in the positive selection assay using **N$^{\varepsilon}$-Boc-Lys** as the model ncAA (**Fig. 9-7, b**).

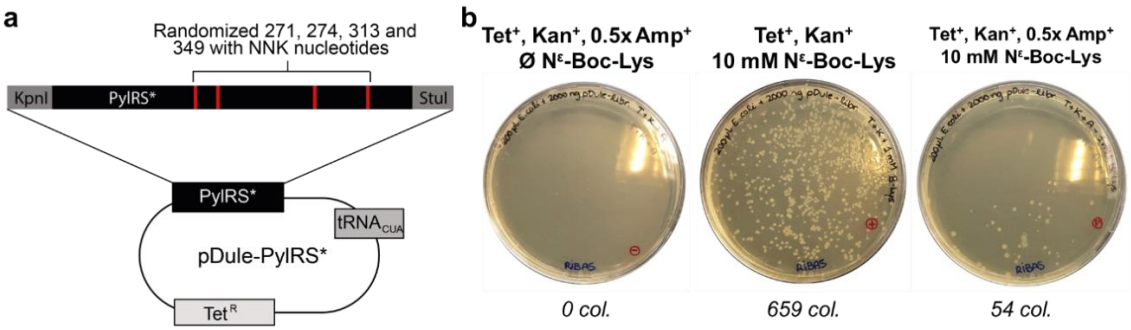


Figure 9-7: Positive selection assay with the library built using site-saturation mutagenesis. a) Map of the used plasmid to generate the library of enzymes with degenerate codons at 4 defined

positions. KpnI and StuI are the inserted flanking sequences for restriction enzymes at the 5' and 3' ends. **b)** LB agar plates resulting from the assay. The colony count is annotated underneath each plate. (0.5x Amp = 25 µg/mL)

As expected, the percentage of colonies able to incorporate **N$^{\varepsilon}$-Boc-Lys** and survive in the presence of ampicillin (8% compared to the total colony count in the positive control) is lower than the obtained when using the previous smaller, rationally designed enzyme library. In order to validate the obtained results, we followed the same approach and isolated 10 colonies to sequence their PCR-amplified PylRS gene (**Table 9-2**). As expected, an increased variability at the four randomized positions was identified, showcasing the advantages of utilizing a larger library in accessing a greater number of active mutant enzymes. Grouping each position separately, and listing the identified mutations pointed towards certain clearly beneficial substitutions, such as Leu274Met, Cys313Ala or Tyr349Phe (**Fig. 9-8**). Importantly, all three substitutions had already been introduced in the smaller PylRS* library (**Experimental Section 9.3.1.**), which validates the rational mutations incorporated in the smaller library. However, other unexpected mutations, such as Tyr270Pro or Cys313Pro, are interesting variants to be considered for further PylRS engineering.

Table 9-2: List of the sequenced PylRS* genes from 10 isolated colonies using the site-saturated mutagenesis generated library towards incorporation of N$^{\varepsilon}$-Boc-Lys.

Residue*	270	274	313	349
*Mb*PylRS-wt	Tyr	Leu	Cys	Tyr
Isolated colony #1	Pro	Val	Pro	Ile
Isolated colony #2	Tyr	Met	Ala	Phe
Isolated colony #3	Leu	Leu	Gly	Ser
Isolated colony #4	Tyr	Met	Ala	Phe
Isolated colony #5	Tyr	Leu	Ala	Phe
Isolated colony #6	Ile	His	Cys	Arg
Isolated colony #7	Asn	Gly	Ala	Leu
Isolated colony #8	Asn	Met	Ala	Phe
Isolated colony #9	Pro	Val	Pro	Ile
Isolated colony #10	Pro	Glu	Ala	Phe

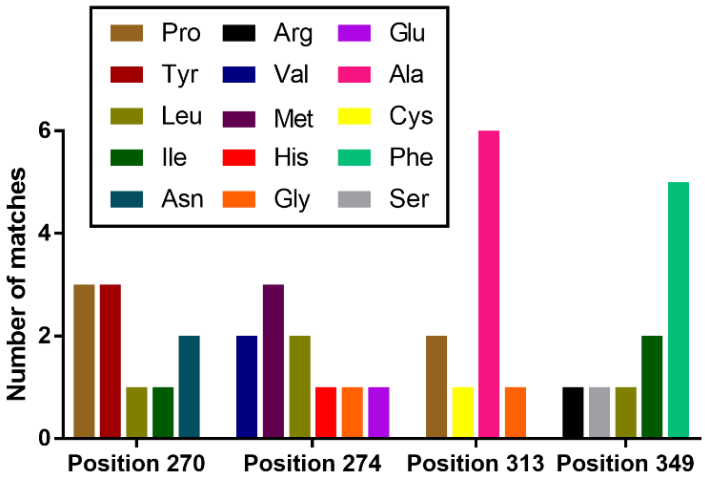


Figure 9-8: Number of times a given mutation was identified at each of the four randomized positions.

Results show how in the positive control plates (lacking ampicillin), the colony count is significantly lower than the number of colonies obtained with the smaller library of 20 PylRS* variants. This can be due to the presence of PylRS* sequences that upon constitutive expression become lethal for the cell and compromise its viability.

Despite the interesting mutations identified, a limitation for the use of the randomly generated library is the low coverage of the different enzyme variants compared to the total number of enzyme variants. Out of the possible codon combinations arising from the site-saturation mutagenesis at the four different positions, $20^4 = 1.6 \cdot 10^5$ different combinations are theoretically accessed. Even by subtracting a calculated 3% of stop codon representation in the library,[34,35] the total number of enzyme variants is at least two orders of magnitude higher than the transformation efficiencies obtained in the used experimental conditions, resulting in a bottleneck for the selection assay. We have improved these numbers switching the bacterial transformation protocol from heat shock to electroporation, using a higher number of initial electrocompetent cell count or increasing the size of the culturing plates from 60 x 15 mm to 100 x 15 mm. However, the colony count in the absence of ampicillin (*positive control*) is still suboptimal. Our next steps ought to be towards solving this limitation before implementing a robust protocol to check for ncAA incorporation by PylRS.

9.3. Summary:

In order to genetically incorporate a ncAA bearing a tetrazine functionality in eukaryotic organisms, we aimed to use the PylRS*/tRNA pair form the archaeon *Methanosarcina barkeri*. Even though we were unable to isolate an active orthogonal pair for the tetrazine-containing amino acids we tested, we developed a positive selection assay that was used with either a small library of 20 representative enzyme variants or with a randomized library prepared using site-saturation mutagenesis. In both cases, we could identify active PylRS* mutants towards **N$^{\varepsilon}$-Boc-Lys**, which we used as a model ncAA, meaning that this methodology could be further utilized in directed evolution experiments.

Similarly to the ongoing work outlined in **Chapter 8**, the research placement in Prof. Dieter Söll's laboratory at Yale University, which was planned from June to August 2020 and was cancelled due to the COVID19 pandemic, would have probably been useful to improve our selection system for active PylRS* variants and increase its applicability.

9.4. Bibliography:

1 G. M. Maestrojuan and D. R. Boone. Characterization of Methanosarcina barkeri MST and 227, Methanosarcina mazei S-6T, and Methanosarcina vacuolata Z-761T. *Int. J. Syst. Bacteriol.* 1991, **41**, 267–274.

2 M. P. Bryant and D. R. Boone. Emended Description of Strain MST(DSM 800T), the Type Strain of Methanosarcina barkeri. *Int. J. Syst. Bacteriol.* 1987, **37**, 169–170.

3 S. A. Burke, S. L. Lo and J. A. Krzycki. Clustered Genes Encoding the Methyltransferases of Methanogenesis from Monomethylamine. *J. Bacteriol.* 1998, **180**, 3432–3440.

4 L. Paul, D. J. Ferguson and J. A. Krzycki. The Trimethylamine Methyltransferase Gene and Multiple Dimethylamine Methyltransferase Genes of Methanosarcina barkeri Contain In-Frame and Read-Through Amber Codons. *J. Bacteriol.* 2000, **182**, 2520–2529.

5 B. Hao, W. Gong, T. K. Ferguson, C. M. James, J. A. Krzycki and M. K. Chan. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* 2002, **296**, 1462–1466.

6 G. Srinivasan, C. M. James and J. A. Krzycki. Pyrrolysine encoded by UAG in archaea: Charging of a UAG-decoding specialized tRNA. *Science* 2002, **296**, 1459–1462.

7 W. Wan, J. M. Tharp and W. R. Liu. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* 2014, **1844**, 1059–1070.

8 T. Suzuki, C. Miller, L.-T. Guo, J. M. L. Ho, D. I. Bryson, Y.-S. Wang, D. R. Liu and D. Söll. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* 2017, **13**, 1261–1266.

9 K. Nozawa, P. O'Donoghue, S. Gundllapalli, Y. Araiso, R. Ishitani, T. Umehara, D. Söll and O. Nureki. Pyrrolysyl-tRNA synthetase–tRNAPyl structure reveals the molecular basis of orthogonality. *Nature* 2009, **457**, 1163–1167.

10 T. Suzuki, C. Miller, L. T. Guo, J. M. L. Ho, D. I. Bryson, Y. S. Wang, D. R. Liu and D. Söll. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* 2017, **13**, 1261–1266.

11 S. V. Melnikov and D. Söll. Aminoacyl-tRNA Synthetases and tRNAs for an Expanded Genetic Code: What Makes them Orthogonal? *Int. J. Mol. Sci.* 2019, **20**, 1929.

12 A. Dumas, L. Lercher, C. D. Spicer and B. G. Davis. Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci.* 2015, **6**, 50–69.

13 R. Ishii, O. Nureki, T. Yanagisawa, R. Fukunaga and S. Yokoyama. Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon Methanosarcina mazei. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 2006, **62**, 1031–1033.

14 V. Flügel, M. Vrabel and S. Schneider. Structural Basis for the Site-Specific Incorporation of Lysine Derivatives into Proteins. *PLoS One* 2014, **9**, e96198.

15 J. M. Kavran, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Soll and T. A. Steitz. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci.* 2007, **104**, 11268–11273.

16 T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto and S. Yokoyama. Crystallographic Studies on Multiple Conformational States of Active-site Loops in Pyrrolysyl-tRNA Synthetase. *J. Mol. Biol.* 2008, **378**, 634–652.

17 T. Yanagisawa, T. Sumida, R. Ishii and S. Yokoyama. A novel crystal form of pyrrolysyl-tRNA synthetase reveals the pre- and post-aminoacyl-tRNA synthesis conformational

states of the adenylate and aminoacyl moieties and an asparagine residue in the catalytic site. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2013, **69**, 5–15.

18 L. Guo, Y. Wang, A. Nakamura, D. Eiler, J. M. Kavran and M. Wong. Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. , DOI:10.1073/pnas.1419737111.

19 Y.-S. Wang, W. K. Russell, Z. Wang, W. Wan, L. E. Dodd, P.-J. Pai, D. H. Russell and W. R. Liu. The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of l-phenylalanine and its derivatives. *Mol. Biosyst.* 2011, **7**, 714.

20 H. S. Jang, S. Jana, R. J. Blizzard, J. C. Meeuwsen and R. A. Mehl. Access to Faster Eukaryotic Cell Labeling with Encoded Tetrazine Amino Acids. *J. Am. Chem. Soc.* 2020, **142**, 7245–7249.

21 S. V. Mayer, A. Murnauer, M. Wrisberg, M. Jokisch and K. Lang. Photo-induced and Rapid Labeling of Tetrazine-Bearing Proteins via Cyclopropenone-Caged Bicyclononynes. *Angew. Chem., Int. Ed.* 2019, **58**, 15876–15882.

22 P. O'Donoghue, J. Ling, Y.-S. Wang and D. Söll. Upgrading protein synthesis for synthetic biology. *Nat. Chem. Biol.* 2013, **9**, 594–598.

23 J. W. Chin. Expanding and reprogramming the genetic code. *Nature* 2017, **550**, 53–60.

24 A. Crnković, O. Vargas-Rodriguez and D. Söll. Plasticity and constraints of tRNA aminoacylation define directed evolution of aminoacyl-tRNA synthetases. *Int. J. Mol. Sci.* , DOI:10.3390/ijms20092294.

25 M. J. Schmidt and D. Summerer. 2018, vol. 1728, pp. 97–111.

26 L. Wang and P. G. Schultz. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* 2001, **8**, 883–890.

27 T. Umehara, J. Kim, S. Lee, L.-T. Guo, D. Söll and H.-S. Park. N -Acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N -acetyl lysine specificity in vitro and in vivo. *FEBS Lett.* 2012, **586**, 729–733.

28 M. Amiram, A. D. Haimovich, C. Fan, Y.-S. Wang, H.-R. Aerni, I. Ntai, D. W. Moonan, N. J. Ma, A. J. Rovner, S. H. Hong, N. L. Kelleher, A. L. Goodman, M. C. Jewett, D. Söll, J. Rinehart and F. J. Isaacs. Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. *Nat. Biotechnol.* 2015, **33**, 1272–1279.

29 D. I. Bryson, C. Fan, L. T. Guo, C. Miller, D. Söll and D. R. Liu. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* 2017, **13**, 1253–1260.

30 J. M. Tharp and W. R. Liu. Synthetases pick up the pace. *Nat. Chem. Biol.* 2017, **13**, 1205–1206.

31 T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto and S. Yokoyama. Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* 2008, **371**, 818–822.

32 T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto and S. Yokoyama. Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode Nε-(o-Azidobenzyloxycarbonyl) lysine for Site-Specific Protein Modification. *Chem. Biol.* 2008, **15**, 1187–1197.

33 M. A. Mena and P. S. Daugherty. Automated design of degenerate codon libraries. *Protein Eng. Des. Sel.* 2005, **18**, 559–561.

34 M. T. Reetz and J. D. Carballeira. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* 2007, **2**, 891–903.

35 M. T. Reetz, S. Prasad, J. D. Carballeira, Y. Gumulya and M. Bocola. Iterative Saturation Mutagenesis Accelerates Laboratory Evolution of Enzyme Stereoselectivity: Rigorous Comparison with Traditional Methods. *J. Am. Chem. Soc.* 2010, **132**, 9144–9152.

9.5. Experimental Section

9.5.1. Rational PylRS* library generation:

A rational library of 20 PylRS* variants was generated using QuikChange Multi Site-Directed Mutagenesis Kit (Agilent) following manufacturer's protocol (**Table 9-3**).

Table 9-3: List of mutations of each PylRS* variant in the rationally generated library:

Residue*	271	274	313	349
*Mb*PylRS-wt	Tyr	Leu	Cys	Tyr
*Mb*PylRS* #1	Tyr	Met	Ala	Phe
*Mb*PylRS* #2	Ala	Met	Ala	Phe
*Mb*PylRS* #3	Tyr	Met	Cys	Phe
*Mb*PylRS* #4	Tyr	Met	Val	Phe
*Mb*PylRS* #5	Tyr	Met	Ala	Tyr
*Mb*PylRS* #6	Tyr	Met	Val	Tyr
*Mb*PylRS* #7	Ala	Met	Cys	Phe
*Mb*PylRS* #8	Ala	Met	Val	Phe
*Mb*PylRS* #9	Ala	Met	Ala	Tyr
*Mb*PylRS* #10	Ala	Met	Val	Tyr
*Mb*PylRS* #11	Tyr	Leu	Ala	Phe
*Mb*PylRS* #12	Ala	Leu	Ala	Phe
*Mb*PylRS* #13	Tyr	Leu	Cys	Phe
*Mb*PylRS* #14	Tyr	Leu	Val	Phe
*Mb*PylRS* #15	Tyr	Leu	Ala	Tyr
*Mb*PylRS* #16	Tyr	Leu	Val	Tyr
*Mb*PylRS* #17	Ala	Leu	Cys	Phe
*Mb*PylRS* #18	Ala	Leu	Val	Phe
*Mb*PylRS* #19	Ala	Leu	Ala	Tyr
*Mb*PylRS* #20	Ala	Leu	Val	Tyr

Briefly, commercially available pDule-Abk (Addgene), bearing a PylRS* with Leu274Met, Cys313Ala and Tyr349Phe mutations, was taken as the scaffold upon which mutagenic primers were generated to introduce the relevant mutations. After PCR amplification with the mutagenic primers and digestion of the parental plasmids with DpnI restriction enzyme, chemically competent *E. coli* DH5α cells were transformed by heat shock and plated in LB agar plates supplemented with tetracycline at 10 μg/mL final concentration. After 24 h incubation at 37 ºC, colonies

were grown overnight in LB supplemented with tetracycline at 10 μg/mL, and the plasmids were purified using Miniprep columns (Qiagen). Mutations were verified by DNA sequencing (Eurofins).

9.5.2. Randomly generated PylRS* library:

600 ng (6 μL of a 100 ng/μL stock solution in water) of the synthetic PylRS* gene (ThermoFisher) were mixed with the corresponding buffer M 10x (2 μL) and double distilled water (ddH_2O, 10 μL), and were double digested with KpnI and StuI restriction enzymes (Takara), 1 μL of each, for 1 hour at 37ºC. After performing a PCR clean-up protocol (Qiagen), the corresponding fragments were mixed with pDule-Abk plasmid previously digested with the same restriction enzymes (1 μg of pDule-Abk, 1.5 μL of a 650 ng/μL solution; 1 μL of KpnI and 1 μL StuI; 2 μL buffer M; 14.5 μL ddH_2O) for 1 hour at 37ºC, and subsequently purified through an agarose gel followed by a gel purification protocol (Qiagen). The ligation reaction was performed overnight at 16ºC, using 15 ng of the digested plasmid (5 μL of a 3 ng/μL solution), 45 ng of the PylRS* insert (5 μL of a 9 ng/μL solution), 1 μL of T4 DNA ligase, 2 μL of the corresponding T4 buffer 10x, and 7 μL of ddH_2O. The sequence of the synthetic gene, with each special feature, is included:

TATAGGTACCCTCGGGTTGTCAGCCTGTCCCGCTTATAAGATCATACGCCGTTATACGTT
GTTTACGCTTTGAGGAATCCCATATGGATAAAAAACCATTAGATGTTTTAATATCTGCGA
CCGGGCTCTGGATGTCCAGGACTGGCACGCTCCACAAAATCAAGCACCATGAGGTCTCAA
GAAGTAAAATATACATTGAAATGGCGTGTGGAGACCATCTTGTTGTGAATAATTCCAGG
AGTTGTAGAACAGCCAGAGCATTCAGACATCATAAGTACAGAAAAACCTGCAAACGATG
TAGGGTTTCGGACGAGGATATCAATAATTTTCTCACAAGATCAACCGAAAGCAAAAACA
GTGTGAAAGTTAGGGTAGTTTCTGCTCCAAAGGTCAAAAAAGCTATGCCGAAATCAGTT
TCAAGGGCTCCGAAGCCTCTGGAAAATTCTGTTTCTGCAAAGGCATCAACGAACACATCC
AGATCTGTACCTTCGCCTGCAAAATCAACTCCAAATTCGTCTGTTCCCGCATCGGCTCCTG
CTCCTTCACTTACAAGAAGCCAGCTTGATAGGGTTGAGGCTCTCTTAAGTCCAGAGGATA
AAATTTCTCTAAATATGGCAAAGCCTTTCAGGGAACTTGAGCCTGAACTTGTGACAAGAA
GAAAAAACGATTTTCAGCGGCTCTATACCAATGATAGAGAAGACTACCTCGGTAAACTCG
AACGTGATATTACGAAATTTTTCGTAGACCGGGGTTTTCTGGAGATAAAGTCTCCTATCC
TTATTCCGGCGGAATACGTGGAGAGAATGGGTATTAATAATGATACTGAACTTTCAAAA
CAGATCTTCCGGGTGGATAAAAATCTCTGCTTGAGGCCAATGCTTGCCCCGACTCTGNNK
AACTATNNKCGAAAACTCGATAGGATTTTACCAGGCCCAATAAAAATTTTCGAAGTCGG
ACCTTGTTACCGGAAAGAGTCTGACGGCAAAGAGCACCTGGAAGAATTTACTATGGTGA
ACTTCNNKCAGATGGGTTCGGGATGTACTCGGGAAAATCTTGAAGCTCTCATCAAAGAG
TTTCTGGACTATCTGGAAATCGACTTCGAAATCGTAGGAGATTCCTGTATGGTCNNKGG
GGATACTCTTGATATAATGCACGGGGACCTGGAGCTTTCTTCGGCAGTCGTCGGGCCAGT

TTCTCTTGATAGAGAATGGGGTATTGACAAACCATGGATAGGTGCAGGTTTTGGTCTTG
AACGCTTGCTCAAGGTTATGCACGGCTTTAAAAACATTAAGAGGGCATCAAGGTCCGAAT
CTTACTATAATGGGATTTCAACCAATCTATAACTGCAGTTTCAAACGCTAAATTGCCTGA
TGCGCTACGCTTATCAGGCCTTAAT

Green: Extra clumps for restriction enzyme

Yellow: Restriction sites (KpnI and StuI)

Blue: Non-ORF plasmid sequence

Red: Degenerate nucleotides

The ligated product was transformed in DH5α *E. coli* cells and grown for 16 hours at 37ºC with shaking in LB media supplemented with tetracycline at 10 µg/mL final concentration. A plasmid purification Maxiprep (Qiagen) was performed following manufacturer's protocol, yielding 500 µL of DNA material at ca. 1000 ng/µL. To check for correct cloning, the resulting plasmid was sequenced, yielding the expected sequence together with a mixture of codons in the NNK positions as expected.

9.5.3. Positive selection assay

9.5.3.1. Bacterial transformation for the small PylRS* library:

100 µL of commercially available *E. coli* BL21 (DE3) chemically competent cells (ThermoFisher) were co-transformed with 150 ng of the mixture of 20 pDule-PylRS* mutants and 150 ng of the selection plasmid (pBAD-GFP150TAG_Amp183TAG) by heat shock. After 1 hour at 37 ºC in recovery media, bacteria were seeded in the corresponding LB agar plates: ***positive control*** (Tet^+ Kan^+), ***negative control*** (Tet^+ Kan^+ Amp^+) and ***problem*** (Tet^+ Kan^+ Amp^+ $ncAA^+$). Used concentrations were: 10 µg/mL for tetracycline, 25 µg/mL for kanamycin, and 50 (1x) or 25 (0.5x) µg/mL for ampicillin. Pictures and colony count were taken 40 h post-incubation at 37 ºC.

9.5.3.2. Bacterial transformation for the randomly generated PylRS* library:

500 µL of commercially available *E. coli* BL21 (DE3) chemically competent cells (ThermoFisher) were co-transformed with 2500 ng of the randomly generated pDule-PylRS* library and 2500 ng of the selection plasmid (pBAD-GFP150TAG_Amp183TAG) by heat shock. Then, the protocol was followed as in the previous section.

9.5.3.3. Colony isolation and PylRS* variant identification by DNA sequencing:

A colony PCR protocol was followed with 10 randomly chosen colonies in the ***problem plate***. The obtained PCR products were cleaned up using a PCR Purification Kit (Qiagen) and i) loaded onto a 1% agarose gel and ii) DNA sequenced (Eurofins) to identify the positively selected mutations.

9.5.4. *M. barkeri* and *M. mazei* PylRS alignment:

To use the crystal structure of *Mm*PylRS as a scaffold for *Mb*PylRS mutagenesis, an alignment of both was performed in T-COFFEE Multiple Sequence Alingment software.

```
Mbarkeri   ---------------------------------MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIY
Mmazei     MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARALRHHKYRKTCKRC

cons                                        :  :       ...   *   :    :*:  . .*

Mbarkeri   IEMACGDHLVVNNSRSCRTARAFRHHKYRKTCKRCRVSDEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPK
Mmazei     RVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQE

cons            . : .:.::.. :*:  .:  .  .  *:.  ..        ..**: ::       :* :  :  :

Mbarkeri   SVSRAPKPLENSVSAKASTNTSRSVPSPAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDKISLNMAKPF
Mmazei     SVSVPASVSTSISSISTGATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPF

cons       *** ...   .  * .:.:.:*  * . ::. .. *.*..*.*.*:**:** **:*.**.*:*:**** .***

Mbarkeri   RELEPELVTRRKNDFQRLYTNDREDYLGKLERDITKFFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQI
Mmazei     RELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQI

cons       ****.**:.***:*:*::*:::**:*******:**:***************** **:*****:*********

Mbarkeri   FRVDKNLCLRPMLAPTLYNYLRKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVNFCQMGSGCTRENLE
Mmazei     FRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLE

cons       ******:********.********** **.******:******************:***************

Mbarkeri   ALIKEFLDYLEIDFEIVGDSCMVYGDTLDIMHGDLELSSAVVGPVSLDREWGIDKPWIGAGFGLERLLKVMH
Mmazei     SIITDFLNHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKH

cons       ::*.:**::* ***:***************:**************:.********************** *

Mbarkeri   GFKNIKRASRSESYYNGISTNL
Mmazei     DFKNIKRAARSESYYNGISTNL

cons       .*******:*************
```

Figure 9-9: T-COFFEE alignment of *M. barkeri* and *M. mazei* PylRS wild-type. The difference of 35 residues in the annotation of both enzymes is due to the missing initial stretch in *Mb*PylRS. The consensus sequence is highlighted in red.

Chapter 10

Conclusions

This PhD dissertation has been divided in three main blocks: firstly, the novel synthesis of differently substituted 3-bromo-1,2,4,5-tetrazines and their use as reagents in S_NAr and Pd-catalyzed cross-coupling reactions; secondly, the development of a simple strategy to attain chemical protein modification with tetrazines and the application of the obtained proteins as bioorthogonal activators of prodrugs; and finally, an advanced method to generate tetrazinyl-labeled proteins based on the manipulation of the genetic code of certain organisms.

* * *

The synthesis of 1,2,4,5-tetrazines is challenging, and relies on methods that usually employ high temperatures, strong oxidation conditions and metal catalysis. Thus, late-stage tetrazine functionalization is sometimes inefficient due to functional group incompatibilities in the required reaction conditions.

- ✓ We have developed an alternative method for the synthesis of 3-bromo-1,2,4,5-tetrazines that is free from oxidants and metal catalysis, does not generate volatile intermediates, and where all reactions proceed at room temperature or 55 ºC. Altogether, this has enabled the synthesis of the previously unreported simplest bromotetrazine, namely 3-bromo-1,2,4,5-tetrazine (**2**), and the production of 3-bromo-6-methyl-1,2,4,5-tetrazine (**1**) and 3-bromo-6-phenyl-1,2,4,5-tetrazine (**3**) in higher overall yields.
- ✓ 3-bromotetrazines **1** and **2** are excellent electrophilic reagents to undergo S_NAr reactions with different aliphatic and aromatic heteroatoms acting as nucleophiles under mild conditions (compounds **8a-f** and **9a-f**). A set of tetrazine-containing amino acids (**10a-e** and **11a-e**) have been synthesized and used to unravel the kinetics of the subsequent iEDDA cycloaddition with a model strained cyclooctyne, revealing a substituent effect in the speed of the reaction, with tetrazine ethers being the most reactive.
- ✓ Tetrazine ethers also exhibit a fluorescent behavior compared to tetrazines with other substituents. We have exploited this feature to incorporate the tyrosine derivative **10d** into a model peptide by solid-phase peptide synthesis and used it in cell culture as a dual fluorescent and clickable peptide that loses

its emissive properties upon iEDDA cycloaddition, showing for the first time their applicability *in vivo*.

- ✓ 3-Bromotetrazines **1** and **3** are suitable reagents to undergo Sonogashira-type cross-coupling reactions directly at position 3 of the tetrazine ring. The corresponding alkynyl-tetrazines (**12a-p**) are underrepresented compounds in the literature that can be exploited to develop new interesting applications. However, we particularly used them as intermediates for the synthesis of alkyl-tetrazines, a highly challenging class of compounds, via a hydrogenation and re-oxidation protocol (**14b,f**). Using this methodology, different unnatural amino acids bearing the alkynyl- and alkyl-tetrazine moiety (**17a,b** and **19a,b**) or potential compounds to be used as CLIPTACs (**20a-d**) have been prepared.

In order to fully exploit the applicability of tetrazines, their incorporation into proteins is an attractive goal to expand protein function beyond the chemical repertoire endowed by natural amino acids. Therefore, the simplest way to modify proteins is through the reactivity of certain amino acid side chains.

- ✓ We have successfully used 3-bromotetrazine **2** as a reagent for protein labeling. LC-MS/MS studies on a model protein (RNAse A) labeled with different equivalents of **2** under different pH values demonstrated a selective preference of **2** towards lysines (*chemoselectivity*), thereby resulting in a homogeneous population of 3-aminotetrazines at the protein surface.
- ✓ We have demonstrated that the formed tetrazinyl-lysines can undergo *click-to-release* reactions with suitable *trans*-cyclooctene carbamates precursors to liberate cargos of choice. In particular, we have used tetrazinyl-lysine **11a** as a model to determine the reaction parameters with **TCO-Dox**, a doxorubicin derivative acting as a prodrug. In physiological conditions (human serum at 37 ºC), ca. 50% liberation was accomplished after 15 h post-reaction.
- ✓ To showcase a potential application of these modified proteins with tetrazines to trigger *click-to-release* reactions *in vivo*, we have labeled the therapeutically relevant monoclonal antibody Trastuzumab and performed a bioorthogonal prodrug activation experiment in cell culture. We envisage that our approach can lead to the development of targeted drug delivery methods combining the

antigen-specific behavior of certain proteins with the ability of tetrazines to trigger bioorthogonal activation of prodrugs.

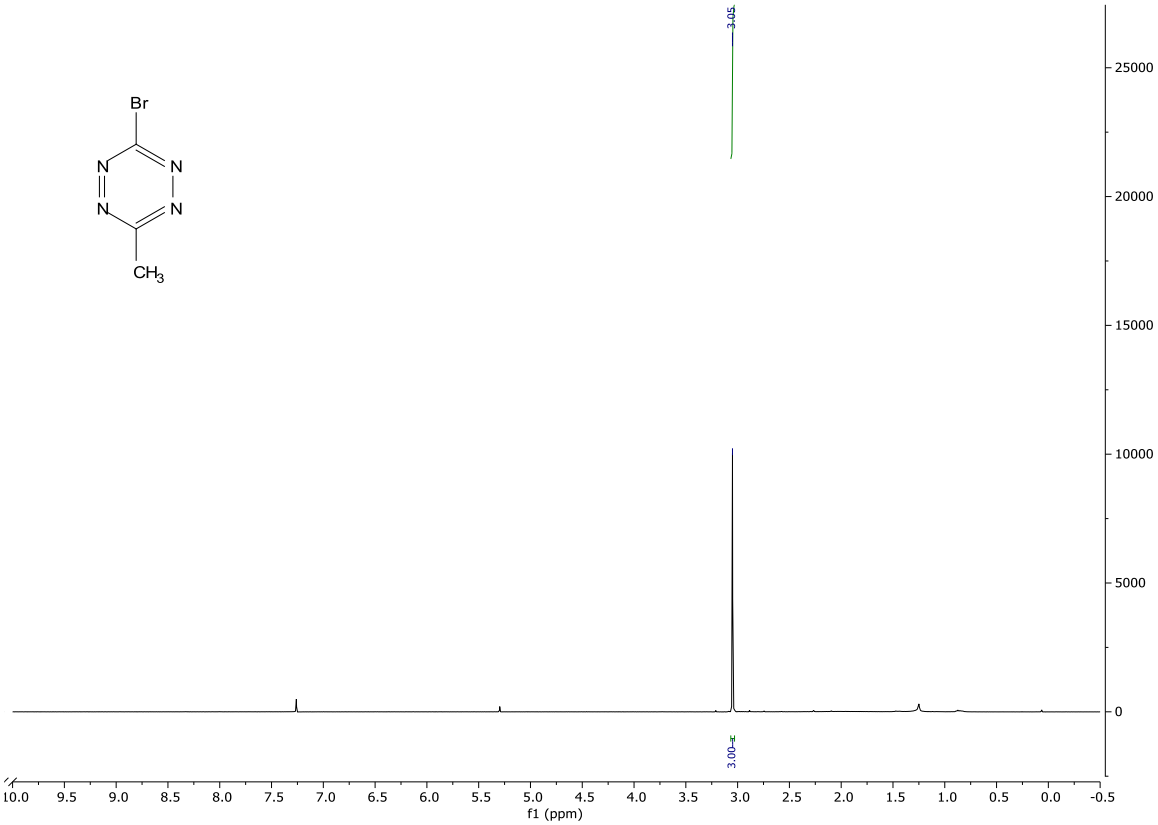
An advanced method for protein modification is the genetic incorporation of unnatural amino acids *in vivo* through the manipulation of the translation machinery of a given organism to expand its genetic code beyond the 22 proteinogenic amino acids. Despite the challenging endeavor it represents, the selectivity in the site of incorporation and the minimal perturbation of the generated proteins are desirable features for subsequent applications.

- ✓ We have been able to genetically encode the Phenylalanine derivative **11b** as a ncAA in *E. coli* using a previously existing TyrRS*/tRNA pair from the archaeon *Methanocaldococcus jannaschii*. The use of a tailormade autoinduction expression media lacking Tyrosine and Cysteine was critical to reduce competing ligands for the enzyme and achieve expression of a model protein with **11b** incorporated at a specific position (GFP-**11b**).
- ✓ The ncAA **11b** can also trigger bioorthogonal *click-to-release* reaction with **TCO-Dox**, which can be a useful attribute to produce nanobodies bearing a genetically encoded **11b** able to activate prodrugs specifically at the site of action.
- ✓ In order to genetically incorporate a ncAA bearing a tetrazine functionality in eukaryotic organisms, we aimed to use the PylRS*/tRNA pair form the archaeon *Methanosarcina barkeri*. Even though we were unable to isolate an active orthogonal pair for the tetrazine-containing amino acids we tested, we developed a positive selection assay that was used with either a small library of 20 representative enzyme variants or with a randomized library prepared using site-saturation mutagenesis. In both cases, we could identify active PylRS* mutants towards **N$^{\varepsilon}$-Boc-Lys**, which we used as a model ncAA, meaning that this methodology could be further utilized in directed evolution experiments.

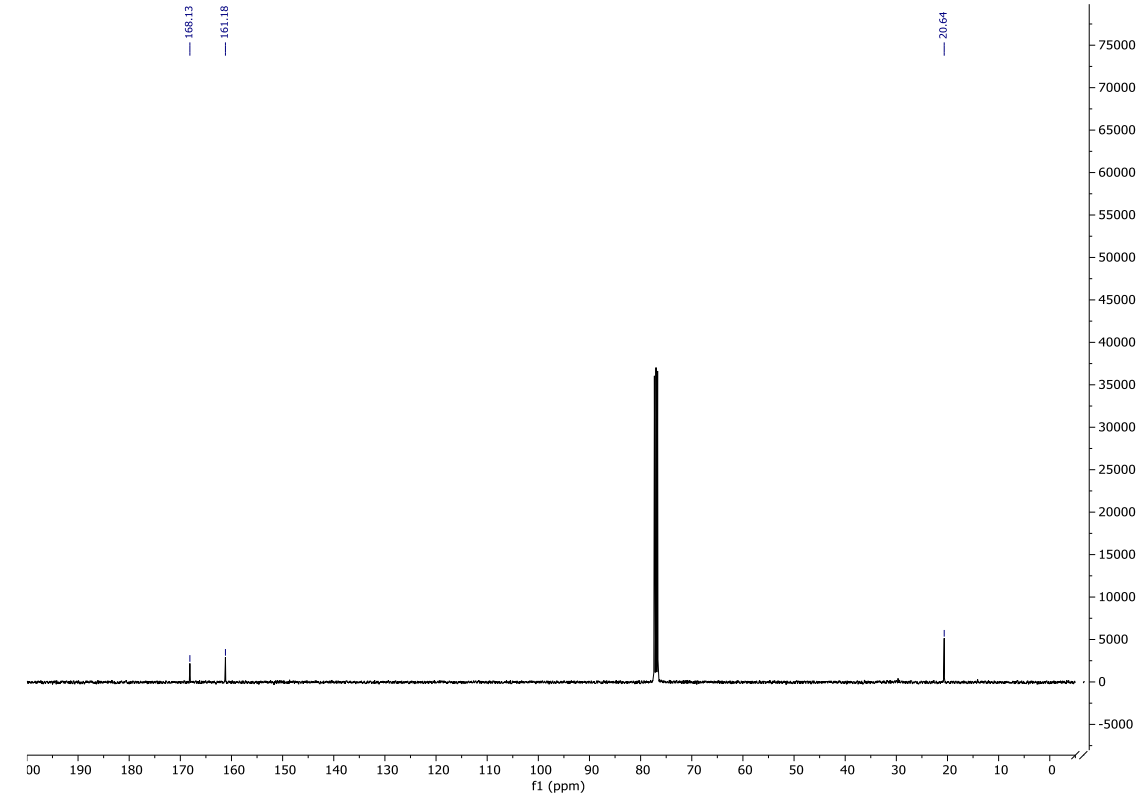
Annex I

Selected NMR spectra

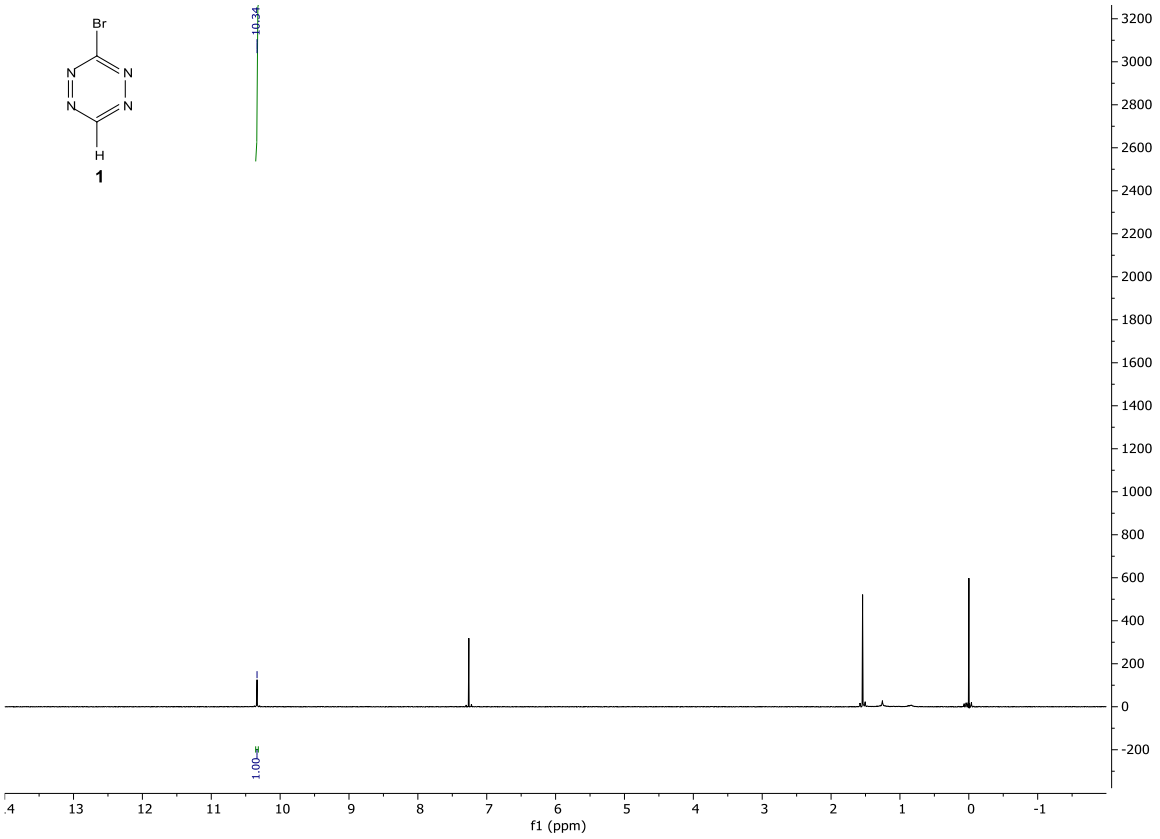
^{1}H NMR (400 MHz, $CDCl_3$) of **1**



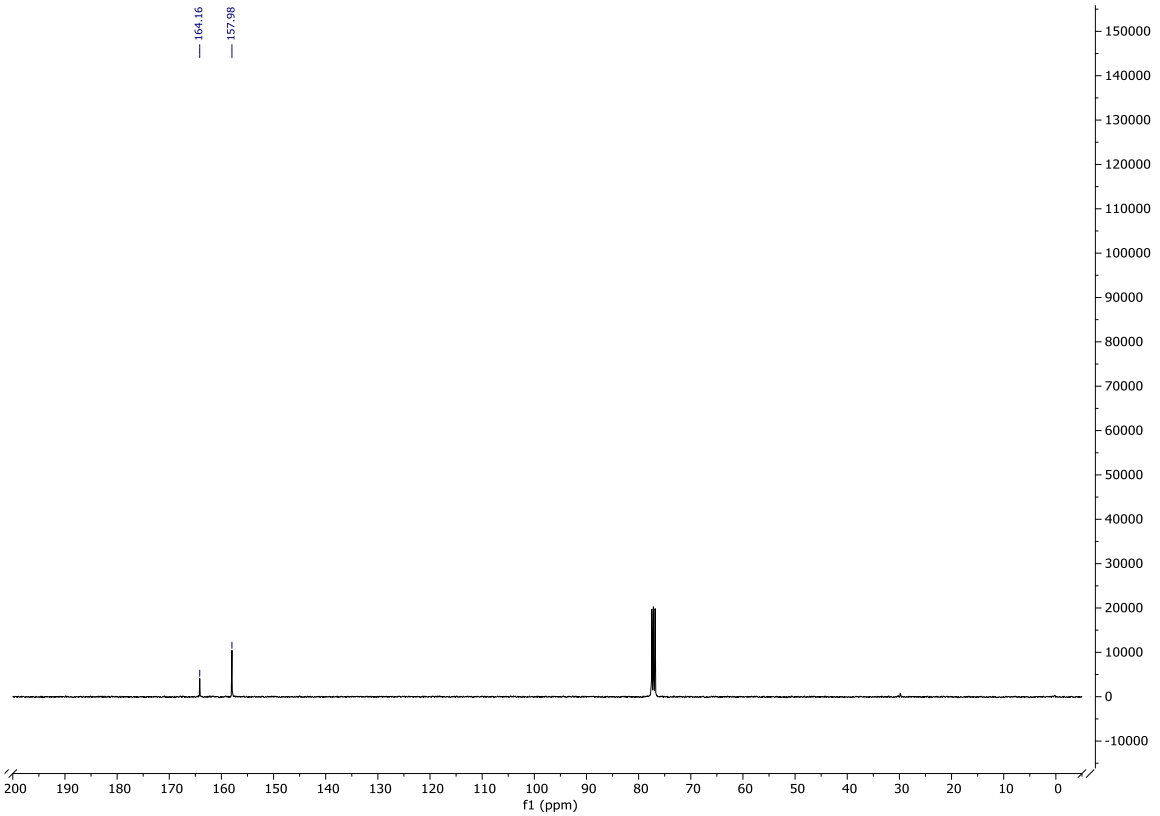
^{13}C NMR (101 MHz, $CDCl_3$) of **1**



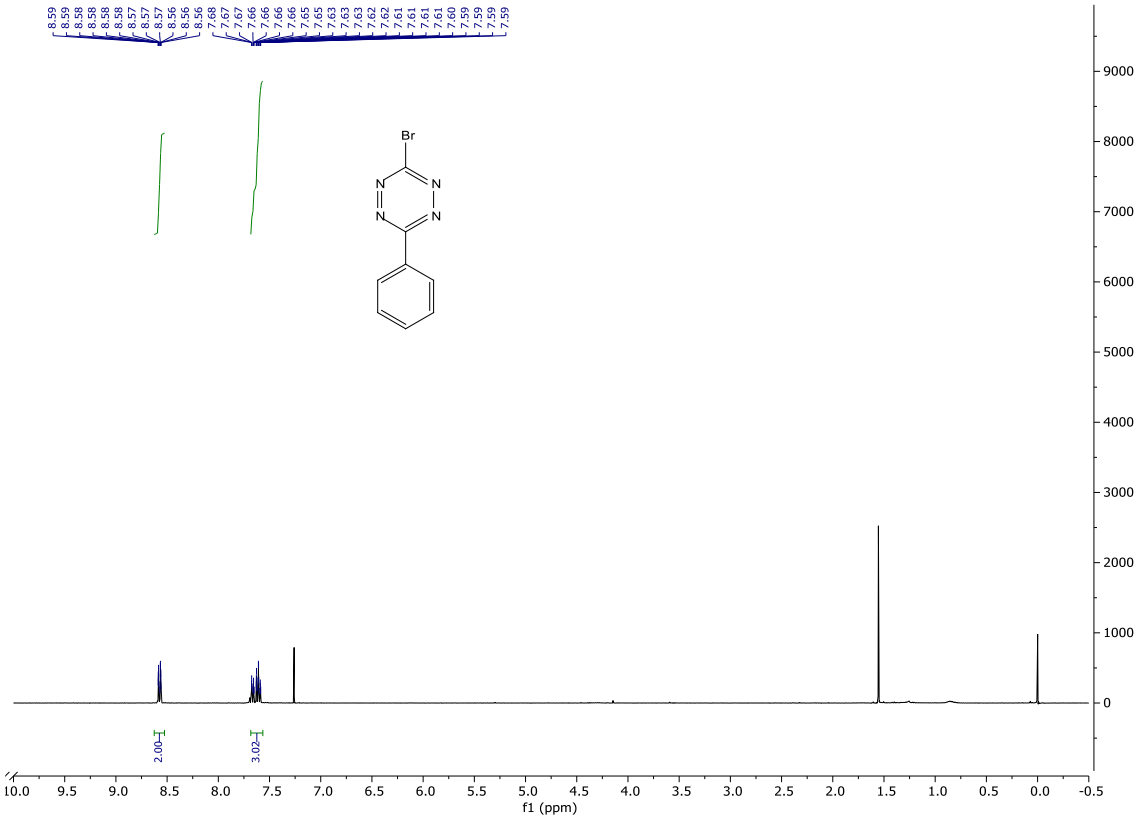
^{1}H NMR (400 MHz, $CDCl_3$) of **2**



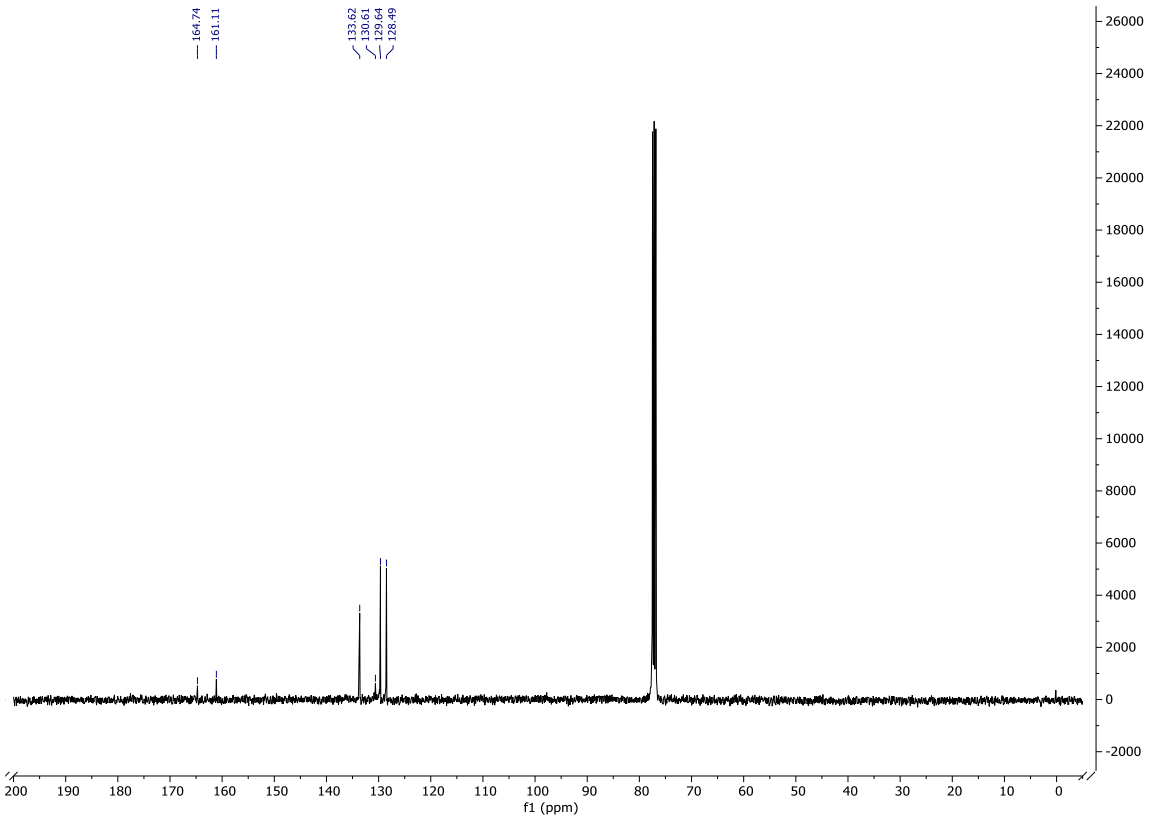
^{13}C NMR (101 MHz, $CDCl_3$) of **2**



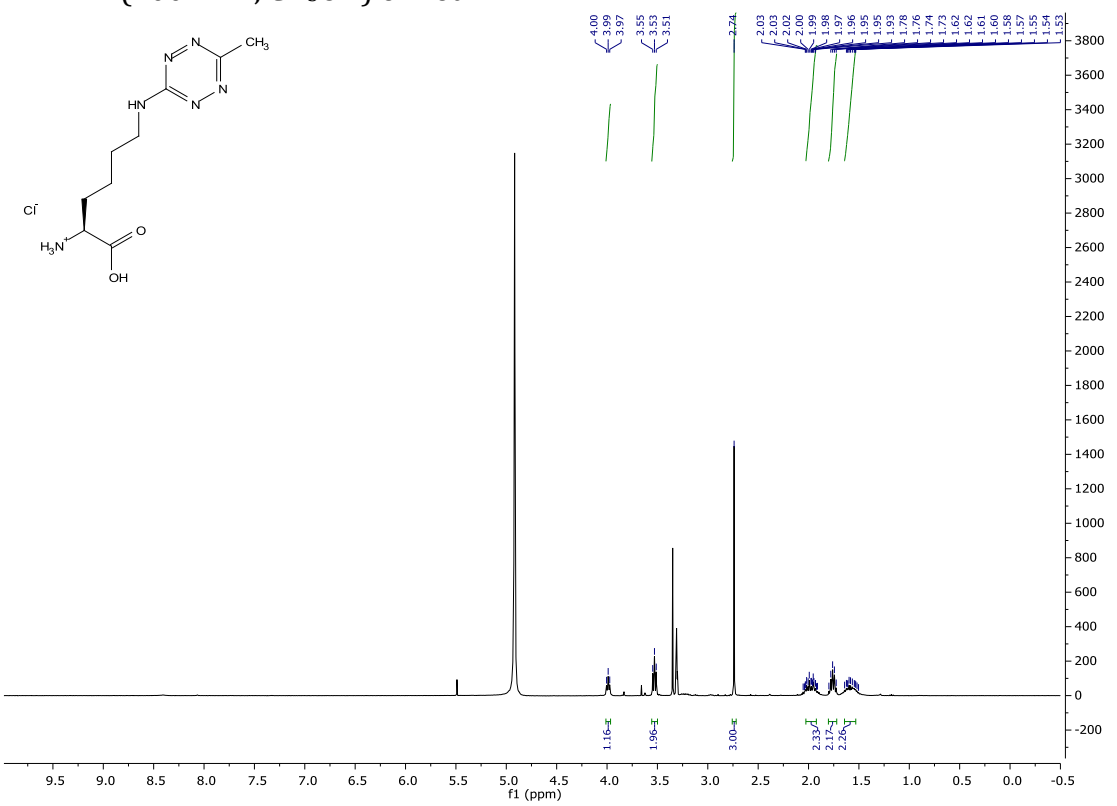
^{1}H NMR (400 MHz, $CDCl_3$) of **3**



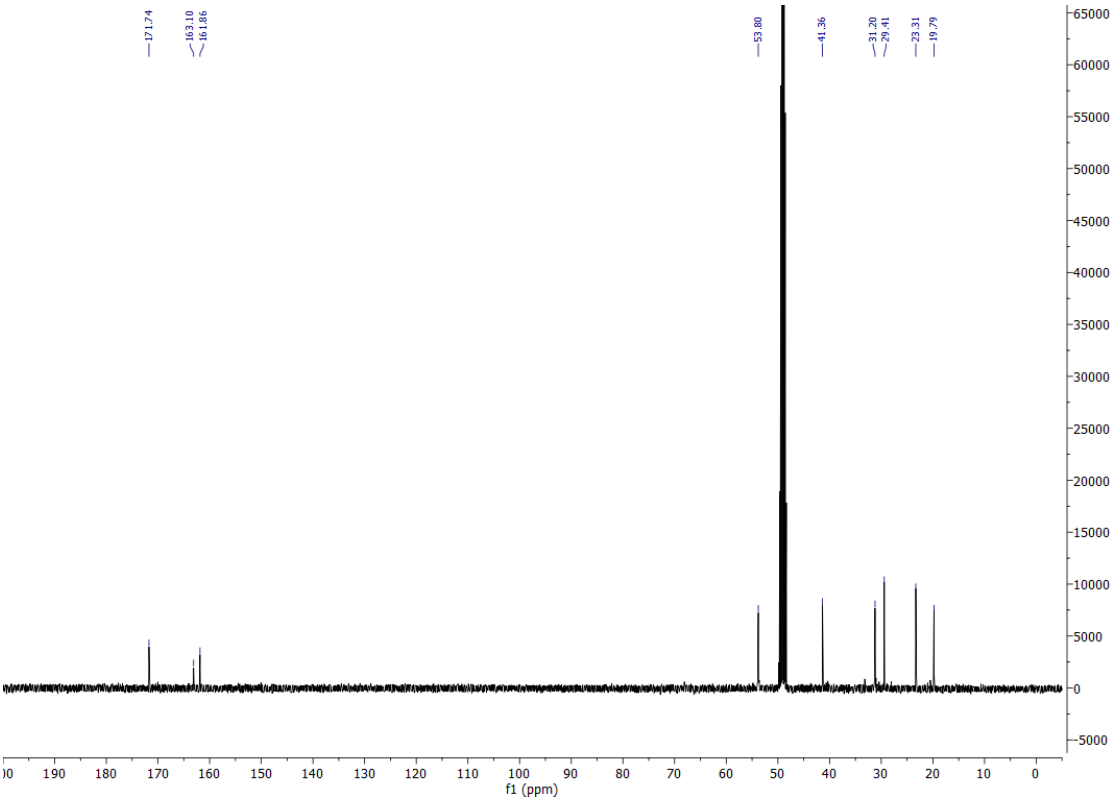
^{13}C NMR (101 MHz, $CDCl_3$) of **3**



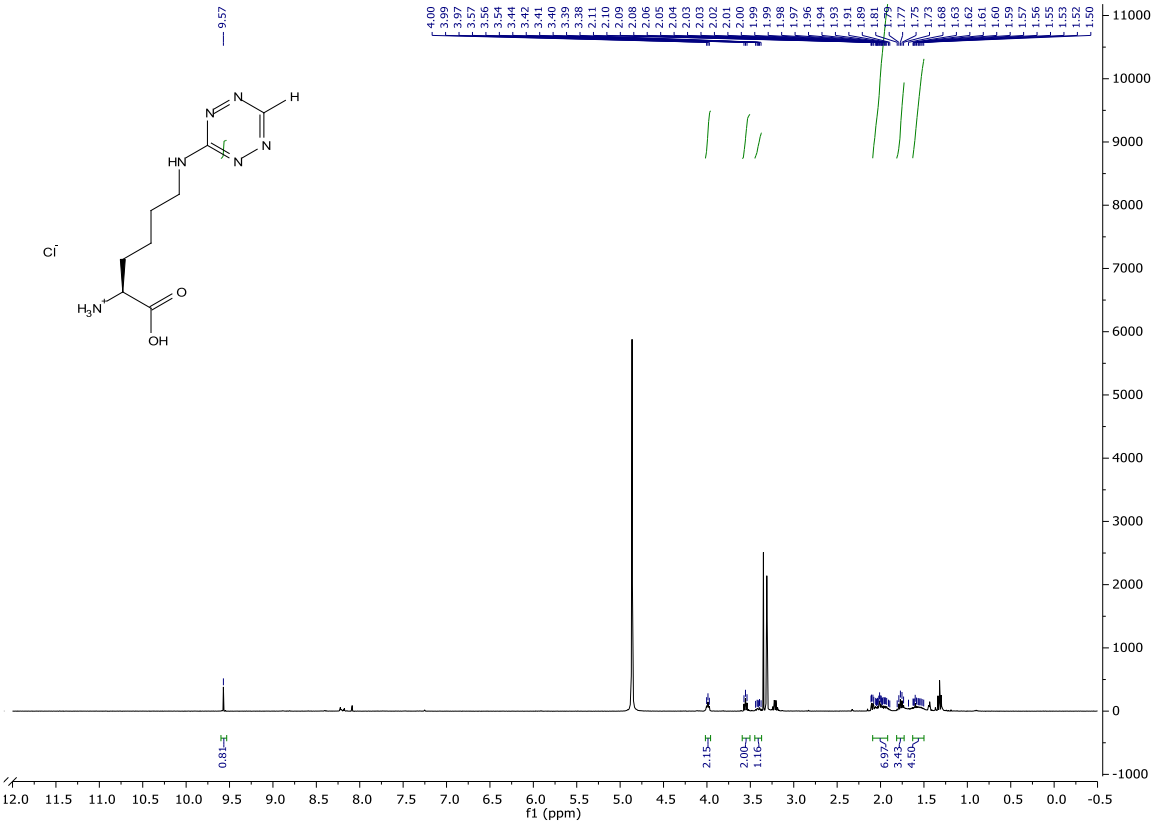
^{1}H NMR (400 MHz, CD_3OD) of **10a**



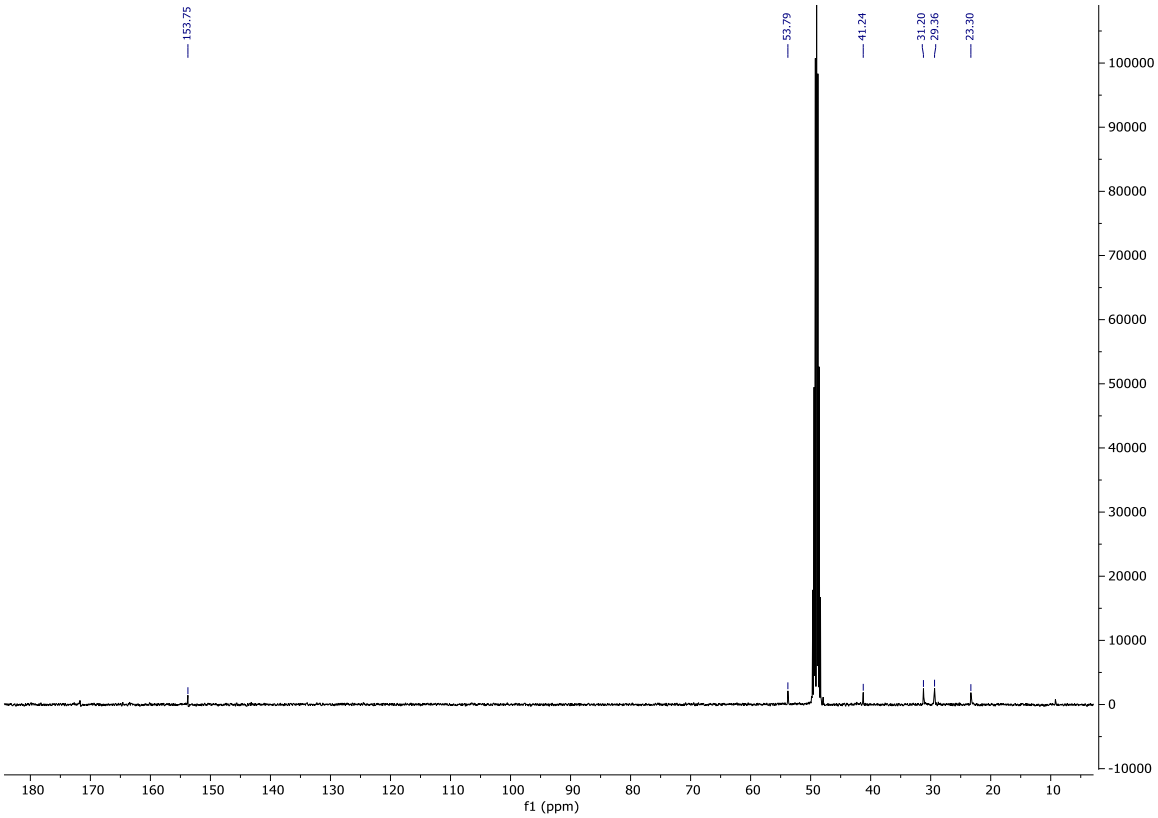
^{13}C NMR (101 MHz, CD_3OD) of **10a**



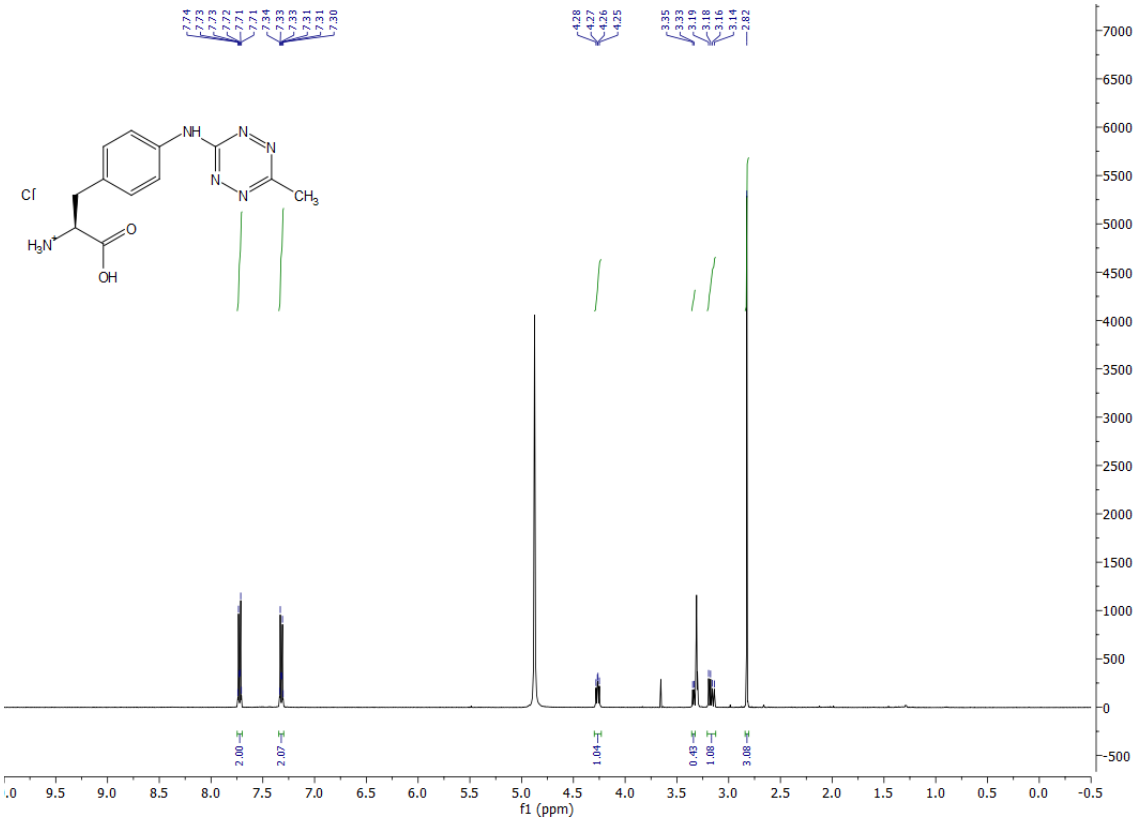
^{1}H NMR (400 MHz, CD_3OD) of **11a**



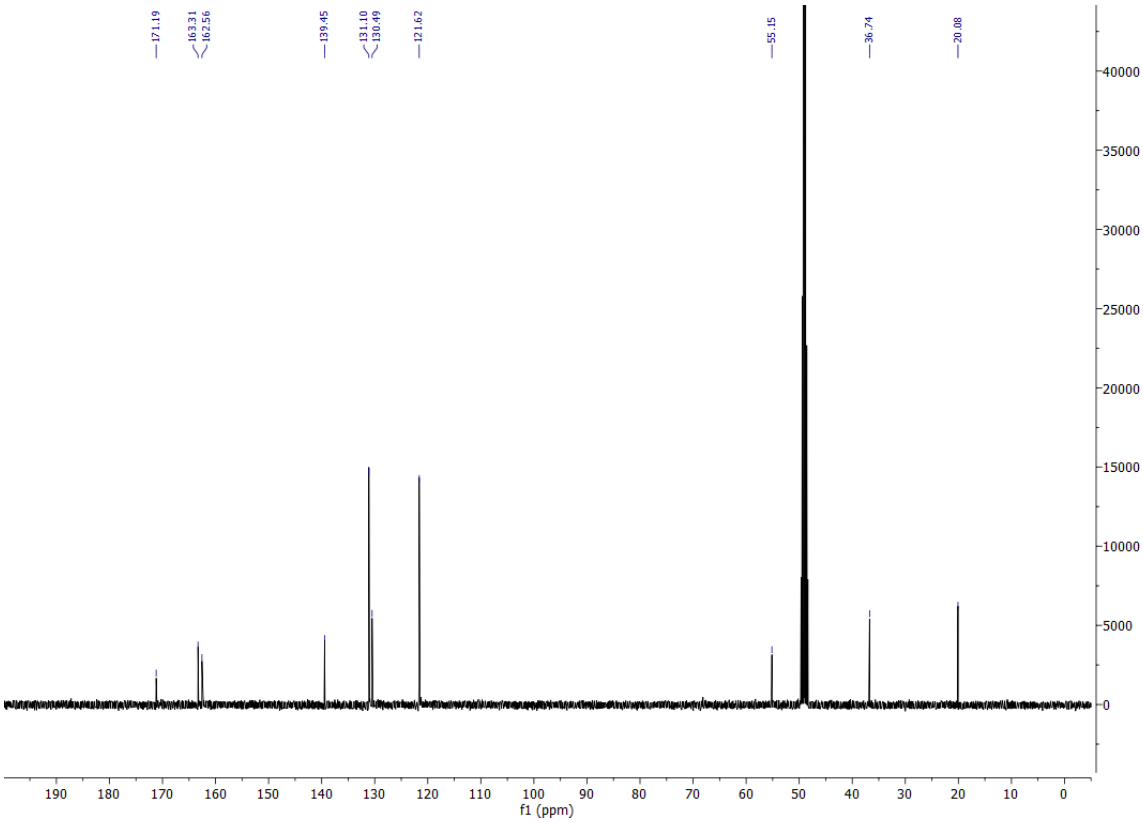
^{13}C NMR (101 MHz, CD_3OD) of **11a**



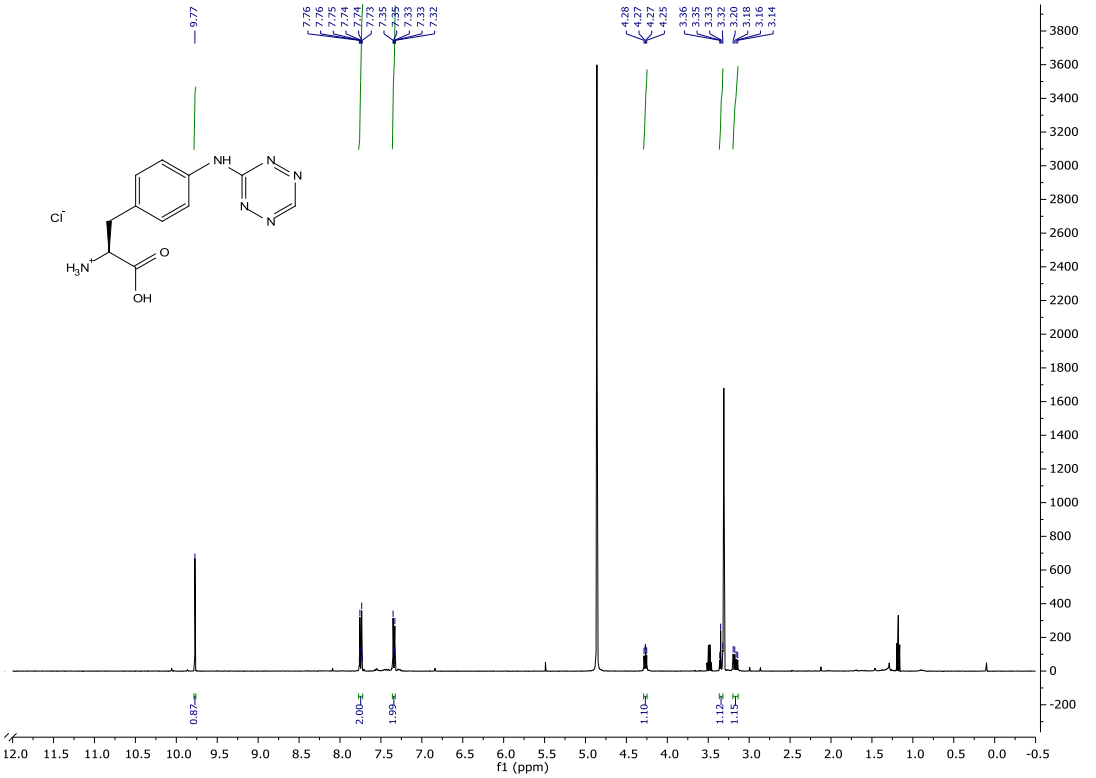
^{1}H NMR (400 MHz, CD_3OD) of **10b** (or **Tet-v1.0**)



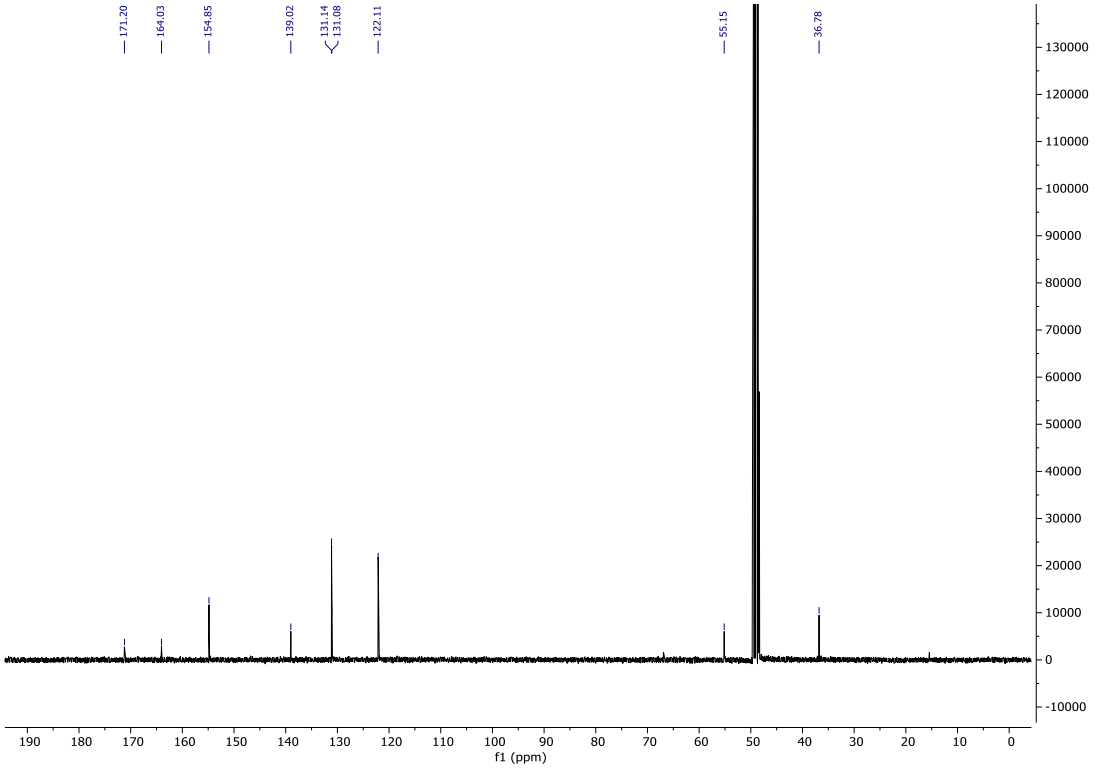
^{13}C NMR (101 MHz, CD_3OD) of **10b** (or **Tet-v1.0**)



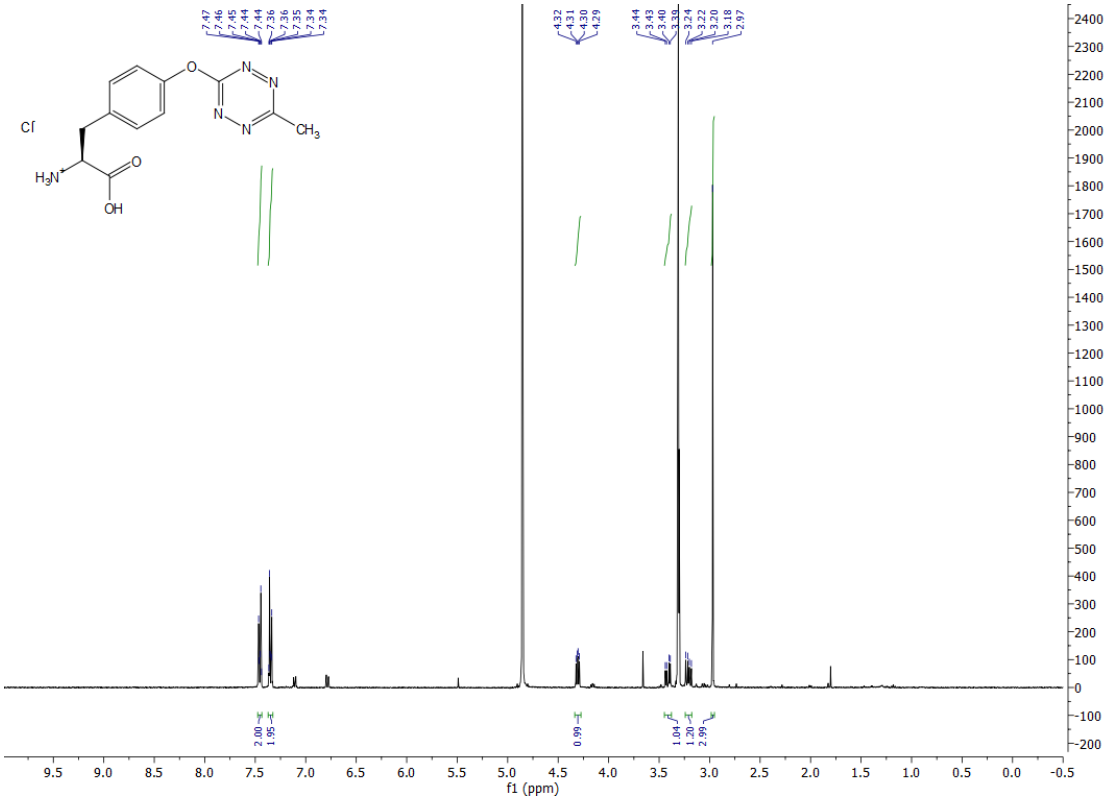
^{1}H NMR (400 MHz, CD_3OD) of **11b**



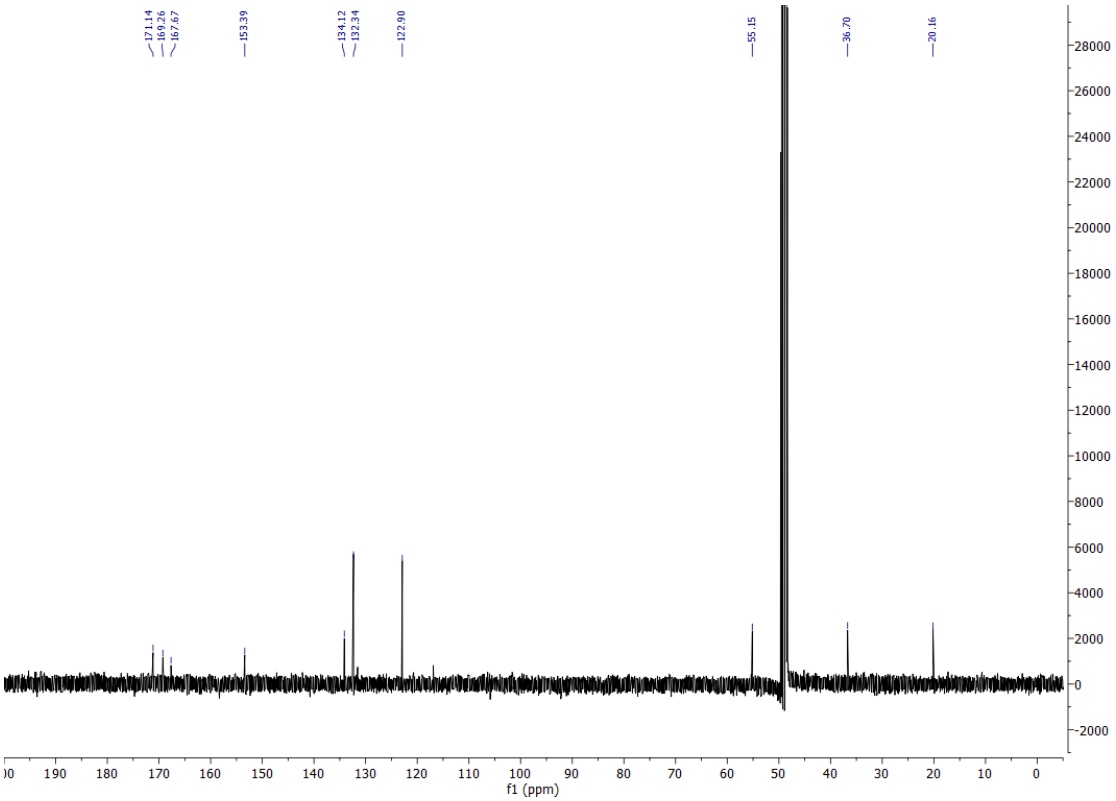
^{13}C NMR (101 MHz, CD_3OD) of **11b**

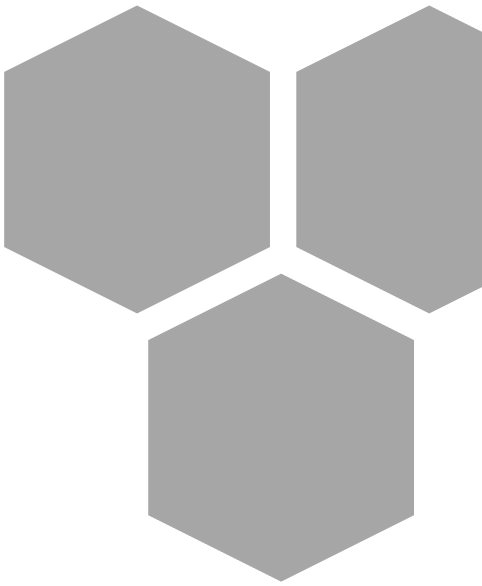


^{1}H NMR (400 MHz, CD_3OD) of **10d**



^{13}C NMR (101 MHz, CD_3OD) of **10d**





Annex II

Publications

Bioconjugate Chemistry

pubs.acs.org/bc

Article

Synthesis and Application of 3-Bromo-1,2,4,5-Tetrazine for Protein Labeling to Trigger Click-to-Release Biorthogonal Reactions

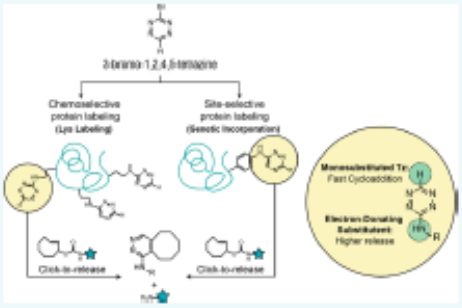
Enric Ros, Marina Bellido, Xavier Verdaguer, Lluís Ribas de Pouplana,* and Antoni Riera*

Cite This: *Bioconjugate Chem.* 2020, 31, 933−938

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: 3-Bromo-1,2,4,5-tetrazine has been synthesized in an oxidant- and metal-free method. The synthesis is scalable and relies on inexpensive starting materials. 3-Bromo-1,2,4,5-tetrazine can undergo nucleophilic aromatic substitutions with differently substituted heteroatoms under mild conditions. In particular, its excellent reactivity has been used to attain chemoselective protein labeling. The resulting labeled lysines can react with strained dienophiles to trigger fast click-to-release (CtR) biorthogonal reactions. The characterization of the CtR reaction in physiological conditions and a therapeutically relevant example with the monoclonal antibody Trastuzumab to showcase its application is presented. Finally, 3-bromo-1,2,4,5-tetrazine has been used to achieve site-selective protein labeling through the genetic incorporation of the first unnatural amino acid bearing an unsubstituted 1,2,4,5-tetrazin-3-yl functionality, which can also undergo CtR reactions.



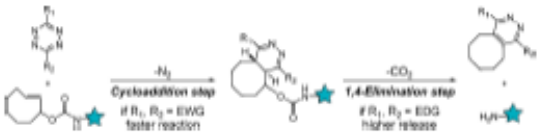
INTRODUCTION

Biorthogonal reactions have had a major impact on the development of chemical biology tools[1,2] and therapeutic strategies.[3,4] In the latter case, the use of these types of reactions to activate prodrugs with spatiotemporal control, employing a reaction commonly referred to as click-to-release (CtR), has enabled several new drug delivery approaches.[5,6] Out of the different biorthogonal reactions developed for triggered release, the most convenient in terms of selectivity, exceptional kinetics, and absence of metal catalysts is based on the inverse electron-demand Diels−Alder (iEDDA) cycloaddition[7] that takes place between a 1,2,4,5-tetrazine fragment (Tz) and a range of dienophiles, namely, vinyl ethers,[8] *trans*-cyclooctene,[3] or benzonorbornadiene[9] derivatives.

The CtR reaction consists of two steps: an initial cycloaddition step, in which the Tz fragment reacts with the dienophile (usually a strained alkene for biorthogonal purposes), followed by an elimination step, in which a cargo linked to the dienophile is released. The parameters influencing these reactions have been intensely studied in recent years.[10,11] It is well stablished that electron-withdrawing groups (EWGs) on the Tz ring favor the initial cycloaddition by decreasing the lowest unoccupied molecular orbital (LUMO) of the Tz.[12] However, EWGs also limit the extent to which the cargo is released in the second step[11] (Scheme 1).

Monosubstituted Tz are known to react much faster than disubstituted ones except in the case of two EWG substituents. For instance, monosubstituted tetrazines react much faster than tetrazines with the same substituent and an additional methyl group. This effect was shown to be partially due to electronic and steric effects,[13] but also because of distortion energies arising from the aliphatic substituent.[14]

Scheme 1. Click-to-Release Reaction between a 1,2,4,5-Tetrazine and *trans*-Cyclooct-2-en-1-yl Carbamate



In light of these observations, we hypothesized that a monosubstituted Tz with an EDG would be an ideal candidate to act as a biorthogonal releaser. The small steric effect would speed up the cycloaddition reaction, whereas the EDG would favor the elimination step. For this reason, we directed our efforts to finding an electrophilic monosubstituted Tz reagent that could be easily used in late-stage functionalization of biologically relevant biomolecules, with a special emphasis on protein labeling.

In previous reports where the CtR reaction was employed using a protein as a carrier for targeted delivery, *trans*-cyclooct-

Received: January 28, 2020
Revised: February 12, 2020
Published: February 14, 2020

https://dx.doi.org/10.1021/acs.bioconjchem.0c00052
Bioconjugate Chem. 2020, 31, 933−938

2-en-1-yl (TCO) linked through a carbamate to the cargo of interest was directly bioconjugated to the protein, and the activator Tz was administered in a subsequent step (Figure 1, top). The substituents on the tetrazine ring varied from the simplest dimethyl-Tz[3] to more elaborate conjugates to increase the clearance time *in vivo* (Figure 1, top).[15,16]

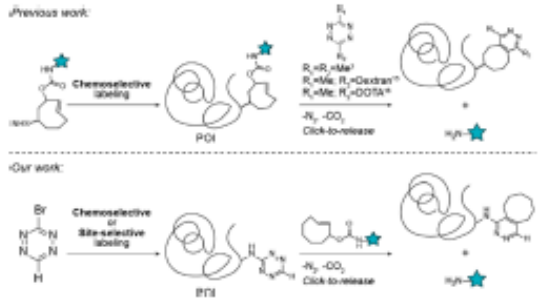


Figure 1. Overview of the strategies to achieve targeted drug release via the CtR reaction. Top: TCO-labeled proteins and Tz used as activators to trigger the drug release (blue star). Bottom: incorporation of monosubstituted amino-Tz to activate a circulating inactive TCO-drug conjugate (this work). NHS: N-Hydroxysuccinimide; POI: Protein of interest; DOTA: Tetraazacyclododecane-1,4,7,10-tetraacetic acid.[16]

These TCO−drug conjugates were typically incorporated in the protein of interest through N-hydroxysuccinimide (NHS) esters, the classic functional group employed for chemoselective lysine labeling. While this is a suitable approach,[17] the synthesis of TCO−carbamates bearing NHS esters for bioconjugation is challenging and typically low-yielding.[15,16] On the other hand, the site-selectivity is still a challenge that could be overcome through the genetic incorporation of unnatural amino acids by amber codon reassignment.[18,19]

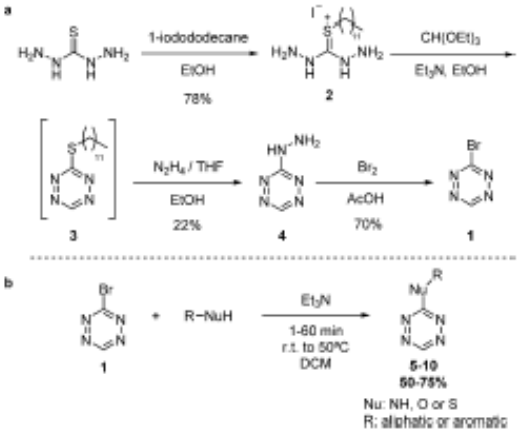
We considered that using the Tz for protein labeling rather than the strained dienophile would be beneficial to trigger the CtR reaction. The dienophiles commonly used in CtR biorthogonal reactions tend to be bulky, apolar molecules that can be masked by hydrophobic interactions with the labeled protein,[20] and in the case of TCO itself, isomerization to the *cis* form can take place under physiological conditions,[21] both factors limiting the release yield. Instead, a monosubstituted Tz would be a polar fragment with a tendency for being at the solvent exposed region of the labeled protein. Moreover, the stability *in vivo* can be tuned depending on their substituents, the EDG-substituted Tz being much more stable and reactive on the release reaction than those with EWGs.[22]

Herein we describe the synthesis of a Tz reagent, 3-bromo-1,2,4,5-tetrazine **1**, and its use in chemoselective protein labeling. We demonstrate the capability of the formed monosubstituted Tz-tagged proteins to undergo fast CtR reactions, and their application in a therapeutically relevant context. Finally, we expand the use of **1** to achieve a site-selective protein labeling by genetic incorporation of the first ever noncanonical amino acid bearing a monosubstituted Tz, opening a new way to produce proteins able to trigger drug release *in vivo*.

RESULTS AND DISCUSSION

Our initial goal was to synthesize a suitable Tz synthon that enabled the late-stage functionalization of a wide range of nucleophilic groups to form the corresponding monosubstituted Tz. We envisaged that **1** was a fitting candidate to undergo S_NAr reactions due to the high electrophilic character of position 3, largely known for various halogenated Tz.[23] We designed the synthetic route of **1** shown in Scheme 2a. We

Scheme 2. Synthetic Route to 3-Bromo-1,2,4,5-Tetrazine (1) and S_NAr Reactivity with Nucleophiles



started with the alkylation of thiocarbohydrazide with 1-iodododecane. We reasoned that the long dodecyl chain would prevent the formation of critically volatile compounds in the following reactions. The subsequent cyclization of the iodide salt **2** with triethyl orthoacetate proceeded under mild conditions, and as expected, the corresponding 3-(dodecylthio)-Tz **3** was easily isolated as a nonvolatile oil. Strikingly, and in contrast to the strong oxidation conditions (i.e., $NaNO_2$/TFA under reflux) usually required for the Tz core formation,[24] dodecylthio tetrazine **3** was obtained uneventfully without the addition of any oxidizing reagent. Most probably, a dehydrogenation step to generate the oxidized tetrazine takes place. Compound **3** was then treated without further purification with a 1 M solution of hydrazine in THF at room temperature yielding tetrazine **4** as a red solid in moderate yield. The reaction likely proceeds via a S_N(ANRORC) mechanism at position 6 instead of a direct substitution at position 3 of the ring,[25] thereby allowing the successful substitution despite the large dodecylthio chain. Finally, the hydrazine derivative **4** was brominated to achieve **1** in good yield as a pink solid (Scheme 2a). Despite the relatively low overall yield of the process (12%), the inexpensive reagents involved combined with the absence of metal catalysts or strong oxidants makes the methodology practical and scalable.

The reactivity of **1** was then tested against a range of aliphatic and aromatic heteroatoms acting as nucleophiles. All substitutions proceeded rapidly under mild conditions to form the corresponding substituted amines, ethers, and sulfides with moderate to good yields (Scheme 2b), demonstrating the highly electrophilic character of **1**.

Considering these results, we decided to exploit the excellent reactivity of **1** with nucleophilic functional groups under mild

https://dx.doi.org/10.1021/acs.bioconjchem.0c00052
Bioconjugate Chem. 2020, 31, 933−938

conditions as a potential labeling strategy for proteins. To determine the potential reactive residues, a mixture of 7 N_{α}-Fmoc-protected amino acids with nucleophilic side chains were mixed with **1** in PBS pH = 7.2 (20% acetonitrile). No significant labeling was detected when only 1 equiv of **1** was added, whereas at 10 equiv a preferential labeling for lysine was observed, which extended to tyrosine when the number of equivalents was increased to 50 (SI Section 2.1).

To further validate the chemoselectivity toward lysine, we studied the labeling pattern of **1** in a model protein. Ribonuclease A (RNase A) was chosen due to its relatively small size (13.6 kDa), high stability, and limited number of lysine residues on its surface (10 in total), and because it had been previously used in bioconjugation experiments, where the most reactive lysines were determined.[26] The solubility of **1** in different aqueous buffers is an advantage in comparison with other labeling reagents, where the use of cosolvents like DMF or DMSO is usually needed. We attempted the labeling using 10 or 50 equiv of **1** at 3 different pH values (6.0, 7.4 or 8.0) (Figure 2a). After 30 min of reaction time, the different protein samples were analyzed intact by LC-MS (Figure 2b). As expected, less modified residues were observed at more acidic pH values, where a higher number of protonated amines impede their nucleophilic attack (Figure 2c).[27] Although the LC-MS/MS studies revealed that we were unable to target a single residue to achieve a site-selective labeling, the tryptic digestion of the full-length protein showed that lysines were primarily targeted (>95% of the modified peptides were Lys-labeled). This experiment shows the chemoselectivity of **1** versus lysines in a protein, in good agreement with the selectivity observed with a mixture of Fmoc-amino acids (Figure 2d). Additionally, circular dichroism revealed that the protein folding seemed to be completely unaffected by the labeling with 10 equiv, and minimally perturbed when using 50 (Figure 2e).

We then wanted to test the selectivity in a protein having available cysteine thiols. To this end, we reacted bovine serum albumin with **1** under the standard conditions. Somewhat surprisingly, we did not fid any significant substitution on the free nucleophilic cysteine 58, whereas several lysine residues readily incorporated the tetrazine fragment (SI Section 5.2.2). Therefore, **1** is able to discriminate between lysines and cysteines, with the reaction being highly chemoselective.

In light of the chemical features that the tetrazine fragment require to undergo CtR reactions with strained dienophiles, we envisaged that the resulting Tz-labeled lysines could be much more suitable as biorthogonal activators. We chose the lysine derivative **11** as a model of the protein labeled lysines (Figure 3a) and TCO−Doxorubicin carbamate[3,15] (TCO−Dox) as the clickable counterpart to study the parameters of the CtR reaction. The *in vitro* release of doxorubicin under physiological conditions (human serum at 37 °C) was determined by UPLC. We successfully found the liberation of approximately 50% of the theoretical amount of doxorubicin (Figure 3c), in line with several reported values to date in the literature.[3,9,11,28]

On the other hand, in order to study the kinetics of the cycloaddition step we used the strained dienophile (1*R*,8*S*,9*s*)-bicyclo[6.1.0]non-4-yn-9-ylmethanol (BCN−OH). The reaction between **11** and BCN−OH was monitored by UV−vis spectroscopy. We found a second-order rate constant (k_2) for the cycloaddition step with a k_2 of 1.073 M^{-1} s^{-1} (Figure 3b), slower than most EWG-substituted Tz but still in the range of biorthogonal reactions.[2] Interestingly, when using the methylated analog of our reagent, namely, 3-bromo-6-methyl-1,2,4,5-tetrazine[29,30] (**1Me**), the corresponding lysine derivative (compound **11Me**) released doxorubicin much more slowly (Figure 3c), probably due to the smaller k_2 of the cycloaddition step (0.013 M^{-1} s^{-1}, almost 2 orders of magnitude slower than **11**) (SI Section 4.1).[14] This different reactivity highlights the advantageous applicability of **1** compared to its methylated analog **1Me**.

To showcase the potential of reagent **1** for chemoselective protein labeling and subsequent CtR drug liberation, we selected the monoclonal antibody Trastuzumab, a front line treatment in HER2$^+$ breast cancer.[31] First, Trastuzumab was conveniently labeled using 50 equiv of **1** at pH 7.2 to afford Trastuzumab-**1**, which was characterized by LC-MS (Figure 4a; SI Section 5.1.2). Then, BT474 Her2$^+$ cells were incubated with the labeled antibody at 250 nM, followed by the addition of TCO−Dox (Figure 4b). A significant decrease compared to the cells incubated with control (Trastuzumab) was obtained

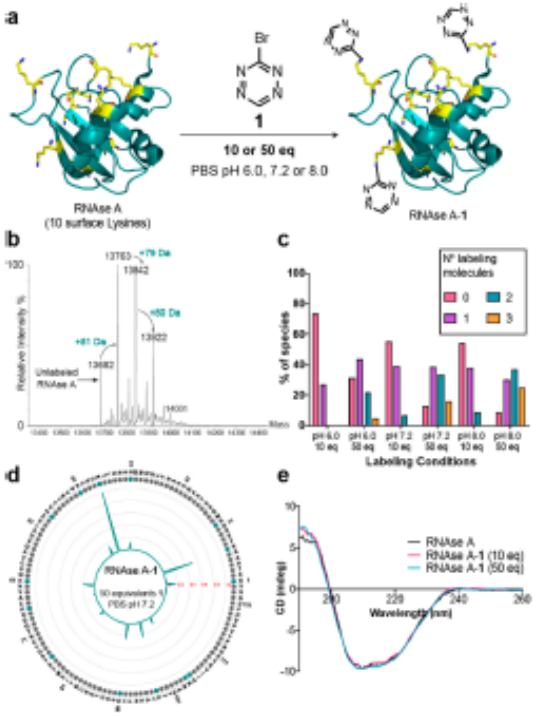


Figure 2. Chemoselective lysine labeling of RNase A. (a) Reaction scheme of RNase A (PDB file: 1FS3) with different equivalents of **1** (10 or 50 equiv) at different pH values (6.0, 7.2, or 8.0). (b) Intact protein analysis by mass spectrometry of a labeled sample of RNase A with 50 equiv of **1** at pH 7.2. The mass differences between each major peak correlate with the addition of a single molecule of **1**, presenting the typical pattern of proteins labeled with NHS esters. (c) Quantification of the relative abundance of the different species of labeled RNase A. Ratios were obtained using MassLynx Software, ver. 4.1.SCN704 and MaxEnt1 algorithm (Waters Inc.). (d) Relative ratios of the modified (labeled) vs unmodified residues of RNase A-1 (50 equiv, pH 7.2) obtained with the digested protein. Blue dots represent the residues identified as labeled in at least 1 digested peptide (SI Section 5.2). (e) Circular dichroism spectra of RNase A, RNase A-1 (10 equiv), and RNase A-1 (50 equiv) in PBS pH 7.2 (where NaCl was replaced by NaF).

https://doi.org/10.1021/acs.bioconjchem.0c00052
Bioconjugate Chem. 2020, 31, 933−938

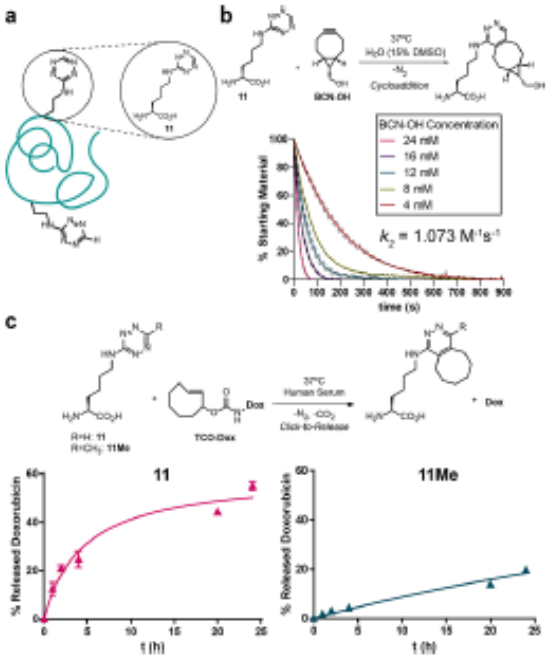


Figure 3. *In vitro* characterization of the CtR between Tz-labeled lysines and TCO−Dox. (a) Compound **11** has been used as a model to characterize both the cycloaddition step and the overall CtR reaction kinetics. (b) Second-order rate constant (k_2) determination of the cycloaddition step between **11** and (1*R*,8*S*,9*s*)-bicyclo[6.1.0]-non-4-yn-9-ylmethanol (BCN−OH), which was used as a model strained dienophile, by UV−vis spectroscopy. Five different k_{obs} were obtained using a controlled excess of the strained dienophile under pseudo-first-order conditions, which were subsequently used to extrapolate the k_2 of the reaction. (c) Kinetics of the CtR reaction between **11** (left) or **11Me** (right) and TCO−Dox in physiological conditions (human serum at 37 °C). Free doxorubicin formation was followed by UPLC (after protein precipitation) at different time points, corrected by a stability factor and quantified by the AUC using a calibration curve.

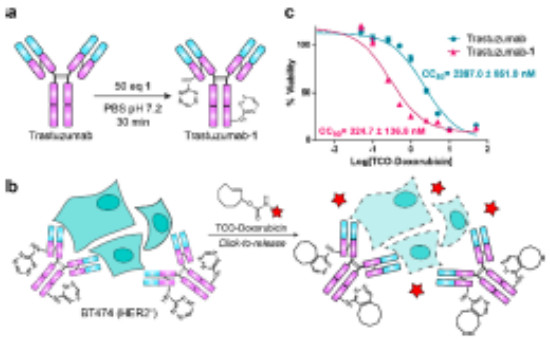


Figure 4. Bioorthogonal liberation of doxorubicin and subsequent cytotoxicity. (a) Labeling of Trastuzumab using 50 equiv of **1** in PBS pH 7.2. (b) Release of doxorubicin (red star) by the reaction of Trastuzumab-**1** and TCO−Dox in BT474 (HER2$^+$) cell culture. (c) Comparison of the cytotoxic concentration 50 (CC_{50}) of TCO−Dox with the labeled/unlabeled Trastuzumab. Cell viability was determined by WST-1 reagent (Roche).

in the CC_{50}, thus demonstrating the release of doxorubicin and its cytotoxic effect after 72 h of incubation (Figure 4c). Most importantly, at 250 nM neither the wild-type antibody nor the antibody with a covalently linked doxorubicin (Trastuzumab-Dox, SI Section 2.4) showed any cytotoxic effect (SI Section 7.1). Moreover, when a nonreleasable version of TCO−Dox (using BCN−Dox instead) was employed under the same experimental conditions, no significant differences between incubating with Trastuzumab or Trastuzumab-**1** were observed (SI Section 7.1.3). As expected, the cytotoxic effect obtained for the released doxorubicin was smaller than the observed CC_{50} when incubating with free doxorubicin (CC_{50} = 29.7 ± 10.4 nM, SI Section 7.1.1). This is in good agreement with the observed ca. 50% *in vitro* doxorubicin release. However, the near 10-fold increase in toxicity obtained when incubating **TCO−Dox** with Trastuzumab-**1** compared to Trastuzumab clearly demonstrates the potential of this strategy to attain specific liberation of active drugs at the site of action.

Since the site-selectivity of **1** in the late-stage incorporation of the tetrazine fragment into proteins was poor, we decided to investigate an alternative strategy to achieve chemoselective and site-selective protein incorporation of a tetrazine through genetic code expansion. To maintain the *in vivo* release values observed in the case of labeled lysines, we focused on preserving the monosubstituted amino-Tz moiety. For this reason, we synthesized compound **12**, a tyrosine analogue suitable for genetic incorporation using mutant *Methanocaldococcus jannaschii* Tyrosyl-tRNA synthetases/tRNA^{CUA} pairs (*Mj*TyrRS*/tRNA^{CUA}) in *E. coli*. The preparation of amino acid **12** was achieved by the reaction of N_α-Boc-4-aminophenylalanine with **1**, followed by Boc deprotection using HCl 4 M in dioxane (Figure 5a). First of all, we demonstrated that the release values from the reaction of **12** with TCO−Dox were comparable to the ones obtained with **11** (Figure 5b). Next, we tested 3 different previously reported *Mj*TyrRS*/tRNA^{CUA} pairs bearing a different set of mutations near the amino acid binding domain: Tet-v1.0[32] and Tet-v2.0[33] are both used to incorporate non-natural amino acids with methyl-Tz groups, whereas pCNF is known as one of the most promiscuous MjTyrRS in terms of substrate specificity.[34] GFP carrying an amber stop (TAG) codon at position 150 (GFP150TAG) was used as the model protein to test the incorporation of **12**. Unexpectedly, all 3 pairs seemed to express the desired protein (GFP-**12**), albeit with different efficiencies (SI Section 8.1). For that reason, we proceeded with Tet-v2.0 as it was the MjTyrRS*/tRNA^{CUA} pair showing the highest expression of the recombinant protein (36.6 mg protein/L culture, 8% expression compared to the obtained GFP wild-type under the same conditions), as confirmed by intact protein mass spectrometry (Figure 5c,d). Conjugation of GFP-**12** with TCO−Cy5 demonstrated that the Tz ring was reactive after the protein expression and purification protocol (Figure 5e), and intact protein analysis of GFP-**12** in reaction with BCN−OH showed that all recombinant protein proceeded rapidly to the expected product (GFP-12+BCN) (SI Section 5.1.4), demonstrating minimal or nonexistent undesired amino acid misincorporation.

■ CONCLUSIONS

To summarize, we have reported a metal- and oxidant-free synthetic route to obtain 3-bromo-1,2,4,5-tetrazine, a monosubstituted, brominated Tz. Its excellent reactivity with different nucleophiles prompted us to investigate its use for

https://dx.doi.org/10.1021/acs.bioconjchem.0c00052
Bioconjugate Chem. 2020, 31, 933−938

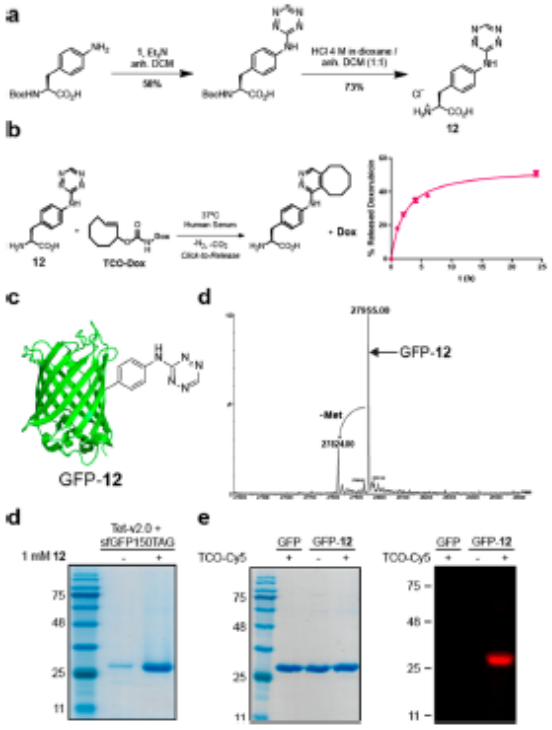


Figure 5. Genetic incorporation of the monosubtituted amino-Tz **11**. (a) Synthesis of the non-natural amino acid **12**. (b) Kinetics of the click-to-release reaction between **12** and TCO−Dox in physiological conditions (human serum at 37 °C) by UPLC. (c) Intact protein analysis of the model protein GFP (PDB file: 2B3P) with **12** site-selectively incorporated at position 150. (d) Coomassie stained SDS-PAGE gel of the expression of GFP150TAG in *E. coli* cotransformed with Tet-v2.0[33] (a mutant version of MjTyrRS). The faint band on the expression where no amino acid **12** was added indicates background expression due to read-through of the stop codon. (e) Coomassie stained and fluorescence imaging of an SDS-PAGE gel with GFP-**12** reacted with TCO−Cy5.

biomolecule labeling, particularly as a simple and straightforward method to incorporate biorthogonal handles into proteins. LC-MS/MS studies revealed that the compound is chemoselective for lysines, thereby forming stable monosubtituted amino-Tz at the protein surface while maintaining the folding. Due to the chemical features of the formed labeled lysines, we have investigated their use for CtR reactions and successfully applied it in a therapeutically relevant context in cell culture. Finally, to finely control the labeling sites of a target protein of interest, we have been able to genetically incorporate compound **12** site-selectively in a model protein and shown its ability to undergo CtR reactions under physiological conditions.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.bioconjchem.0c00052.

General synthetic methods, characterization of compounds and NMR spectra, protein labeling methodology, MS techniques and data processing, reaction kinetics, cellular assays, and protein expression and purification protocol (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Antoni Riera − *Institute for Research in Biomedicine (IRB Barcelona), Barcelona Institute of Science and Technology, 08028 Barcelona, Spain; Departament de Química Inorgànica i Orgànica, Secció Orgànica, Universitat de Barcelona, Barcelona 08028, Spain;* orcid.org/0000-0001-7142-7675; Email: antoni.riera@irbbarcelona.org

Lluís Ribas de Pouplana − *Institute for Research in Biomedicine (IRB Barcelona), Barcelona Institute of Science and Technology, 08028 Barcelona, Spain; Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08010, Spain;* Email: lluis.ribas@irbbarcelona.org

Authors

Enric Ros − *Institute for Research in Biomedicine (IRB Barcelona), Barcelona Institute of Science and Technology, 08028 Barcelona, Spain;* orcid.org/0000-0001-6269-3058

Marina Bellido − *Institute for Research in Biomedicine (IRB Barcelona), Barcelona Institute of Science and Technology, 08028 Barcelona, Spain;* orcid.org/0000-0002-8035-8648

Xavier Verdaguer − *Institute for Research in Biomedicine (IRB Barcelona), Barcelona Institute of Science and Technology, 08028 Barcelona, Spain; Departament de Química Inorgànica i Orgànica, Secció Orgànica, Universitat de Barcelona, Barcelona 08028, Spain;* orcid.org/0000-0002-9229-969X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.bioconjchem.0c00052

Author Contributions

All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank institutional funding from the Spanish Ministry of Economy, Industry and Competitiveness (MINECO, CTQ2017-87840-P to A.R. and BIO2015-64572-R to L.R.dP) through the Centres of Excellence Severo Ochoa Award, and from the CERCA Programme of the Catalan Government. E.R. thanks the support of funding from La Caixa Foundation (ID 100010434) and La Caixa Ph.D. Fellowship. We thank Mass Spectrometry and Proteomics Core Facility of IRB Barcelona, which is a member of Proteored, PRB3-ISCIII. We also thank Dr Macarena Sánchez for her technical support and fruitful discussions and Ryan Mehl's group, from Oregon State University, for providing the plasmids necessary for the genetic incorporation of unnatural amino acids.

■ REFERENCES

(1) Murrey, H. E., Judkins, J. C., Am Ende, C. W., Ballard, T. E., Fang, Y., Riccardi, K., Di, L., Guilmette, E. R., Schwartz, J. W., Fox, J. M., et al. (2015) Systematic evaluation of bioorthogonal reactions in live cells with clickable haloTag ligands: Implications for intracellular imaging. *J. Am. Chem. Soc. 137*, 11461−11475.

(2) Oliveira, B. L., Guo, Z., and Bernardes, G. J. L. (2017) Inverse electron demand Diels−Alder reactions in chemical biology. *Chem. Soc. Rev. 46*, 4895−4950.

https://dx.doi.org/10.1021/acs.bioconjchem.0c00052
Bioconjugate Chem. 2020, 31, 933−938

(3) Versteegen, R. M., Rossin, R., Ten Hoeve, W., Janssen, H. M., and Robillard, M. S. (2013) Click to release: instantaneous doxorubicin elimination upon tetrazine ligation. *Angew. Chem., Int. Ed. 52*, 14112−14116.
(4) Reiner, T., and Zeglis, B. M. (2014) The inverse electron demand Diels-Alder click reaction in radiochemistry. *J. Labelled Compd. Radiopharm. 57*, 285−290.
(5) Rossin, R., Van Duijnhoven, S. M. J., Ten Hoeve, W., Janssen, H. M., Kleijn, L. H. J., Hoeben, F. J. M., Versteegen, R. M., and Robillard, M. S. (2016) Triggered drug release from an antibody-drug conjugate using fast "Click-to-Release" chemistry in mice. *Bioconjugate Chem. 27*, 1697−1706.
(6) Mejia Oneto, J. M., Khan, I., Seebald, L., and Royzen, M. (2016) In vivo bioorthogonal chemistry enables local hydrogel and systemic pro-drug to treat soft tissue sarcoma. *ACS Cent. Sci. 2*, 476−482.
(7) Oliveira, B. L., Guo, Z., and Bernardes, G. J. L. (2017) Inverse electron demand Diels−Alder reactions in chemical biology. *Chem. Soc. Rev. 46*, 4895−4950.
(8) Jiménez-Moreno, E., Guo, Z., Oliveira, B. L., Albuquerque, I. S., Kitowski, A., Guerreiro, A., Boutureira, O., Rodrigues, T., Jiménez-Osés, G., and Bernardes, G. J. L. (2017) Vinyl ether/tetrazine pair for the traceless release of alcohols in cells. *Angew. Chem., Int. Ed. 56*, 243−247.
(9) Xu, M., Tu, J., and Franzini, R. M. (2017) Rapid and efficient tetrazine-induced drug release from highly stable benzonorbornadiene derivatives. *Chem. Commun. 53*, 6271−6274.
(10) Carlson, J. C. T., Mikula, H., and Weissleder, R. (2018) Unraveling tetrazine-triggered bioorthogonal elimination enables chemical tools for ultrafast release and universal cleavage. *J. Am. Chem. Soc. 140*, 3603−3612.
(11) Fan, X., Ge, Y., Lin, F., Yang, Y., Zhang, G., Shu, W., Ngai, C., Lin, Z., Zheng, S., Wang, J., et al. (2016) Optimized tetrazine derivatives for rapid bioorthogonal decaging in living cells. *Angew. Chem., Int. Ed. 55*, 14046−14050.
(12) Mayer, S., and Lang, K. (2017) Tetrazines in inverse-electron-demand Diels-Alder cycloadditions and their use in biology. *Synthesis 49*, 830−848.
(13) Yang, J., Liang, Y., Šečkute, J., Houk, K. N., and Devaraj, N. K. (2014) Synthesis and reactivity comparisons of 1-methyl-3-substituted cyclopropene mini-Tags for tetrazine bioorthogonal reactions. *Chem. - Eur. J. 20*, 3365−3375.
(14) Rossin, R., Versteegen, R. M., Wu, J., Khasanov, A., Wessels, H. J., Steenbergen, E. J., Ten Hoeve, W., Janssen, H. M., Van Onzen, A. H. A. M., Hudson, P. J., et al. (2018) Chemically triggered drug release from an antibody-drug conjugate leads to potent antitumour activity in mice. *Nat. Commun. 9*, 1−11.
(15) Rossin, R., Van Duijnhoven, S. M. J., Ten Hoeve, W., Janssen, H. M., Kleijn, L. H. J., Hoeben, F. J. M., Versteegen, R. M., and Robillard, M. S. (2016) Triggered drug release from an antibody-drug conjugate using fast "Click-to-Release" chemistry in mice. *Bioconjugate Chem. 27*, 1697−1706.
(16) Rossin, R., Versteegen, R. M., Wu, J., Khasanov, A., Wessels, H. J., Steenbergen, E. J., Ten Hoeve, W., Janssen, H. M., Van Onzen, A. H. A. M., Hudson, P. J., et al. (2018) Chemically triggered drug release from an antibody-drug conjugate leads to potent antitumour activity in mice. *Nat. Commun. 9*, 1−11.
(17) Boutureira, O., and Bernardes, G. J. L. (2015) Advances in chemical protein modification. *Chem. Rev. 115*, 2174−2195.
(18) O'Donoghue, P., Ling, J., Wang, Y.-S., and Söll, D. (2013) Upgrading protein synthesis for synthetic biology. *Nat. Chem. Biol. 9*, 594−598.
(19) Dumas, A., Lercher, L., Spicer, C. D., and Davis, B. G. (2015) Designing logical codon reassignment - expanding the chemistry in biology. *Chem. Sci. 6*, 50−69.
(20) Rahim, M. K., Kota, R., and Haun, J. B. (2015) Enhancing reactivity for bioorthogonal pretargeting by unmasking antibody-conjugated *trans*-Cyclooctenes. *Bioconjugate Chem. 26*, 352−360.
(21) Rossin, R., Van Den Bosch, S. M., Ten Hoeve, W., Carvelli, M., Versteegen, R. M., Lub, J., and Robillard, M. S. (2013) Highly reactive *trans*-cyclooctene tags with improved stability for Diels-Alder chemistry in living systems. *Bioconjugate Chem. 24*, 1210−1217.
(22) Maggi, A., Ruivo, E., Fissers, J., Vangestel, C., Chatterjee, S., Joossens, J., Sobott, F., Staelens, S., Stroobants, S., Van Der Veken, P., et al. (2016) Development of a novel antibody−tetrazine conjugate for bioorthogonal pretargeting. *Org. Biomol. Chem. 14*, 7544−7551.
(23) Zhou, Q., Audebert, P., Clavier, G., Miomandre, F., and Tang, J. (2014) New unsymmetrical alkyl-s-tetrazines: original syntheses, fluorescence and electrochemical behaviour. *RSC Adv. 4*, 7193−7195.
(24) Fields, S. C., Parker, M. H., and Erickson, W. R. (1994) A simple route to unsymmetrically substituted 1,2,4,5-tetrazines. *J. Org. Chem. 59*, 8284−8287.
(25) Counotte-Potman, A., van der Plas, H., van Veldhuizen, B., and Landheer, C. (1981) Occurrence of the SN(ANRORC) mechanism in the hydrazination of 1,2,4,5-tetrazines. *J. Org. Chem. 46*, 5102−5109.
(26) Chen, X., Muthoosamy, K., Pfisterer, A., Neumann, B., and Weil, T. (2012) Site-selective lysine modification of native proteins and peptides via kinetically controlled labeling. *Bioconjugate Chem. 23*, 500−508.
(27) Matos, M. J., Oliveira, B. L., Martínez-Sáez, N., Guerreiro, A., Cal, P. M. S. D., Bertoldo, J., Maneiro, M., Perkins, E., Howard, J., Deery, M. J., et al. (2018) Chemo- and regioselective lysine modification on native proteins. *J. Am. Chem. Soc. 140*, 4004−4017.
(28) Davies, S., Oliveira, B. L., and Bernardes, G. J. L. (2019) Development of a self-immolative linker for tetrazine-triggered release of alcohols in cells. *Org. Biomol. Chem. 17*, 5725−5730.
(29) Counotte-Potman, A., and van der Plas, H. (1978) Degenerate ring transformations in reactions of 1,2,4,5-tetrazines with hydrazine. *J. Heterocycl. Chem. 15*, 445−448.
(30) Wieczorek, A., Werther, P., Euchner, J., and Wombacher, R. (2017) Green- to far-red-emitting fluorogenic tetrazine probes − synthetic access and no-wash protein imaging inside living cells. *Chem. Sci. 8*, 1506−1510.
(31) Hudis, C. A. (2007) Trastuzumab — mechanism of action and use in clinical practice. *N. Engl. J. Med. 357*, 39−51.
(32) Seitchik, J. L., Peeler, J. C., Taylor, M. T., Blackman, M. L., Rhoads, T. W., Cooley, R. B., Refakis, C., Fox, J. M., and Mehl, R. A. (2012) Genetically encoded tetrazine amino acid directs rapid site-specific in vivo bioorthogonal ligation with *trans*-cyclooctenes. *J. Am. Chem. Soc. 134*, 2898−2901.
(33) Blizzard, R. J., Backus, D. R., Brown, W., Bazewicz, C. G., Li, Y., and Mehl, R. A. (2015) Ideal bioorthogonal reactions using a site-specifically encoded tetrazine amino acid. *J. Am. Chem. Soc. 137*, 10044−10047.
(34) Young, D. D., Young, T. S., Jahnz, M., Ahmad, I., Spraggon, G., and Schultz, P. G. (2011) An evolved aminoacyl-TRNA synthetase with atypical polysubstrate specificity. *Biochemistry 50*, 1894−1900.

https://dx.doi.org/10.1021/acs.bioconjchem.0c00052
Bioconjugate Chem. 2020, 31, 933−938

ChemComm

ROYAL SOCIETY OF CHEMISTRY

COMMUNICATION

View Article Online
View Journal

Cite this: DOI: 10.1039/d0cc03482g

Received 15th May 2020,
Accepted 3rd August 2020

DOI: 10.1039/d0cc03482g

rsc.li/chemcomm

Synthesis of 3-alkyl-6-methyl-1,2,4,5-tetrazines *via* a Sonogashira-type cross-coupling reaction†

Enric Ros,[a] Amparo Prades,[a] Dominique Forson,[a] Jacqueline Smyth,[a] Xavier Verdaguer,[ab] Lluis Ribas de Pouplana *[ac] and Antoni Riera *[ab]

1,2,4,5-Tetrazines have become extremely useful tools in chemical biology. However, the synthesis of some challenging substrates such as asymmetrically disubstituted alkyltetrazines is still a limitation for the widespread use of this class of compounds. Herein we describe an efficient route to these compounds based on the Sonogashira coupling of 3-bromo-6-methyl-1,2,4,5-tetrazine and 3-bromo-6-phenyl-1,2,4,5-tetrazine with terminal alkynes. The preparation of the starting reagents has also been optimized. The alkynyl products have been used as intermediates for the synthesis of dialkyl-tetrazines through a sequence of hydrogenation and re-oxidation with unprecedented yields. The synthetic applicability of this new approach is showcased through the preparation of several unnatural amino acids bearing alkynyl- and alkyl-1,2,4,5-tetrazine fragments.

Throughout the 20th century, the interest in 1,2,4,5-tetrazines remained almost completely limited to natural product synthesis,[1] coordination chemistry[2] and explosives applications.[3] However, with the advent of click chemistry in the beginning of the 2000s[4] and their first use as bioorthogonal reagents with strained alkenes through the inverse electron-demand Diels–Alder (iEDDA) cycloaddition in 2008,[5,6] the interest in tetrazines increased exponentially. Their use as advantageous research tools in live cell imaging,[7] bioconjugation experiments[8] or modern drug discovery techniques[9] has exemplified their huge potential. Furthermore, therapeutic applications for nuclear medicine,[10] *in vivo* PROTAC assembly[11] or prodrug activation[12] have expanded their use in the pharmaceutical industry. The iEDDA cycloaddition endows some characteristics that allow its widespread application as a unique chemical biology tool: high selectivity, superfast kinetics in water and absence of metal catalysis.[13]

Despite great efforts directed towards the development of robust synthetic methods for the generation of tetrazines, their preparation is still challenging. For instance, tetrazines are typically synthesized through the condensation of two nitriles in the presence of hydrazine, followed by an oxidation step of the obtained dihydrotetrazine.[13] While this is a useful strategy to attain several products, it is less efficient when generating non-symmetrical alkyltetrazines or compounds with labile functional groups due to the strong reaction conditions required.[14]

To overcome these limitations, new ways to functionalize useful substrates with tetrazines have arisen in the last two decades, employing modern metal-catalyzed cross-coupling reactions. A potential approach is to use aryltetrazines, in which the C–C bond formation reaction takes place onto the aryl substituent.[15–17] Alternatively, direct coupling at position 3 or 6 of the tetrazine ring is preferable because it leads to the introduction of minimal perturbation in the attached substrate (Fig. 1).

The first example on direct tetrazine cross-coupling employed the Sonogashira reaction between terminal alkynes and electronically enriched chloro-tetrazines.[18] Suzuki and Stille-type couplings with chloro-[19,20] and methylthio-tetrazines[21] have also been reported to generate similar 3-alkylamino-6-aryltetrazines. Unfortunately, the amino substituent in the reported tetrazines limits their bioorthogonal use due to slow reaction kinetics.[13]

An improvement in the bioorthogonal applicability for the products of these reactions was reported when a set of fluorescein and Oregon green derivatives were synthesized *via* a Stille coupling, employing 3-bromo-6-methyl-1,2,4,5-tetrazine (**1a**) as reagent.[22] However, the resulting aryltetrazines were obtained in low yields, and no examples with alkyl substituents in the tetrazine ring were described. Also, the relatively inefficient synthesis reported for **1a** limits the broader use of this synthetic approach. Recently, the 6-methyltetrazin-3-yl fragment has also been introduced *via* a Ag-mediated Liebeskind–Srogl reaction to arylboronic acids.[23]

[a] *Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain. E-mail: antoni.riera@irbbarcelona.org, lluis.ribas@irbbarcelona.org*

[b] *Departament de Química Inorgànica i Orgànica, Secció Orgànica, Universitat de Barcelona, Martí i Franquès 1, Barcelona E-08028, Spain*

[c] *Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluis Companys, 23, Barcelona 08010, Spain*

† Electronic supplementary information (ESI) available: General synthetic methods, characterisation of products and NMR spectra. See DOI: 10.1039/d0cc03482g

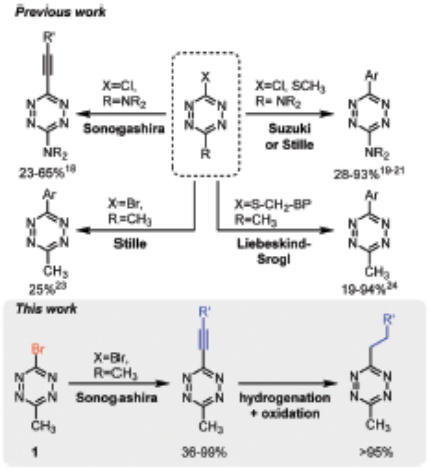
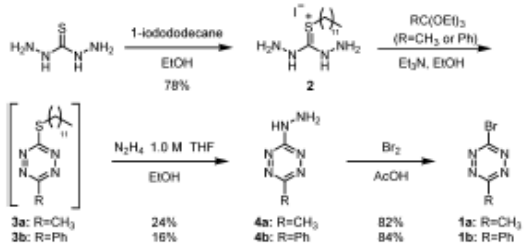


Fig. 1 Overview of the reported metal-catalysed cross-couplings with 1,2,4,5-tetrazines.

The revision of existing literature on direct cross-coupling reactions in the tetrazine ring indicates that chlorine is more inert than bromine as the participating halogen,[24] thus requiring electronically enriched tetrazines and strong reaction conditions; in contrast, bromine can react under milder conditions, enabling the access to a broader spectrum of useful substrates.[22] Furthermore, given that substituents in position 3- and 6- of the tetrazine ring determine its reactivity and stability, we envisaged that 3-alkyl-6-methyl-1,2,4,5-tetrazines could strike the right balance between both parameters. Consequently, we directed our efforts towards the synthesis of **1a**.

Herein, we describe an optimized synthesis of **1a**, achieving a 15% overall yield, the highest reported to date for this reagent.[22,25] We demonstrate the modular character of the synthetic route with the preparation of 3-bromo-6-phenyl-1,2,4,5-tetrazine (**1b**). We describe the Sonogashira couplings of **1a–b** with terminal alkynes, and the use of the products as intermediates for the synthesis of 3-alkyl-6-methyl-1,2,4,5-tetrazines, reaching unprecedented yields for this type of compounds. The methodology is applied to the synthesis of unnatural amino acids bearing the 6-methyltetrazin-3-yl fragment.

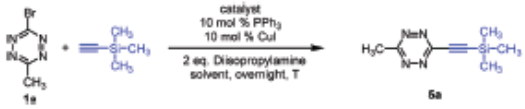
Although **1a** was first described in 1978,[25] its use as a reagent to introduce the tetrazine fragment was hampered by the low yields of the previously reported syntheses.[22,25] We followed a synthetic route analogous to the one recently published by us to generate the unsubstituted 3-bromo-1,2-4,5-tetrazine.[26] Thiocarbohydrazide was alkylated with 1-iodododecane to afford the iodide salt **2**, which was then reacted with triethyl orthoacetate in the presence of triethylamine to generate 3-(dodecylthio)-6-methyl-1,2,4,5-tetrazine (**3a**) (Scheme 1). Notably, **3a** was formed at room temperature and without the need for any oxidizing agent. Without further purification, intermediate **3a** was reacted with a solution of hydrazine in THF to achieve the hydrazinyl-methyltetrazine **4a**, which was finally brominated to achieve the final product **1a** as a red, crystalline solid. This synthesis, in addition to the larger overall yield (15%) compared to the highest reported to date (2%),[22] has the advantages of being scalable, with non-volatile intermediates, and not requiring any metal catalysis or oxidants. An analogous route employing triethyl orthobenzoate was followed to generate **1b** with comparable yields, demonstrating the versatility of the synthesis to generate bromotetrazines.



Scheme 1 Synthesis of bromotetrazines **1a** and **1b**.

With an efficient synthesis of **1a** in hand, we then focused on screening different Pd-catalyzed reactions. Initial attempts evidenced that **1a** underwent undesired reactions when using inorganic bases or high reaction temperatures. Hence, we decided to test a Sonogashira-type reaction under milder conditions. We initially tested the coupling between **1a** and trimethylsilylacetylene to generate the coupled product 3-methyl-6-((trimethylsilyl)ethynyl)-1,2,4,5-tetrazine (**5a**). Using 5 mol% of Pd^{II} precatalysts (either $PdCl_2$ or $Pd(OAc)_2$) (Table 1, entries 1 and 2) in toluene, the desired product **5a** was formed in moderate yields. The use of $Pd(PPh_3)_4$ as catalyst resulted in lower yields (Table 1, entry 3). However, an excellent 75% yield was obtained using $PdCl_2(PPh_3)_2$ (Table 1, entry 4). Using more polar solvents, like dichloromethane or tetrahydrofurane (Table 1, entries 5 and 6) or increasing the temperature (Table 1, entries 7 and 8), worsened the yields. We concluded that, under mild conditions, **1a** is an excellent reagent to undergo Sonogashira couplings with terminal alkynes.

Table 1 Screening of reaction conditions for Sonogashira coupling of **1a**

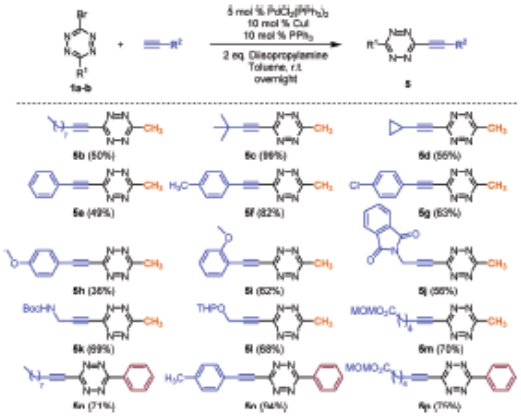


Entry	Catalyst	*T*	Solvent	Conv.[a] [%]	Yield[b] [%]
1	$PdCl_2$	r.t.	Toluene	>99	57
2	$Pd(OAc)_2$	r.t.	Toluene	>99	54
3	$Pd(PPh_3)_4$	r.t.	Toluene	>99	23
4	$PdCl_2(PPh_3)_2$	r.t.	Toluene	>99	75
5	$PdCl_2(PPh_3)_2$	r.t.	DCM	>99	56
6	$PdCl_2(PPh_3)_2$	r.t.	THF	>99	5
7	$PdCl_2(PPh_3)_2$	50	Toluene	>99	66
8	$PdCl_2(PPh_3)_2$	100	Toluene	>99	53

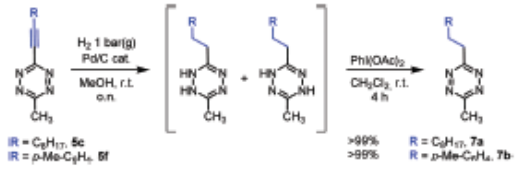
The reaction was performed in a degassed, sealed vial with 0.057 mmol **1a** [0.15 M], using 5 mol% of catalyst loading and 2 eq. of TMS-acetylene as the model alkyne. [a] Determined by ^{1}H NMR spectroscopy. [b] Yield determined by ^{1}H NMR spectroscopy using mesitylene as an internal standard.

With the optimized reaction conditions in hand, we explored the substrate scope using a wide range of terminal alkynes (Scheme 2). Aliphatic alkynes afforded acetylenic tetrazines **5b–d** in good to excellent yields. Likewise, aromatic alkynes with different substitution patterns (phenyl, tolyl, *p*-chlorophenyl, *o*- and *p*-methoxyphenyl) were tested yielding the desired products (**5e–i**) in moderate to good yields. As it could be anticipated, the coupling with nucleophilic substrates (alcohols or amines) did not afford desired products, giving mainly the nucleophilic aromatic substitution products. Likewise, free carboxylic acids did not yield the desired products, leading to decomposition of the tetrazine ring. Therefore, we needed to use protecting groups. Acetylenic amines were protected either as phthalimides (**5j**) or *tert*-butyl carbamates (**5k**); hydroxyls were protected as tetrahydropyranyl ethers (**5l**) and acetylenic esters were tested with good results (**5m**). Coupling between **1b** and different terminal alkynes also lead to the desired products **5n–p** in good yields. Deprotection of compounds **5j–m** was next studied. Basic conditions were incompatible with the tetrazine ring. Fortunately, under acidic conditions (diluted solution of HCl in dioxane) **5k–m** led to the desired free amine, alcohol and carboxylic acid (**6a–c**) (Section S2.4, ESI†).

Alkynyl-tetrazines **5a–p** and **6a–c** are an interesting new class of compounds that hold a potential use in bioorthogonal iEDDA reactions, for instance in the generation of rigid fluorogenic probes.[16] However, we envisaged they could also be useful intermediates to attain alkyltetrazines by hydrogenation of the triple bond. In particular, asymmetrically substituted 3-alkyl-6-methyl-1,2,4,5-tetrazines are a synthetically challenging class of compounds and, to date, there are only a limited number of reliable methods for their synthesis.[27–30] The general use of hydrazine, which acts both as a good nucleophile and reductant, entails some functional group incompatibilities.[31] Furthermore, the generation of asymmetrically substituted tetrazines might be a challenge due to the generation of a statistical mixture of products difficult to separate.



Scheme 2 Substrate scope of the Sonogashira reaction between **1a–b** and aliphatic or aromatic alkynes.

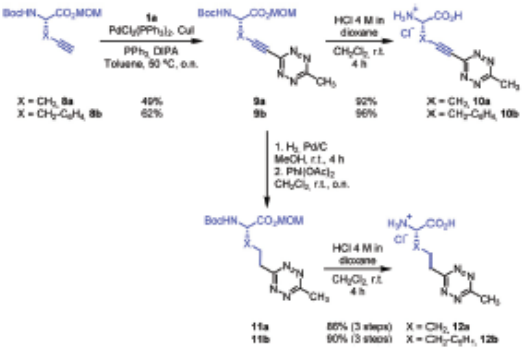


Scheme 3 Stepwise hydrogenation/re-oxidation in alkynyltetrazines **5c** and **5f** taken as model (aliphatic and aromatic) compounds.

Due to these limitations, we decided to test the hydrogenation of the triple bond on two of the Sonogashira coupling products (**5c** and **5f**, Scheme 3) as model compounds. When each of them was subjected to hydrogenation at 1 bar(g) of pressure with Pd/C as catalyst, the characteristic pink coloration of tetrazines disappeared immediately in both cases. ^{1}H NMR spectroscopy of the reaction crude showed a mixture of compounds present, consistent with the formation of the different dihydrotetrazine isomers. When we attempted the re-oxidation to the fully functional tetrazine on the reaction crude employing the typical $NaNO_2/H^+$ preparation -usually employed to oxidize dihydrotetrazines-[13] we observed the formation of the desired product with moderate yields, together with degradation products. Gratifyingly, when the re-oxidation was performed with an hypervalent iodine reagent [(diacetoxyiodo)benzene],[30] full conversion to the desired products **7a** and **7b** was achieved. The two-step sequence of hydrogenation/oxidation gave almost quantitative yields.

Finally, to showcase the versatility of the reported synthetic strategy on biologically relevant molecules, the synthesis of different unnatural amino acids was performed (Scheme 4). Tetrazine-containing amino acids are interesting compounds in chemical biology, and several have been genetically incorporated to proteins *in vivo*.[26,32–35]

Therefore, following our previous findings, *N*-Boc and MOM esters were selected as appropriate protecting groups. Thus, *N*-Boc protected MOM esters of L-propargylglycine (**8a**) and 4-ethynyl-L-phenylalanine (**8b**) were synthesised and reacted with **1a** under the optimized Sonogashira conditions to afford the methyltetrazinyl-amino acid derivatives **9a–b**. Interestingly,



Scheme 4 Synthesis of unnatural amino acids **10a–b** and **12a–b**.

heating the reaction mixture at 50 °C was necessary in both cases to afford completion. The direct deprotection of these intermediates under acidic conditions yielded the hydrochloride salts **10a–b**, containing an alkynyltetrazine functionality. Alternatively, when intermediates **10a–b** were subjected to the sequence of hydrogenation and re-oxidation, followed by deprotection of the Boc carbamate and MOM ester, the (dialkyl)tetrazine amino acids **12a–b** were obtained.

In summary, we report an optimised synthetic route to 3-bromotetrazines **1a** and **1b**, useful tetrazine precursors to undergo Sonogashira-type cross-couplings with a set of structurally different terminal alkynes. The obtained products can be hydrogenated to afford asymmetrically substituted (dialkyl)-tetrazines in excellent yields. The versatility of the different reactions, together with the functional group tolerance provided by easily removable protective groups, enabled the synthesis of a set of 6-methyltetrazin-3-yl unnatural amino acids.

We anticipate this methodology can be further exploited to afford numerous new 1,2,4,5-tetrazines derived from the large catalogue of available alkynes, thereby achieving a wide range of potentially useful compounds for chemical biology.

This work was supported by grants from FEDER/Ministerio de Ciencia, Innovación y Universidades (MICINN)-Agencia Estatal de Investigación (CTQ2017-87840-P to A. R.) and the Spanish Ministry of Economy, Industry and Competitiveness (MINECO, BIO2015-64572-R to L. R. dP). IRB Barcelona is the recipient of institutional funding from MICINN through the Centres of Excellence Severo Ochoa award and from the CERCA Program of the Catalan Government. E. R. thanks the support of funding from La Caixa PhD Fellowship.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 D. L. Boger and J. Hong, *J. Am. Chem. Soc.*, 2001, **123**, 8515–8519.
2 W. Kaim, *Coord. Chem. Rev.*, 2002, **230**, 127–139.
3 D. E. Chavez and M. A. Hiskey, *J. Heterocycl. Chem.*, 1998, **35**, 1329–1332.
4 H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2001, **40**, 2004–2021.
5 M. L. Blackman, M. Royzen and J. M. Fox, *J. Am. Chem. Soc.*, 2008, **130**, 13518–13519.
6 N. K. Devaraj, R. Weissleder and S. A. Hilderbrand, *Bioconjugate Chem.*, 2008, **19**, 2297–2299.
7 G. B. Cserép, A. Herner and P. Kele, *Methods Appl. Fluoresc.*, 2015, **3**, 042001.
8 J. Šečkuté, J. Yang and N. K. Devaraj, *Nucleic Acids Res.*, 2013, **41**, e148.
9 A. Rutkowska, D. W. Thomson, J. Vappiani, T. Werner, K. M. Mueller, L. Dittus, J. Krause, M. Muelbaier, G. Bergamini and M. Bantscheff, *ACS Chem. Biol.*, 2016, **11**, 2541–2550.
10 T. Reiner and B. M. Zeglis, *J. Labelled Compd. Radiopharm.*, 2014, 57, 285–290.
11 H. Lebraud, D. J. Wright, C. N. Johnson and T. D. Heightman, *ACS Cent. Sci.*, 2016, **2**, 927–934.
12 X. Ji, Z. Pan, B. Yu, L. K. De La Cruz, Y. Zheng, B. Ke and B. Wang, *Chem. Soc. Rev.*, 2019, **48**, 1077–1094.
13 B. L. Oliveira, Z. Guo and G. J. L. Bernardes, *Chem. Soc. Rev.*, 2017, **46**, 4895–4950.
14 G. Clavier and P. Audebert, *Chem. Rev.*, 2010, **110**, 3299–3314.
15 H. Xiong, Y. Gu, S. Zhang, F. Lu, Q. Ji, L. Liu, P. Ma, G. Yang, W. Hou and H. Xu, *Chem. Commun.*, 2020, **56**, 4692–4695.
16 A. Wieczorek, T. Buckup and R. Wombacher, *Org. Biomol. Chem.*, 2014, **12**, 4177–4185.
17 G. Knorr, E. Kozma, A. Herner, E. A. Lemke and P. Kele, *Chem. – Eur. J.*, 2016, **22**, 8972–8979.
18 Z. Novák and A. Kotschy, *Org. Lett.*, 2003, **5**, 3495–3497.
19 F. Pop, J. Ding, L. M. L. Daku, A. Hauser and N. Avarvari, *RSC Adv.*, 2013, **3**, 3218–3221.
20 A. M. Bender, T. C. Chopko, T. M. Bridges and C. W. Lindsley, *Org. Lett.*, 2017, **19**, 5693–5696.
21 N. Leconte, A. Keromnes-Wuillaume, F. Suzenet and G. Guillaumet, *Synlett*, 2007, 204–210.
22 A. Wieczorek, P. Werther, J. Euchner and R. Wombacher, *Chem. Sci.*, 2017, **8**, 1506–1510.
23 W. D. Lambert, Y. Fang, S. Mahapatra, Z. Huang, C. W. A. Ende and J. M. Fox, *J. Am. Chem. Soc.*, 2019, **141**, 17068–17074.
24 P. Fitton and E. A. Rick, *J. Organomet. Chem.*, 1971, **28**, 287–291.
25 A. Counotte-Potman and H. van der Plas, *J. Heterocycl. Chem.*, 1978, **15**, 445–448.
26 E. Ros, M. Bellido, X. Verdaguer, L. Ribas de Pouplana and A. Riera, *Bioconjugate Chem.*, 2020, **31**, 933–938.
27 J. Yang, M. R. Karver, W. Li, S. Sahu and N. K. Devaraj, *Angew. Chem., Int. Ed.*, 2012, **51**, 5222–5225.
28 Y. Qu, F. X. Sauvage, G. Clavier, F. Miomandre and P. Audebert, *Angew. Chem., Int. Ed.*, 2018, 57, 12057–12061.
29 W. Mao, W. Shi, J. Li, D. Su, X. Wang, L. Zhang, L. Pan, X. Wu and H. Wu, *Angew. Chem., Int. Ed.*, 2019, **58**, 1106–1109.
30 Y. Xie, Y. Fang, Z. Huang, A. M. Tallon, C. W. A. Ende and J. M. Fox, *Angew. Chem., Int. Ed.*, 2020, DOI: 10.1002/anie.202005569.
31 H. Wu and N. K. Devaraj, *Acc. Chem. Res.*, 2018, **51**, 1249–1259.
32 J. L. Seitchik, J. C. Peeler, M. T. Taylor, M. L. Blackman, T. W. Rhoads, R. B. Cooley, C. Refakis, J. M. Fox and R. A. Mehl, *J. Am. Chem. Soc.*, 2012, **134**, 2898–2901.
33 R. J. Blizzard, D. R. Backus, W. Brown, C. G. Bazewicz, Y. Li and R. A. Mehl, *J. Am. Chem. Soc.*, 2015, **137**, 10044–10047.
34 S. V. Mayer, A. Murnauer, M. K. von Wrisberg, M. L. Jokisch and K. Lang, *Angew. Chem., Int. Ed.*, 2019, **58**, 15876–15882.
35 H. S. Jang, S. Jana, R. J. Blizzard, J. C. Meeuwsen and R. A. Mehl, *J. Am. Chem. Soc.*, 2020, **142**, 7245–7249.

Trends in Biotechnology

CellPress
REVIEWS

Review

Learning from Nature to Expand the Genetic Code

Enric Ros [1,3] Adrian Gabriel Torres [1,3] and Lluís Ribas de Pouplana [1,2,*]

The genetic code is the manual that cells use to incorporate amino acids into proteins. It is possible to artificially expand this manual through cellular, molecular, and chemical manipulations to improve protein functionality. Strategies for *in vivo* genetic code expansion are under the same functional constraints as natural protein synthesis. Here, we review the approaches used to incorporate noncanonical amino acids (ncAAs) into designer proteins through the manipulation of the translation machinery and draw parallels between these methods and natural adaptations that improve translation in extant organisms. Following this logic, we propose new nature-inspired tactics to improve genetic code expansion (GCE) in synthetic organisms.

Highlights

Nature has evolved different strategies to expand the repertoire of translatable proteins using the standard genetic code.

The standard genetic code can be expanded through artificial cellular, molecular, and chemical manipulations.

Strategies found in nature to enhance protein synthesis and functionality have been applied to improve genetic code expansion (GCE).

A key aspect for efficient GCE relies on improving tRNA functionality.

Novel strategies inspired on nature can be devised to further optimize GCE.

The Amino Acid Repertoire of the Genetic Code Can Be Modified

The standard **genetic code** (see Glossary) is the 'dictionary' that cells use to translate genetic information into **proteins**. It is defined by sequences of nucleotide triplets (**codons**) that specify the **amino acids** that will be used for protein synthesis (see later). Key components of the **translation** machinery are **tRNAs** and the enzymes that aminoacylate them, the **aminoacyl tRNA synthetases** (AARS). Aminoacylated tRNAs pair with their complementary codons in the **ribosome**, thus delivering the amino acids to the growing protein chain (Box 1).

The genetic code is almost universal, degenerate (i.e., a given amino acid can be encoded by more than one codon), and designed to accommodate 20 chemically distinct natural amino acids and three stop signals in 64 codons (Box 1). Other amino acids are incorporated naturally, but these require important modifications of the basic translation machinery, possibly due to the limitations that the relatively simple tRNA structure imposes upon the increase in the informational content of the system [1]. Despite these limitations, the genetic code can be artificially manipulated to reduce [2], exchange [3], or expand [4] the standard amino acid repertoire it decodes. This can be achieved by growing strains under permanent metabolic selection [2], by using experimental set-ups with **auxotrophic** strains that exploit the natural substrate tolerance of AARS [3], or by genetically reassigning codons.

Genetic code engineering to exchange **canonical amino acids** with **ncAAs** (i.e., those being natural or **unnatural amino acids**) or to expand the standard set of amino acids *in vivo*, is an extremely active area of **synthetic biology**, referred to here as **genetic code expansion (GCE)**. The functional versatility that ncAAs can provide, in addition to the controlled site of incorporation in the protein sequence, provide countless promising biotechnological applications (see later).

In nature, a limited set of **post-translational modifications** (PTMs) affecting protein backbones or amino acid side chains modulate protein stability, activity, or subcellular localization [5,6]. Likewise, ncAAs can be designed to alter protein backbones or to carry amino acid side chain modifications, so that their incorporation into protein sequences can mimic natural PTMs [7]. For instance, marine

[1]Institute for Research in Biomedicine, The Barcelona Institute of Science and Technology, Barcelona, Catalonia, 08028, Spain
[2]Catalan Institution for Research and Advanced Studies, Barcelona, Catalonia, 08010, Spain
[3]These authors contributed equally to this work

*Correspondence: lluis.ribas@irbbarcelona.org (L. Ribas de Pouplana).

Trends in Biotechnology

CellPress REVIEWS

Box 1. The Main Components of the Translation Machinery

mRNAs carry the genetic information necessary for protein synthesis, encoded in nucleotide triplets called 'codons'. There are 64 possible codons in the genetic code. Three of these codons mark the end of translation ('stop codons'): amber (UAG), ochre (UAA), and opal (UGA). The other 61 codons encode 20 amino acids (Figure IA,B).

tRNAs are adaptor molecules that recognize codons on mRNAs and translate them into a specific amino acid (Figure IA). A nucleotide triplet on the tRNA (anticodon) specifically interacts with the mRNA codons. Some tRNAs recognize more than one codon by forming non-Watson–Crick interactions (tRNA 'wobbling'). At their 3′-end, tRNAs are charged with their cognate amino acid (AA) by aminoacyl-tRNA-synthetases (AARS). Importantly, tRNAs need to be chemically modified post transcriptionally (black stars) to become fully active. Modifications primarily affect tRNA maturation, aminoacylation, structure, and codon:anticodon pairing.

Ribosomes (Figure IA) are large ribonucleoprotein complexes. They comprise two subunits, each containing a different number of rRNAs and ribosomal proteins. Ribosomes present the mRNA for decoding, distinguish between correct and incorrect tRNA species, catalyze peptide bond formation and release the newly formed polypeptide. There are three tRNA-binding sites within the ribosome: the 'acceptor' (A; that receives the incoming charged tRNA), 'peptidyl' (P; where peptide bond formation among contiguous AAs occur) and 'exit' (E; where the uncharged tRNA leaves the ribosome) sites.

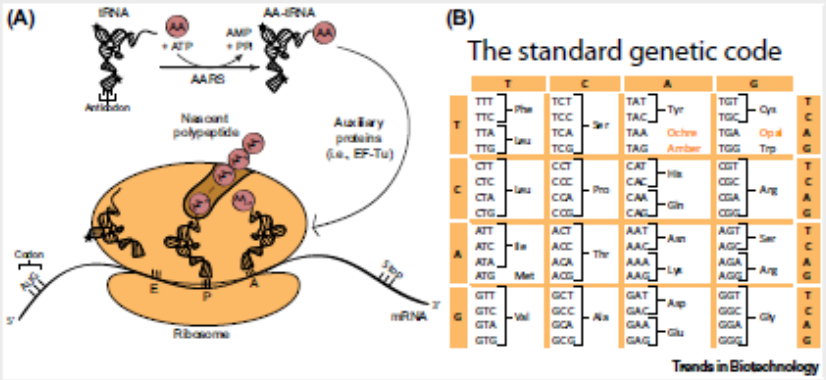


Figure I. Translation and the Genetic Code. (A) Schematic representation of the translation machinery. (B) Table representing the sequence of every codon and the correspondence to their cognate amino acid. Ochre, amber, and opal represent stop codons. Abbreviations: AA, amino acid; AARS, aminoacyl tRNA synthetase.

mussel foot proteins are natural water-resistant bioadhesives that can be used to develop healthcare materials [8]. Their adhesive properties depend upon post-translational hydroxylation of amino acids that can only be performed by eukaryotic cells [9]. The incorporation of hydroxyproline as a ncAA enables the synthesis of functional proteins in bacteria, making their production more cost- and time-effective [9].

However, synthetically incorporating ncAAs into proteins has applications beyond PTM mimicking. For example, they can be used to trigger protein cleavage at specific positions [10], an approach used to engineer autmaturing microbial enzymes for the food-processing industry [11]. They can also be used to increase protein solubility and half-life through bioconjugation with, for example, polyethylene glycol (PEG) chains [12]. In laboratory research, ncAAs can be used for *in vivo* live cell imaging [13], as probes for nuclear magnetic resonance [14], or to characterize new protein–protein interactions in their native environment [15]. In biomedicine, ncAAs can be used to generate light-activated therapeutics [16], *in vivo* activation of prodrugs [17], or immunogenic epitopes in vaccine development [18]. Indeed, various antibody–drug conjugates bearing ncAAs are currently in clinical trials [19].

Glossary

Amino acid: building blocks of proteins. They contain an amino group, a carboxylic acid group, and a unique side chain.
Aminoacyl tRNA synthetase (AARS): enzyme that catalyzes aminoacylation.
Aminoacylation: incorporation of an amino acid with its cognate tRNA.
Anticodon: unit of three nucleotides in the tRNA that recognizes a given codon through complementary base pairing.
Auxotrophic strain: a variety of an organism that is unable to synthesize a specific organic compound necessary for growth.
Canonical amino acids: natural amino acids decoded by the standard codon: amino acid association rule of the genetic code.
Codon: unit of three nucleotides in mRNAs that encodes a specific amino acid.
Codon reassignment: use of a given codon to encode a different amino acid than that defined by the genetic code.
Codon usage: frequency at which codons are used in translation.
Efficiency (of translation): rate at which a given mRNA is translated.
Extant: currently existing.
Fidelity (of translation): recognition accuracy between each codon and the incorporated amino acid.
Genetic code: set of rules that govern the translation of codons into specific amino acids.
Genetic code engineering: process by which the genetic code is modified so that codon:amino acid correspondences are different to the natural ones.
Genetic code expansion (GCE): artificially modified genetic code able to encode more than the 22 natural amino acids. Here, we also consider the exchange of canonical for noncanonical amino acids by genetic code engineering as part of GCE.
Mistranslation: inappropriate decoding of codons that alters protein sequences upon translation.
mRNA: RNA derived from gene transcription that defines, based on its codon composition, the amino acid sequence of proteins.
Natural amino acids: the set of 22 naturally occurring proteinogenic amino acids. Two of them (pyrrolysine and selenocysteine) are not universally present.

Trends in Biotechnology

CellPress
REVIEWS

Thus, the development of new ncAAs has the potential to create limitless variants of building blocks for proteins, and their incorporation via GCE offers numerous biotechnological applications. However, GCE is a complex endeavor with multiple technical constraints that currently limit its applications. Here, we review the natural mechanisms of translation optimization used by **extant** organisms to enhance the functionality of their proteomes, and compare them to the strategies used in synthetic biology for GCE (Figure 1, Key Figure). From this comparison, new approaches can be formulated to improve GCE. We focus on strategies resulting from optimizing tRNA functionality, and do not cover methods that are not inspired by natural adaptations of the translation machinery, such as the use of four- and five-nucleotide codons [20,21] or unnatural base pairs [22,23].

Natural and Synthetic Gene Translation

Genetic translation involves four main steps: initiation (recognition of the first codon on **messenger RNAs** by ribosomes), elongation (translation of successive codons into an amino acid sequence), termination (ribosomal detection of a stop codon), and recycling (release of the polypeptide chain and ribosomal disassembly). Each step requires many auxiliary proteins [24] and needs to be tightly controlled and optimized for **fidelity** and **efficiency** [25]. Fidelity is achieved through precise recognition mechanisms between all components of the process (Box 1). Efficiency is influenced by the rates of initiation and elongation, a key determinant being transcript codon composition versus abundance of different tRNAs (**codon usage**) [25].

To the same extent, effective generation of synthetic proteins via GCE requires an efficient and faithful translation machinery. This is typically achieved through the use of an **orthogonal translation system** (OTS) (Box 2) able to incorporate the desired ncAA into the protein(s) of interest, without affecting (or be affected by) the natural translation machinery of the cell. Most *in vivo* strategies for GCE require the reassignment of the identity of a codon to the desired ncAA. Thus, most OTSs include a tRNA specific for the reassigned codon and an engineered AARS that specifically aminoacylates this tRNA with the ncAA of interest. Most often, orthogonal AARS/tRNA pairs are derived from species phylogenetically distant to the host, and can be engineered to accept a variety of ncAAs [26] (Table 1). Ideally, the orthogonal AARS should be unreactive towards natural amino acids and endogenous tRNAs; the tRNA should only be a substrate for the **orthologous** AARS; and the ncAA should not be recognized by the AARS of the host. These requirements are identical to those placed upon the components of natural translation machineries by natural selection. Thus, natural strategies that improve the translation of the genetic code and enhance protein synthesis can also be applied to improve GCE.

Nature's Laboratory for Novel Protein Synthesis

The total number of possible codon combinations on mRNAs represents an astronomical amount of feasible protein sequences, only a small fraction of which are found in nature [27,28]. Most soluble protein structures evolved through the assembly of a limited number of small domains and motifs [29], which explains the relatively limited conformational space presented by extant proteomes. A second factor contributing to this uniformity may be limitations of the translation machinery with regards to certain codon combinations, as evidenced by the discovery of adaptations in ribosomes, mRNAs, tRNAs, AARS, and elongation factors that facilitate the translation of particular codon combinations (Figure 1) [30].

For example, the synthesis of proteins containing poly-proline stretches requires an elongation factor (EF-P in bacteria and eIF5A in eukaryotes) that prevents ribosome stalling at two or more contiguous proline codons [31] (Figure 1A). Likewise, synthesis of hydrophobic proteins by mitochondrial ribosomes requires a remodeled ribosomal polypeptide exit channel to facilitate insertion of these proteins into mitochondrial membranes [32] (Figure 1B).

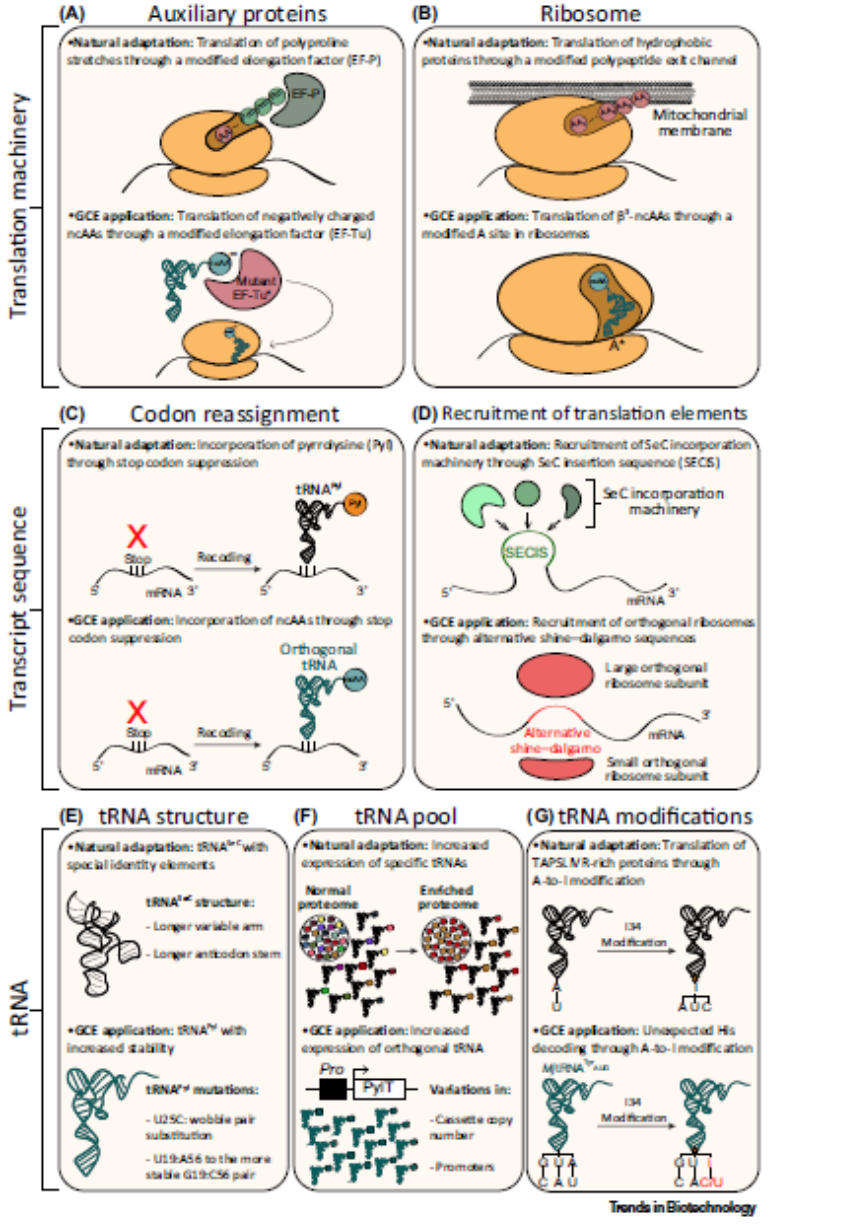
Noncanonical amino acids (ncAAs): amino acids that are incorporated into proteins without following natural codon: amino acid association rules. These are usually unnatural amino acids, but may include natural ones.
Orthogonal translation system (OTS): supplementary translation machinery engineered to permit the participation of ncAAs in translation within a given organism.
Orthologous: evolved from the same common ancestor after a speciation event.
Post-translational modifications: modifications to the protein backbone or the amino acid side chains that occur after ribosomal protein synthesis.
Protein: biopolymer comprising amino acids linked together through peptide bonds.
Ribosome: complex formed by assemblies of rRNA and specific proteins that catalyzes peptide bond formation between amino acids during translation.
Riboswitches: RNA structures present in mRNAs that, upon binding of small molecules, regulate the activity of such mRNA.
Ribozymes: RNA-based enzymes.
Selenocysteine insertion sequence (SECIS): motif present in certain mRNAs that recruits elements required for selenocysteine incorporation into protein sequences.
Suppressor tRNA: tRNA bearing an anticodon sequence complementary to a stop codon used to incorporate an amino acid at that position.
Synthetic biology: a research field aimed at engineering living organisms with novel or enhanced abilities.
Translation: process by which a sequence contained in the mRNA is processed to biosynthesize proteins.
tRNA: nonprotein-coding RNA that decodes the information contained in mRNA codons through base-pairing interactions with its anticodon. Each tRNA carries a specific amino acid to be used in translation.
tRNA base modification: post-transcriptional addition of functional groups to the nucleotides of tRNAs.
Unnatural amino acid: synthetic amino acids absent in nature. They are structurally different from the 22 natural amino acids that can participate in translation.

Trends in Biotechnology

CellPress REVIEWS

Key Figure

Comparison of Strategies to Improve Translation Found in Nature and Applied to Genetic Code Expansion (GCE)



(See figure legend at the bottom of the next page.)

Trends in Biotechnology

CellPress REVIEWS

Nature can also enhance protein functionalities through **codon reassignment** to incorporate new amino acid side chains (Figure 1C). Some methanogenic archaea and bacteria use a specific AARS/tRNA pair [33] to reassign the amber (UAG) stop codon to a sense codon and incorporate the amino acid pyrrolysine (Pyl) into methyltransferases to enhance their activity [34]. Similarly, incorporation of selenocysteine (SeC) into proteins requires several adaptations of the translational apparatus. SeC is an analog of cysteine that contains selenium instead of sulfur, and displays a lowered reducing potential that protects proteins from oxidative stress effects [35]. Use of coded SeC in translation requires a specific tRNA to suppress an opal (UGA) stop codon, a '**selenocysteine insertion sequence**' (SECIS) element that generates a specific secondary structure on the mRNA (Figure 1D), a cognate tRNA initially aminoacylated with Ser, a dedicated SeC synthetase to convert Ser-tRNA^{SeC} into SeC-tRNA^{SeC}, and a specific elongation factor to bring SeC-tRNA^{SeC} to the ribosome [36].

Nature is also capable of reassigning sense codons to generate an ambiguous proteome for enhanced functionality [37]. A remarkable example of this has been described in *Candida albicans*, where a CUG codon is translated with serine or leucine in different proportions according to the environmental needs of the organism [38].

Applying Natural Strategies to GCE

Successful approaches reported to improve GCE show strong parallelisms to Nature's strategies to improve protein functionality. These include the engineering of mRNAs via reassignment of sense and nonsense (stop) codons, modification of translation elongation and termination factors, and the synthetic generation of specialized ribosomes.

Sense codon reassignments usually require low-frequency codons [39,40], as is the case for start codons (AUG) in the mRNA used to incorporate methionine analogs [41]. This strategy can be used to develop tools for monitoring *de novo* protein synthesis. However, sense codon reassignment can also result in indiscriminate incorporation of ncAAs across the proteome, and may induce cellular stress responses [42]. More commonly, in-frame stop codons recognized by engineered **suppressor tRNA** are used to incorporate the ncAA of interest, mirroring the natural examples of Pyl and SeC [33,36]. This strategy allows insertion of the ncAA at the desired position of a synthetic mRNA that will be translated into a target protein [4] (Figure 1C). This approach mimics the natural incorporation of Pyl by tRNA$_{CUA}$ through the recognition of UAG stop codons. By contrast, use of SeC in translation requires not only a tRNA that decodes a stop codon (tRNA$_{UCA}$), but also additional metabolic transformations (i.e., Ser-tRNA^{Sec} to SeC-tRNA^{Sec} conversion), and structural signals in the translated transcript for its efficient incorporation into proteins [36].

Modifications to elongation factors that improve their affinity for suppressor tRNAs charged with ncAA have been designed. For example, modifications in the amino acid-binding domain of elongation factor thermo-unstable (EF-Tu) allow this factor to accommodate ncAAs that serve as mimics of PTMs [7] (Figure 1A). This indicates that, just as the elongation factor EF-P is naturally required for efficient translation of consecutive proline codons [31], improved incorporation of ncAAs through GCE requires specialized elongation factors. In addition, modified components

Figure 1. Adaptations to modulate translation efficiency or expand translation machinery capacities found in nature are depicted ('natural adaptation'). These include improvement of translation factor recognition (A); structural alteration of ribosomes (B); genetic code changes through codon reassignment (C) and insertion of sequences that recruit translation elements to mRNAs (D); and modulation of the cellular tRNA structure (E), population ('tRNA pool') (F), or modification patterns (G). Equivalent strategies applied to GCE are depicted ('GCE application'). Abbreviations: ncAA, noncanonical amino acid; SeC, selenocysteine; TAPSILVR, threonine, alanine, proline, serine, isoleucine, leucine, valine, and arginine.

Trends in Biotechnology

CellPress REVIEWS

Box 2. The Main Components of an Orthogonal Translation System

An engineered OTS for GCE is formed by an orthologous AARS capable of aminoacylating its cognate tRNA with the ncAA of choice (Figure I). While some natural AARS in their wild-type form can incorporate numerous ncAAs efficiently (e.g., PylRS), most often the selected AARS require mutagenesis to expand its substrate recognition and to avoid interactions with its natural amino acid, thereby achieving orthogonality.

In general, tRNA recognition by the orthologous AARS is not dependent on the anticodon loop, which permits the modification of the anticodon sequence to reassign a given codon of choice (depicted as XYZ in Figure I). The aminoacylated tRNA needs to be structurally compatible with auxiliary translation factors and ribosomes from the host organism to successfully incorporate the ncAA to the target protein.

An optimal OTS must not cross-react with the natural amino acids (AA) of the cell, its endogenous AARS, or tRNAs (continuous black arrows in Figure I). If any of these processes take place (loss of orthogonality), a detrimental effect for the functionality of the OTS is produced (dashed-gray arrows in Figure I). On the one hand, if the ncAA is ligated to an endogenous tRNA (either because the orthologous AARS recognizes endogenous tRNAs or because an endogenous AARS can recognize the ncAA), it will result in a proteome-wide misincorporation of the ncAA. By contrast, if the orthogonal tRNA is aminoacylated with a natural amino acid (either because the endogenous AARS recognizes the orthogonal tRNA or because the orthologous AARS can recognize the natural amino acid), that amino acid will be incorporated into the reassigned codon, thereby lowering the fidelity of the OTS. To prevent any of these processes from happening, several rounds of positive and negative selection in the OTS screening for each ncAA are performed.

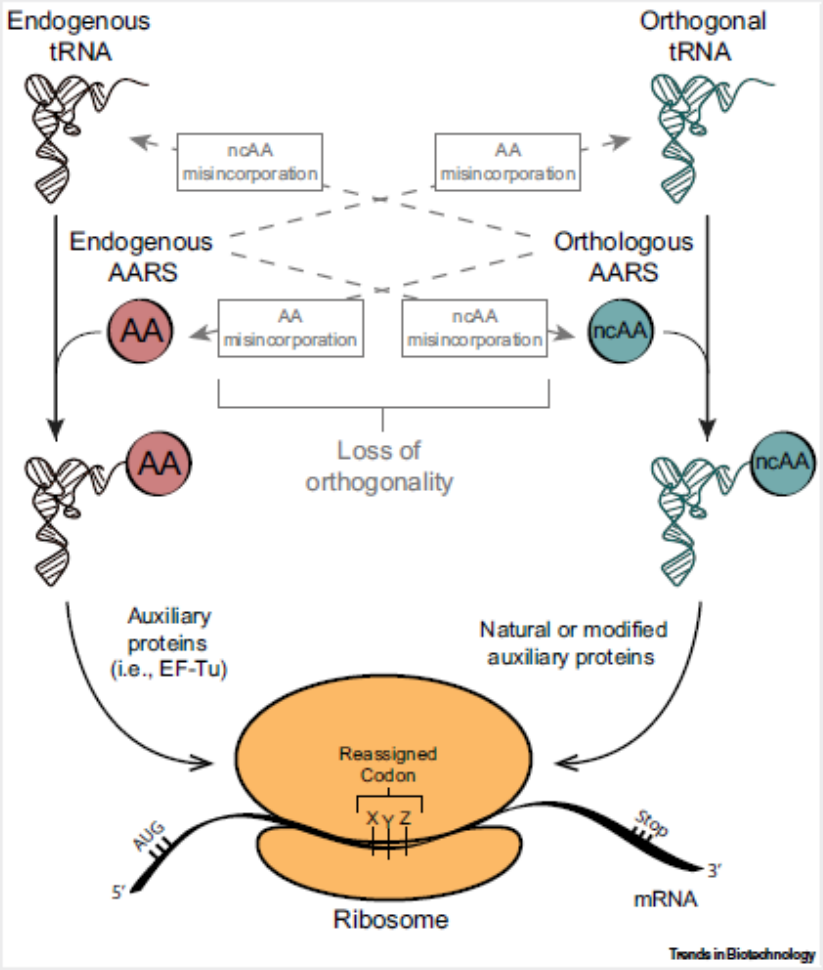
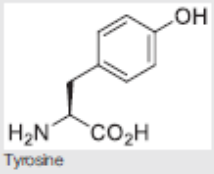
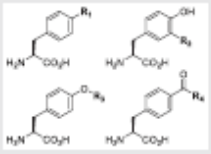
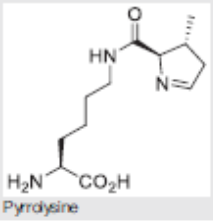
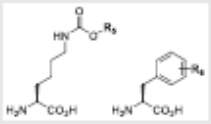
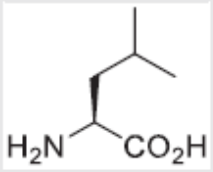
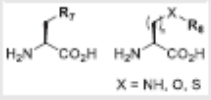


Figure I. Schematic Representation of an Orthogonal Translation System (OTS). Desired and undesired interactions are depicted in black and gray, respectively. Abbreviations: AA, amino acid; AARS, aminoacyl tRNA synthetase; ncAA, noncanonical amino acid.

Table 1. Engineered OTSs[a] and Structure of Incorporated ncAAs[b]

Species of origin	AARS	Orthogonality in	Canonical amino acid	Incorporated ncAA[c]	Refs
Methanocaldococcus jannaschii	Tyrosyl-RS (TyrRS)	Prokaryotes	 Tyrosine		[17,68,73]
Escherichia coli		Yeast (and mammalian cells if coexpressed with *Bacillus stearothermophilus* $tRNA^{Tyr}_{CUA}$)			[74,90]
Methanosarcina barkeri/mazei	Pyrrolysyl-RS (PylRS)	*E. coli*, yeast, and mammalian cells	 Pyrrolysine		[67,70,72,75]
E. coli	Leucyl-RS (LeuRS)	Yeast and *Caenorhabditis elegans*	 Leucine	 X = NH, O, S	[91,92]

[a]Only the most commonly used AARS–tRNA pairs are represented.
[b]For a detailed summary of incorporated ncAA, see [26].
[c]R_1: *M. jannaschii* = H, CH_3, N_3, NH_2, Ph, CN, I, Br, NO_2, OCF_3, OSO_2H. R_2: *M. jannaschii* = OH, NH_2, I, NO_2; *E. coli* = N_3, I. R_3: *M. jannaschii* = CH_3, CF_3, $C(CH_3)_3$, benzyl, propargyl; *E. coli* = CH_3, propargyl. R_4: *M. jannaschii* = CH_3, phenyl, 4-F-phenyl, 4-I-phenyl, 4-NO_2-phenyl; *E. coli* = CH_3, phenyl. R_5: $C(CH_3)_3$, cyclopentyl, propenyl, butenyl, propargyl, BCN, TCO, norbonenyl, 4-Hbenzyl, 4-NO_2-benzyl. R_6: (*ortho*) H, CH_3, Cl, Br, I, OCH_3, CN, NO_2; (*meta*) CH_3, CF_3, OCH_3, NO_2, CN, N_3, F, Cl, Br, I; (*para*) 2-Cl-ethoxy, 2-Br-ethoxy, 2-I-ethoxy, 2-N_3-ethoxy. R_7: butenyl, penyl, pentenyl, hexyl, 4-OCH_3-phenyl. R_8: (X=NH) naphtalenyl, 7-Ac-naphtalenyl; (X=O) pentenyl, 2-pentenyl; (X=S) 2-NO_2-benzyl.

of the translation termination machinery have been used to prevent premature elongation termination and optimize ncAA incorporation. In prokaryotes, deletion of the release factor that recognizes amber and ochre stop codons (RF1) can be accomplished if the function of the complementary release factor (RF2) is enhanced [43], or if amber stop codons are replaced by ochre stop codons genome-wide [44,45]. Under these conditions, cellular fitness is not compromised. In eukaryotes, release factor 1 (eRF1) recognizes all three stop codons, limiting the possibilities to modulate this function [46]. However, eRF1 mutants were shown to improve the efficiency of multisite ncAA incorporation in mammalian cells [47].

Finally, the ribosome itself has been the subject of modifications to allow the incorporation of ncAAs, following the example of the structural modifications found in mammalian mitochondrial ribosomes [32]. For example, the biosynthesis of proteins carrying β^3-amino acids, which are not tolerated by natural ribosomes, was achieved through the use of variant ribosomes carrying mutations near the aminoacyl-tRNA binding site [48] (Figure 1B). β^3-amino acids can confer resistance to proteolysis, a property used in the design of small peptides for clinical uses [49], but unexploited in the field of protein engineering.

A more drastic approach is the selection of orthogonal ribosomes that only recognize a subset of designed orthogonal mRNAs with specific ribosome-binding sites [50,51] (Figure 1D). Given that

these orthogonal ribosomes are not required for ordinary translation, they can tolerate mutations that would otherwise be lethal. Different variations of this approach have enabled the successful translation of proteins using quadruplet codons [20] or polyproline sequences in the absence of EF-P [52].

Transfer RNAs as a Hub for Protein Synthesis-Enhancement Strategies

tRNAs are important targets of natural adaptations that expand the range of accessible protein sequences to organisms, including: (i) new recognition motifs to generate new tRNA identities; (ii) mechanisms of tRNA pool modulation; and (iii) **tRNA base modifications** for enhanced ribosomal elongation or codon reassignment.

The modification of tRNA structure for the generation of new identities is evident in $tRNA^{Sec}$, which, in contrast to canonical tRNAs, presents several unique structural features, such as a long variable arm and a longer **anticodon** stem required for the specific **aminoacylation** of $tRNA^{Sec}$ by seryl-tRNA synthetase [36] (Figure 1E).

Modulation of the tRNA pool in a cell-specific manner is essential in specialized tissues dedicated to the massive production of specific polypeptides of highly biased composition. For example, the synthesis of silk fibers by the salivary glands of some arthropods requires an extreme enrichment in tRNAs cognate for alanine, glycine, and serine to satisfy the high demand for these amino acids needed by ribosomes dedicated to the translation of silk-forming polypeptides [53,54] (Figure 1F).

Post-transcriptional modifications on tRNAs enhance translation at different levels (preventing **mistranslation**, improving codon recognition, generating codon reassignments, etc.). While some modifications are conserved across the three domains of life, others are unique to subsets of species [55]. For example, tRNAs genetically encoded with adenosine at the wobble position of their anticodon (position 34; A34) are predominant in eukaryotic organisms, where they are modified to inosine (I34) [56–58]. This modification is present in tRNAs cognate for the amino acids threonine, alanine, proline, serine, leucine, isoleucine, valine, and arginine (TAPSLIVR), and it is required for the synthesis of TAPSLIVR-rich sequences, including those containing mucin-like domains (MLDs) [59]. TAPSLIVR-rich stretches are more abundant in Eukarya than in Bacteria, and the eukaryotic genes encoding these proteins prefer codons recognized by I34-tRNAs [58,60,61]. Thus, I34 is a tRNA modification that enhances the translation machinery of eukaryotes to synthesize proteins that are otherwise scarce in prokaryotes (Figure 1G).

Importantly, tRNA modifications can also permit codon reassignments. In some mitochondria, $tRNA^{Met}_{CAU}$ is capable of reading AUA (isoleucine) codons upon modification of its cytosine at position 34 (C34) to 5-formylcytosine (f^5C) [62], thereby generating an alternative translation initiation codon.

Lastly, tRNA modifications have also evolved to improve tRNA aminoacylation. For example, post-transcriptional addition of a guanosine residue at the 5′-end (G^{-1} modification) of $tRNA^{His}$ allows aminoacylation by histidine-tRNA synthetase [63], while 1-methylguanosine at position 37 (m^1G37) of $tRNA^{Asp}$ blocks charging of this tRNA by argininyl-tRNA synthetase [64]. The primary purpose of these identity elements is to prevent unspecific aminoacylation, but they are also important for tRNA synthesis, maturation, recycling, and degradation.

Enhanced ncAA Incorporation by Improving tRNA Functionality

In a fashion similar to natural adaptations, structural changes in orthogonal tRNAs can enhance ncAA incorporation. Evolution of an orthogonal tRNA for the incorporation of two PTM analogs

(phosphoserine and phosphothreonine) resulted in tRNAs with better AARS selectivity, leading to increased translational efficiency [65] and orthogonality [66]. Similarly, rational mutations on the tRNA that improve its interaction with EF-Tu increased incorporation efficiency for several ncAAs [67–69] (Figure 1E). Lastly, in $tRNA^{Pyl}$, the substitution of a canonical U19:A56 base pair by G19:C56, or a U25C mutation designed to generate a stable G:C pair, were shown to increase suppressor efficiency by increasing tRNA stability [47,70–72].

By contrast, natural modulations of the tRNA pool to enhance amino acid incorporation (e.g., adaptations for silk synthesis) are mirrored in the need for fine-tuned expression of orthogonal tRNAs for optimal ncAA incorporation [47] (Figure 1F). The latter is often achieved by boosting the orthogonal tRNA expression through the use of constitutive promoters, such as the *Escherichia coli* $tRNA^{Pro}_{CGG}$ promoter (proK) [71,73], or the eukaryotic RNA polymerase III promoters H1 [74] or U6 [47,75]. However, high orthogonal tRNA expression levels may not always result in a beneficial outcome [70,72,76], indicating that maintaining a balanced cellular homeostasis is essential for the correct function of the OTS.

Analogous to natural translation, the efficiency of orthogonal expression systems can be equally affected by post-transcriptional tRNA modifications (Figure 1G), because the activity of the OTS in the host may depend on the presence of the correct modifications in the orthogonal tRNAs. For instance, O-phosphoserine incorporation in bacteria benefits from the methylation of the suppressor tRNA by the archaeal enzyme Trm5 [77]. In addition, endogenous tRNA modification enzymes of the host organism can also interact with the orthogonal tRNA and modulate its function. For example, deletion of the bacterial enzyme MiaA, responsible for N6-isopentenyladenosine at position 37 (ms^2i^6A37) in native tRNAs, is detrimental to OTSs [76], while its overexpression improves orthogonality [77]. A similar example was described for a family of bacterial tRNA pseudouridine synthases [67,77].

Post-transcriptional tRNA modifications can also affect codon discrimination. For instance, expression of *Methanocaldococcus jannaschii* $tRNA^{AUG}$ in *E. coli* resulted in the decoding of histidine codons CAC and CAU due to an I34 modification introduced by the bacterial adenosine deaminase TadA. Expressing orthogonal tRNAs that are not substrates of TadA prevented A-to-I editing, allowing discrimination of both histidine codons [78].

New Nature-Inspired Avenues for Improved GCE

Growing evidence indicates that ribosomal populations in cells are not homogenous, and this heterogeneity was shown to define the type of messages that each ribosome can translate [79,80]. Thus, orthogonal ribosomes incorporating alternative ribosomal proteins, or specific post-transcriptional modifications, could be used as transcript-specific translation machineries to improve the efficiency of OTSs (Figure 2A) [50,51]. In addition, incorporation of *cis*-acting elements on mRNAs may also be used to enhance their recognition by specialized ribosomes [81], so that only a targeted subpopulation of the translation machinery would be hijacked, thereby minimizing the stress exerted by the OTS on the host organism.

Subcellular compartmentalization is a strategy used by biological systems to increase local concentrations of components of metabolic reactions and improve their efficiency. In the context of GCE, the engineering of AARS/tRNA pairs capable of separating in subcellular compartments could increase aminoacylation efficiency and fidelity (Figure 2B). It was recently shown that a certain degree of compartmentalization can be achieved by fusing protein domains known to form liquid droplets to the orthogonal AARS [82]. It is likely that developing novel ways to produce such subcellular clusterization will further increase the efficiency of the OTS.

Trends in Biotechnology

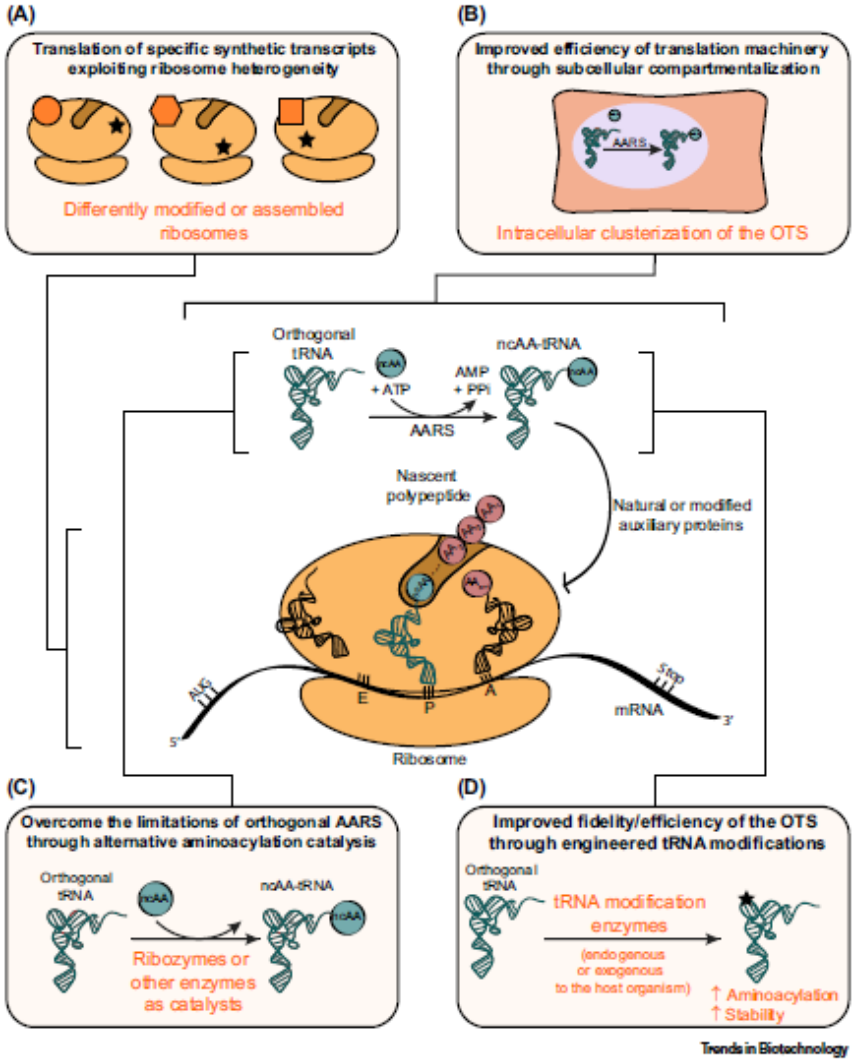


Figure 2. Potential Nature-Inspired Strategies to Increase Orthogonal Translation System (OTS) Efficiency. Expanding ribosome heterogeneity by modulating the level of association of ribosomal proteins (dark-orange circle, hexagon, and square) or post-transcriptional rRNA modifications (black stars) could lead to improved translation of specific synthetic transcripts (A). Generating artificial subcellular compartmentalization of the OTS, for example by fusing liquid droplet-forming protein domains to the orthologous aminoacyl tRNA synthetase (AARS), can improve the efficiency of the biochemical reactions required for translation (B). Alternative ways to aminoacylate tRNAs, such as those based on ribozymes, could further expand the limitations of current orthogonal AARSs (C). Natural post-transcriptional modifications (black stars) occurring on orthologous tRNA can be exploited to improve tRNA stability and aminoacylation (D). Abbreviations: ncAA, noncanonical amino acid;

Selection of orthogonal AARS/tRNA pairs with improved functionality is technically challenging and could benefit from new methods of directed evolution. In bacteria, phage-assisted evolution has proven useful to select AARS/tRNA pairs with enhanced recognition, lower promiscuity, and

Trends in Biotechnology

CellPress
REVIEWS

increased catalytic activity [83,84]. In eukaryotes, exploiting unconventional natural strategies for efficient recognition of tRNAs and tRNA-like structures by AARS may as well be a suitable strategy to generate AARS/tRNA pairs with higher efficiency. For instance, HIV-1 requires the interaction with an AARS/tRNA complex during viral assembly, and uses the tRNA molecule as primer for the reverse transcriptase-catalyzed synthesis of viral DNA [85]. Similarly, the expression of amino acid metabolism-related genes is regulated by T-box **riboswitches** that gauge amino acid availability through the binding of cognate tRNAs [86]. Similar mechanisms that specify tRNA recognition to a readily detectable signal could be used to engineer selection strategies for the identification of new orthologous AARS/tRNA pairs.

The selection of AARS variants capable of recognizing new ncAAs is an important bottleneck in GCE, but aminoacylation reactions are not solely catalyzed by AARS. It is believed that primitive **ribozymes** were responsible for charging amino acids onto tRNAs, or equivalent molecules, before the appearance of AARS. Moreover, flexizymes are a set of artificial ribozymes capable of *in vitro* aminoacylating tRNAs with a range of natural and unnatural amino acids [87]. Developing *in vivo* aminoacylating ribozymes, or other ester-forming enzymes, to aminoacylate orthogonal tRNAs with ncAAs would be a new strategy to achieve ncAA-tRNA ligation while minimizing cross-reactivity with endogenous translation apparatuses (Figure 2C).

Post-transcriptional tRNA modifications can also be further exploited to improve OTSs. Most tRNA modifications to enhance GCE have been applied in the context of codon:anticodon recognition, but their influence upon the aminoacylation efficiency of orthogonal tRNAs with ncAAs has not been explored. Modifications in the acceptor stem and in the anticodon region of the tRNA could be used to improve the specific charging of orthogonal tRNAs with ncAAs of interest (Figure 2D) [63,64]. Lastly, understanding further the modification patterns that orthogonal tRNAs undergo in eukaryotic hosts will pave the way for improving their functionality.

Concluding Remarks and Future Perspectives

The list of known natural adaptations of the translation apparatus to synthesize proteins with special sequences continues to grow, and includes changes in tRNA modifications, alterations to ribosomal structure, differential functionality of translation factors, and the modulation of mRNA codon composition, among others. GCE has mimicked some of these strategies to improve the variety of ncAAs efficiently incorporated into diverse biological systems. Most likely, a vast number of additional nature-inspired strategies exist that can be applied to improve GCE, although several key questions remain to be explored (see Outstanding Questions).

From the protein-engineering perspective, the number of newly incorporated ncAAs with different structures and functional groups is constantly growing. However, the effects of incorporated ncAA upon protein conformation is still a neglected topic. Typically, optimal positions to incorporate ncAAs in proteins are determined by trial and error. Improvements in predictive models of protein dynamics based on machine learning will enable the rational identification of optimal ncAA incorporation sites, minimizing protein perturbation and maximizing desired activities.

At the cellular level, there is currently a limitation on the number of different ncAAs that can be encoded simultaneously within the same host. In this sense, it will be key to develop OTSs capable of containing multiple orthogonal AARS/tRNA pairs that can coexist in biological systems. A recent report showed that the incorporation of three distinct ncAAs could be achieved by triply orthogonal PylRS/tRNA pairs [88]. The discovery of more identity elements on naturally occurring translation systems, as well as better understanding of the processes to which the components of the OTS are subjected to within the host (e.g., modification patterns of orthogonal tRNAs) will

Outstanding Questions

What other protein properties can be modulated by the development of novel ncAAs? Will nature-inspired strategies to modify the translation machinery aid in their efficient incorporation by GCE?

To what extent can the number of different ncAAs simultaneously encoded *in vivo* be expanded? Can natural identity elements be used to generate a larger variety of mutually orthogonal AARS/tRNA pairs? Is the life cycle of orthogonal tRNAs comparable to that of endogenous natural tRNAs? Are they equally produced, processed and modified?

How do tRNA modification enzymes present in host organisms affect the functionality of orthogonal tRNAs? Will it be possible to artificially modulate tRNA modification patterns of orthogonal tRNAs to improve their aminoacylation by orthogonal AARS?

Can nature inspire the development of other means for tRNA aminoacylation (e.g., flexizymes) to be used *in vivo* for GCE as an alternative to orthogonal AARS/tRNA pairs?

Can nature-inspired strategies be applied to improve GCE in complex, multicellular organisms?

Will nature-inspired strategies aid in the development of more efficient methods for selecting orthogonal AARS/tRNA pairs? Will these be applied to other biological systems besides bacteria (e.g., mammalian cells)?

Which other nature-inspired techniques can be used to improve GCE?

likely aid in this task. The possibility of eventually obtaining a nondegenerate genetic code in which all synonymous codons are reassigned to different ncAAs is an exciting scenario that may one day become feasible as more mutually orthogonal AARS/tRNA pairs are described.

Just as natural evolution progressed from unicellular microorganisms to more complex living entities, genomically recoded organisms will experiment the same evolution. In this context, the outburst of novel gene-editing tools (e.g., CRISPR/Cas9 technology), opens a way to engineer higher forms of life with constitutively expanded genetic codes. While there are seminal contributions in this subject [89], the implementation of this approach is in its infancy. However, whole-organism engineering holds great promise to help understand protein function in its native environments, or to mass-produce new biomaterials with enhanced characteristics.

We foresee that GCE will touch upon all aspects of biotechnological manufacturing. Industrial materials will be enhanced in stability, reactivity, and recycling properties. Chemical reactions will be accelerated, modified and better controlled through the use of biological catalysts. Agricultural technologies will benefit from crops with improved yields and novel nutritional properties that might, for example, help reduce the need for animal consumption. Research tools for the visualization and kinetic monitoring of macromolecules *in vivo* will be developed, and targeted therapies based on the enhanced functionalities of modified amino acid side chains will become widespread.

Acknowledgments

This work was funded through a grant from the Spanish Ministry of Economy, Industry and Competitiveness (MINECO, BIO2015-64572-R to L.R.d.P.). E.R. thanks the support of funding from La Caixa Foundation (ID 100010434) and La Caixa PhD Fellowship.

References

1. Saint-Léger, A. *et al.* (2016) Saturation of recognition elements blocks evolution of new tRNA identities. *Sci. Adv.* 2, e1501860
2. Pezo, V. *et al.* (2013) A metabolic prototype for eliminating tryptophan from the genetic code. *Sci. Rep.* 3, 1359
3. Hoesl, M.G. *et al.* (2015) Chemical evolution of a bacterial proteome. *Angew. Chem. Int. Ed.* 54, 10030–10034
4. Chin, J.W. (2017) Expanding and reprogramming the genetic code. *Nature* 550, 53–60
5. Millar, A.H. *et al.* (2019) The scope, functions, and dynamics of posttranslational protein modifications. *Annu. Rev. Plant Biol.* 70, 119–151
6. Müller, M.M. (2018) Post-translational modifications of protein backbones: unique functions, mechanisms, and challenges. *Biochemistry* 57, 177–185
7. Chen, H. *et al.* (2018) Recent development of genetic code expansion for posttranslational modification studies. *Molecules* 23, 1662
8. Pandey, N. *et al.* (2020) Mussel-inspired bioadhesives in healthcare: design parameters, current trends, and future perspectives. *Biomater. Sci.* 8, 1240–1255
9. Lamegda, M. *et al.* (2012) Congeneric bio-adhesive mussel foot proteins designed by modified prolines revealed a chiral bias in unnatural translation. *Biochem. Biophys. Res. Commun.* 421, 646–650
10. Guo, J. *et al.* (2008) Addition of an α-hydroxy acid to the genetic code of bacteria. *Angew. Chem. Int. Ed.* 47, 722–725
11. Ohtake, K. *et al.* (2018) Engineering an automaturing transglutaminase with enhanced thermostability by genetic code expansion with two codon reassignments. *ACS Synth. Biol.* 7, 2170–2176
12. Young, D.D. and Schultz, P.G. (2018) Playing with the molecules of life. *ACS Chem. Biol.* 13, 854–870
13. Sakin, V. *et al.* (2017) A versatile tool for live-cell imaging and super-resolution nanoscopy studies of HIV-1 env distribution and mobility. *Cell Chem. Biol.* 24, 635–645.e5
14. Zhang, F. *et al.* (2018) A genetically encoded 19 FNMR probe for lysine acetylation. *Chem. Commun.* 54, 3879–3882
15. Nguyen, T-A. *et al.* (2018) Expanding the genetic code to study protein-protein interactions. *Angew. Chem. Int. Ed.* 57, 14350–14361
16. Jaditzke, B. *et al.* (2020) Photobodies: light-activatable single-domain antibody fragments. *Angew. Chem. Int Ed* 59, 1506–1510
17. Ros, E. *et al.* (2020) Synthesis and application of 3-bromo-1,2,4,5-tetrazine for protein labeling to trigger click-to-release bioorthogonal reactions. *Bioconjug. Chem* 31, 933–938
18. Fok, J.A. and Mayer, C. (2020) Genetic code expansion strategies for vaccine development. *ChemBioChem* Published online June 30, 2020. https://doi.org/10.1002/cbic.202000343
19. Huang, Y. and Liu, T. (2018) Therapeutic applications of genetic code expansion. *Synth. Syst. Biotechnol.* 3, 150–158
20. Neumann, H. *et al.* (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* 464, 441–444
21. Hohsaka, T. *et al.* (2001) Five-base codons for incorporation of nonnatural amino acids into proteins. *Nucleic Acids Res.* 29, 3646–3651
22. Zhang, Y. *et al.* (2017) A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* 551, 644–647
23. Hoshika, S. *et al.* (2019) Hachimoji DNA and RNA: a genetic system with eight building blocks. *Science* 363, 884–887
24. Rodnina, M.V. (2018) The ribosome in action: tuning of translational efficiency and protein folding. *Protein Sci.* 25, 1390–1406
25. Novoa, E.M. and Ribas de Pouplana, L. (2012) Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet.* 28, 574–581
26. Dumas, A. *et al.* (2015) Designing logical codon reassignment – expanding the chemistry in biology. *Chem. Sci.* 6, 50–69
27. Caetano-Anollés, G. *et al.* (2009) The origin, evolution and structure of the protein world. *Biochem. J.* 417, 621–637
28. Woolfson, D.N. *et al.* (2015) De novo protein design: how do we expand into the universe of possible protein structures? *Curr. Opin. Struct. Biol.* 33, 16–26
29. Alva, V. *et al.* (2015) A vocabulary of ancient peptides at the origin of folded proteins. *Elife* 4, e09410

30. Ribas de Pouplana, L. *et al.* (2017) What froze the genetic code? *Life* 7, 14
31. Lassak, J. *et al.* (2016) Stalino more at polyproline stretches with the translation elongation factors EF-P and IF-5A. *Mol. Microbiol.* 99, 219–235
32. Greber, B.J. *et al.* (2014) Architecture of the large subunit of the mammalian mitochondrial ribosome. *Nature* 505, 515–519
33. Wan, W. *et al.* (2014) Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta Proteins Proteomics* 1844, 1059–1070
34. Gaston, M.A. *et al.* (2011) Functional context, biosynthesis, and genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.* 14, 342–349
35. Zhang, Y. *et al.* (2020) Role of selenoproteins in redox regulation of signaling and the antioxidant system: a review. *Antioxidants* 9, 383
36. Sentio, V.H.B. *et al.* (2018) The unique tRNASec and its role in selenocysteine biosynthesis. *Amino Acids* 50, 1145–1167
37. Ribas de Pouplana, L. *et al.* (2014) Protein mistranslation: friend or foe? *Trends Biochem. Sci.* 39, 355–362
38. Simões, J. *et al.* (2016) The fungus *Candida albicans* tolerates ambiguity at multiple codons. *Front. Microbiol.* 7, 401
39. Mukai, T. *et al.* (2015) Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. *Nucleic Acids Res.* 43, 8111–8122
40. Bohlke, N. and Budisa, N. (2014) Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: rare isoleucine codon AUA as a target for genetic code expansion. *FEMS Microbiol. Lett.* 351, 133–144
41. Mahdavi, A. *et al.* (2016) Engineered aminoacyl-tRNA synthetase for cell-selective analysis of mammalian protein synthesis. *J. Am. Chem. Soc.* 138, 4278–4281
42. Geslain, R. *et al.* (2009) Chimeric tRNAs as tools to induce proteome damage and identify components of stress responses. *Nucleic Acids Res.* 38, e80
43. Johnson, D.B.F. *et al.* (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* 7, 779–786
44. Wannier, T.M. *et al.* (2018) Adaptive evolution of genomically recoded *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 115, 3090–3095
45. Lajoie, M.J. *et al.* (2013) Genomically recoded organisms expand biological functions. *Science* 342, 357–360
46. Lind, C. *et al.* (2017) Origin of the omnipotence of eukaryotic release factor 1. *Nat. Commun.* 8, 1425
47. Schmied, W.H. *et al.* (2014) Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. *J. Am. Chem. Soc.* 136, 15577–15583
48. Melo Czekster, C. *et al.* (2016) *In vivo* biosynthesis of a β-amino acid-containing protein. *J. Am. Chem. Soc.* 138, 5194–5197
49. Porter, E.A. *et al.* (2002) Mimicry of host-defense peptides by unnatural oligomers: antimicrobial β-peptides. *J. Am. Chem. Soc.* 124, 7324–7330
50. Rackham, O. and Chin, J.W. (2005) A network of orthogonal ribosome·mRNA pairs. *Nat. Chem. Biol.* 1, 159–166
51. Wang, K. *et al.* (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.* 25, 770–777
52. Schmied, W.H. *et al.* (2018) Controlling orthogonal ribosome subunit interactions enables evolution of new function. *Nature* 564, 444–448
53. Chevalier, A. and Garel, J-P. (2005) Differential synthesis rates of tRNA species in the silk gland of *Bombyx mori* are required to promote tRNA adaptation to silk messages. *Eur. J. Biochem.* 124, 477–482
54. Li, J. *et al.* (2015) Comparative proteomic analysis of the silkworm middle silk gland reveals the importance of ribosome biogenesis in silk protein production. *J. Proteome* 126, 109–120
55. Grosjean, H. *et al.* (2010) Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett.* 584, 252–264
56. Torres, A.G. *et al.* (2014) A-to-I editing on tRNAs: biochemical, biological and evolutionary implications. *FEBS Lett.* 588, 4279–4286
57. Novoa, E.M. *et al.* (2012) A role for tRNA modifications in genome structure and codon usage. *Cell* 149, 202–213
58. Rafels-Ybern, À. *et al.* (2019) The expansion of inosine at the wobble position of tRNAs, and its role in the evolution of proteomes. *Mol. Biol. Evol.* 36, 650–662
59. Pinzón Martín, S. *et al.* (2019) Mucins and pathogenic mucin-like molecules are immunomodulators during infection and targets for diagnostics and vaccines. *Front. Chem.* 7, 710
60. Rafels-Ybern, À. *et al.* (2015) Distribution of ADAT-dependent codons in the human transcriptome. *Int. J. Mol. Sci.* 16, 17303–17314
61. Rafels-Ybern, À. *et al.* (2018) Codon adaptation to tRNAs with Inosine modification at position 34 is widespread among Eukaryotes and present in two Bacterial phyla. *RNA Biol.* 15, 500–507
62. Nakano, S. *et al.* (2016) NSUN3 methylase initiates 5-formylcytidine biogenesis in human mitochondrial tRNAMet. *Nat. Chem. Biol.* 12, 546–551
63. Rudinger, J. *et al.* (1994) Histidylation by yeast HisRS of tRNA or tRNA-like structure relies on residues -1 and 73 but is dependent on the RNA context. *Nucleic Acids Res.* 22, 5031–5037
64. Pütz, J. *et al.* (1994) A single methyl group prevents the mischarging of a tRNA. *Nat. Struct. Mol. Biol.* 1, 580–582
65. Rogerson, D.T. *et al.* (2015) Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat. Chem. Biol.* 11, 496–503
66. Zhang, M.S. *et al.* (2017) Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nat. Methods* 14, 729–736
67. Fan, C. *et al.* (2015) Rationally evolving tRNA^{Pyl} for efficient incorporation of noncanonical amino acids. *Nucleic Acids Res.* 43, e156
68. Guo, J. *et al.* (2009) Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew. Chem. Int. Ed.* 48, 9148–9151
69. Thyer, R. *et al.* (2015) Evolving tRNA^{Sec} for efficient canonical incorporation of selenocysteine. *J. Am. Chem. Soc.* 137, 46–49
70. Serfling, R. *et al.* (2018) Designer tRNAs for efficient incorporation of non-canonical amino acids by the pyrrolysine system in mammalian cells. *Nucleic Acids Res.* 46, 1–10
71. Chatterjee, A. *et al.* (2013) A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. *Biochemistry* 52, 1828–1837
72. Aloush, N. *et al.* (2018) Live cell imaging of bioorthogonally labelled proteins generated with a single pyrrolysine tRNA gene. *Sci. Rep.* 8, 14527
73. Ryu, Y. and Schultz, P.G. (2006) Efficient incorporation of unnatural amino acids into proteins in *Escherichia coli*. *Nat. Methods* 3, 263–265
74. Chatterjee, A. *et al.* (2013) Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 110, 11803–11808
75. Mukai, T. *et al.* (2008) Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* 371, 818–822
76. Baldridge, K.C. *et al.* (2018) Directed evolution of heterologous tRNAs leads to reduced dependence on post-transcriptional modifications. *ACS Synth. Biol.* 7, 1315–1327
77. Crnković, A. *et al.* (2018) Effects of heterologous tRNA modifications on the production of proteins containing noncanonical amino acids. *Bioengineering* 5, 11
78. Biddle, W. *et al.* (2016) Modification of orthogonal tRNAs: unexpected consequences for sense codon reassignment. *Nucleic Acids Res.* 44, 10042–10050
79. Krogh, N. *et al.* (2016) Profiling of 2′-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res.* 44, 7884–7895
80. Shi, Z. *et al.* (2017) Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Mol. Cell* 67, 71–83.e7
81. Shi, Z. and Barna, M. (2015) Translating the genome in time and space: specialized ribosomes, RNA regulons, and RNA-binding proteins. *Annu. Rev. Cell Dev. Biol.* 31, 31–54

82. Reinkemeier, C.D. *et al.* (2019) Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. *Science* 363, eaaw2644
83. Suzuki, T. *et al.* (2017) Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* 13, 1261–1266
84. Bryson, D.I. *et al.* (2017) Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* 13, 1253–1260
85. Kleiman, L. (2002) tRNA(Lys3): the primer tRNA for reverse transcription in HIV-1. *IUBMB Life* 53, 107–114
86. Kreuzer, K.D. and Henkin, T.M. (2018) The T-box riboswitch: tRNA as an effector to modulate gene regulation. *Microbiol. Spectr.* 6, RWR-0028-2018
87. Fasioura, T. and Suga, H. (2014) Reprogramming the genetic code in vitro. *Trends Biochem. Sci.* 39, 400–408
88. Dunkelmann, D.L. *et al.* (2020) Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. *Nat. Chem.* 12, 535–544
89. Brown, W. *et al.* (2018) Genetic code expansion in animals. *ACS Chem. Biol.* 13, 2375–2386
90. Italia, J.S. *et al.* (2018) Resurrecting the bacterial tyrosyl-tRNA synthetase/tRNA pair for expanding the genetic code of both *E. coli* and eukaryotes. *Cell Chem. Biol.* 25, 1304–1312.e5
91. Steglitz, J.T. *et al.* (2018) A robust and quantitative reporter system to evaluate noncanonical amino acid incorporation in yeast. *ACS Synth. Biol.* 7, 2256–2269
92. Parrish, A.R. *et al.* (2012) Expanding the genetic code of *Caenorhabditis elegans* using bacterial aminoacyl-tRNA synthetase/tRNA pairs. *ACS Chem. Biol.* 7, 1292–1302