UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

UMR 7199- Laboratoire des Systèmes Chimiques Fonctionnels

THÈSE

présentée par

Oleksandr KONIEV

soutenue le **20 mars 2014**

pour obtenir le grade de

Docteur de l'Université de Strasbourg

Discipline / Spécialité : Chimie organique

Development of new bioselective ligation reactions.

THÈSE dirigée par :

WAGNER Alain	Docteur, Université de Strasbourg
BAATI Rachid	Docteur, Université de Strasbourg

RAPPORTEURS :

KOESTER Hubert	Professeur, Université d'Hambourg
MELNYK Oleg	Docteur, Université de Lille

AUTRES MEMBRES DU JURY :

MELY Ives	Professeur, Université de Strasbourg
BECK Alain	Docteur, Laboratoires Pierre Fabre

Hofstadter's Law:

It always takes longer than you expect, even when you take into account Hofstadter's Law.

D. Hofstadter: *Gödel, Escher, Bach: An Eternal Golden Brain.*

ACKNOWLEDGMENTS

This work has been carried out in the Laboratory of Functional Chemo-Systems (LFCS) of the Faculty of Pharmacy at the University of Strasbourg under the supervision of Dr. Alain Wagner and Dr. Rachid Baati and has received funding from the European Community's Seventh Framework Programme (FP7/2009) under grant agreement n° 238434.

First of all I would like to express my special appreciation and thanks to my supervisors Dr. Alain Wagner and Dr. Rachid Baati for their encouraging my research and ideas they have given me. I would like to thank you for your confidence shown to me throughout this project that, I believe, allowed me to grow as a research scientist.

I would also like to express my gratitude to the committee members for their willingness to participate in the examining of my PhD thesis and to come to Strasbourg for the defense: Professor Hubert Köster from the University of Hamburg and Caprotec Bioanalytics, Doctor Oleg Melnyk from the University of Lille, Doctor Alain Beck from Pierre Fabre Laboratories, and Professor Yves Mély from the University of Strasbourg.

I would especially like to thank the members of the Analytical Services of the Faculty of Pharmacy, who were always there to shed the light on sometimes seeming improbable chemical transformations I encountered. Cyril Antheaume for his everyday assistance with NMR in general and for providing first insights into a fabulous world of advanced NMR techniques, which have significantly simplified my work. Pascale Buisine and Patrick Wehrung for their persistent and incessant search for the right ion mass and the corresponding fragments.

Of course, I would also like to thank Michel Mosser, our scaling-up Master, without whom we would never be able to squander fluorescent dyes and strain-promoted alkynes in such a reckless way.

Being an organic chemist by training, I would like to express my gratitude to those who helped me, when I got lost in the woods of biology, proteomics, and combinatorial chemistry. Marc Nothisen for his precious advice during the preparation and punctilious carrying of the experiments in cells. Hélène Muller-Steffner and Isabelle Kuhn for explaining me the pitfalls of IC50 measurements. The LSMBO group – Christine Schaeffer-Reiss, Jean-Marc Strub, Sarah Cianferani, and Alan van Dorsselaer – for making possible the establishment of the TMPP-based screening system. The whole Reymond Group from the University of Bern – Jean-Louis Reymond, Tamis Darbre, and a full bunch of crazy PhD-students (Michaela, Marco, Justus, Rachel, Lars, Julian, Emilyne, and Falco) – for your helping me to get into the world of combinatorial chemistry and solid-phase peptide synthesis. This great month I have spent in your laboratory will surely long remain in my memory.

In the frame of the Marie-Curie training network, I had the opportunity to meet and work with a great number of people from all over Europe. I want to thank you all – brave BioChemLiggers – for the great time we had during our regular meetings. These people have been (but surely are not limited to) Tilmann Heinisch, Anamitra Chatterjee, Todd Hyster and David Gnandt from Thomas Ward's group in Basel, Sandra Enriquez Sansaloni, Veronica López Carillo and Steve Saulnier from the groups of Pier Girgio Cozzi and Francesco Paolucci in Bologna, João André Banha Olivera and Michael Sefkow from Caprotec in Berlin, Elena Benedetto and Dong Li

from Veronique Gouverneur's group in Oxford, Michaela Stach and Marco Bartoloni from the above-mentioned Jean-Louis Reymond's group in Bern, Sisir Nandi and Alessandro Monesi from Marjana Novic's group in Ljubjana, and Valentina Bevilacqua and especially my bosom friend Sergii Kolodych from Frédéric Taran's group in Paris. It is in close collaboration with the group of Frédéric Taran that we have established three novel promising technologies for bioconjugation which today constitute the core of the start-up company under creation.

My sincere thanks also goes to the two project managers: Dr. Louise Chisholm, who unfortunately left us, for her home Britain, and Dr. Jérémie Hagenmuller, who brought BioChemLig to its successful completion and is now experiencing all the pleasant moments and vicissitudes (of which, I hope there are less) of his world tour.

Furthermore, I would like to thank my fellow labmates. Starting from the old crew, Ghyslain, Aurelia, Manue, and Cédric, whom I have known much less than I would like to, because of their leaving LFCS short after my arrival to the lab. Laure and Samira for our lively discussions near the Flash Chromatograph. Antoine for many valuable pieces of advice and the working place I inherited after his departure to Canada. Coraline for a considerable improvement of my historical knowledge of the French monarchs and customs. Old fellows Mathieu and Geoffray for teaching me all the seriousness and importance of keeping lab stuff well organised, combined with a simultaneous dragging me off to the pubs, and organising terrific (in a good sense) house-warmings. Of course, Hélène and Marion to whom I'm especially grateful for their tireless increasing of my motivation to do the laboratory routine with their magic "faut l'faire, Sasha", and coping with my inability to deal with all the administrative things at École Doctorale. Mathias, the most organised german in the lab, without whom it would be plunged into complete chaos. He has proven himself a great colleague and a devoted like-minded stalwart of the pastafarianism. Julien for his almost perfect speaking Russian. Next, our post-docs, Marie-Charlotte for a great catchy-phrase for my publication, Guiseppe for his motivating me going to the gym, and Tam for her unforgettable "tu ne comprends pas".

In addition I would like to thank all those, who have only shortly been part of the LFCS. Thanks Jessie, Patricia, David, Gosha; our trainees Charlotte, Aïssata, Abiba, Elliot, Ricardo, Emra, Valentin, Youssef; and, of course, all the Ukrainians, who seem to become more numerous every single moment. Among them, my faithful pals Sergii (see above), Artim (see below), Ihor, Jenya, Dasha, Mishka, 2 x Julia, Marianna, Vasyl, Vika, Maria, and many more I have certainly forgotten to mention. Especially Artim, who together with Marion took the brunt of reading the first drafts of this manuscript, sometimes completely incompresensible; and Ihor Fedin from the University of Chicago who was always a great example of diligence for me.

Finally, our new arrivals Zolo, Sylvain, and Elsa, as well as Bourguignons: Emilie, Maude, François; all the SECOistes, especially our dream-team: Nico, Amandine, Rémy, and … (drum-roll) Anirban; our UMR7199 all in all, and Jean-Serge particularly – I thank you all, and if I have undeservedly forgotten anyone, feel free to come and complain, you are always welcome.

And last but not the least, I would like to thank my family for supporting me spiritually throughout my life. Thank you.

ABBREVIATIONS

3-BPA	3-Bromopropanamine
AA	Amino acid
AAA	Amino acid analysis
Ab	Antibody
Ac	Acetyl
ACN	Acetonitrile
ADC	Antibody-drug conjugate
APN	3-Arylpropiolonitriles
AU	Arbitrary unit
BHQ	Black hole quencher®
Bn	Benzyl
Boc	*tert*-Butyloxycarbonyl
BODIPY	Boron-dipyrromethene
BPTI	Bovine pancreatic trypsin inhibitor
BSA	Bovine serum albumin
BSOCOES	Bis[2-(succinimidooxycarbonyloxy)ethyl]sulfone
Bu	Butyl
Bz	Benzoyl
CAN	Cerium(IV) ammonium nitrate
CBT	2-Cyanobenzothiazole
CDMS	Chlorodimethylsulfide
Dabsyl	4-(Dimethylamino)azobenzene-4′-sulfonyl
DBAA	α,α' Dibromoadipyl(bis)amide
DCC	Dicyclohexylcarbodiimide
DCM	Dichloromethane
DIC	Diisopropylcarbodiimide
DIPEA	N,N-Diisopropylethylamine
DMA	Dimethyl adipimidate
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMP	Dimethyl pimelimidate
DMPA	2,2-dimethoxy-2-phenylacetophenone
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNPS-Cl	2,4 Dinitrophenylsulfenyl chloride
DNSC	Dansyl chloride
DST	Disuccinimidyl tartarate
DTBP	Dimethyl 3,3´-dithiobispropionimidate
DTNB	5,5'-Dithiobis-(2-nitrobenzoic acid), Elmann's reagent.
DTT	Dithiotreitol
EDC	1-Ethyl-3-(3-dimethyl-aminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced green fluorescent protein
EGS	Ethylene glycol bis[succinimidylsuccinate]
EIC	Extracted ion chromatogram
EPL	Expressed protein ligation
EPR	Electron paramagnetic resonance
ESI	Electrospray ionisation
Et	Ethyl
FAD	Flavin adenine dinucleotide
FDA	US Food and Drug Administration
FG	Functional group
FITC	Fluorescein isothiocyanate
FlAsH	4',5'-Bis(1,3,2-dithioarsolan-2-yl)fluorescein
Fmoc	Fluorenylmethyloxycarbonyl
FRET	Förster resonance energy transfer
GABA	*gamma*-Aminobutyric acid
GCMS	Gas chromatography–mass spectrometry
GFP	Green fluorescent protein
GSH	Glutathione reduced
GuHCl	Guanidinium chloride
HATU	1-[Bis(dimethylamino)methylene]-*1H*-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluoro-phosphate
HAZA	2-(2′-Alkoxy-4′-hydroxyphenylazo)benzoic acid
HBTU	*O*-Benzotriazole-*N,N,N′,N′*-tetramethyl-uronium-hexafluoro-phosphate
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HL	High loading
HMPT	Hexamethylphosphoramide
HPLC	High-performance liquid chromatography
HRMS	High-resolution mass spectrometry
HRP	Horseradish peroxidase
HSA	Human serum albumin
IAC	Iodoacetyl chloride
IC50	Half maximal inhibitory concentration
IFN	Interferon

IgG	Immunoglobulin G
IMAC	Immobilised-metal affinity chromatography
IPA	Isopropyl alcohol
KCL	Kinetically controlled ligation
LC	Liquid chromatography
mAb	Monoclonal antibody
MALDI	Matrix-assisted laser desorption/ionisation
MDA	Malondialdehyde
MMPP	Monoperoxyphthalate
MNBS	Methyl 4-nitrobenzenesulfonate
MPA	3-Mercaptopropionic acid
MPAA	4-(Carboxymethyl)thiophenol
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
Ms	Mesyl
MSH	*o*-Mesitylenesulfonylhydroxylamine
MTS	Methylthiosulfonate
NADP	Nicotinamide adenine dinucleotide phosphate
nanoLC	Nanoscale liquid chromatography
NBD-F	4-fluoro-7-nitro-2,1,3-benzoxadiazole
NBS	*N*-Bromosuccinimide
NCL	Native chemical ligation
NCS	*N*-Chlorosuccinimide
NHS	*N*-hydroxysuccinimide
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
NTA	Nitrilotriacetic acid
OEGMA	Oligo(ethylene glycol) methyl ether methacrylate
OND	Oxanorbornadienedicarboxylate
PB	Phosphate buffered
PBS	Phosphate buffered saline
PEG	Polyethyleneglycol
PLP	Pyridoxal-5-phosphate
Pr	Propyl
PTAD	4-Phenyl-*3H*-1,2,4-triazoline-3,5(4H)-dione
Pybop	Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
PyS	Pyridine-2(*1H*)-thione
QD	Quantum dot
RG	Reactive group
RNA	Ribonucleic acid
RNase	Ribonuclease
RT	Retention time
SAM	Bis(2-sulfanylethyl)amide
sCT	Salmon calcitonin
SEA	*N*-sulfanylethylamilide
SPPS	Solid-phase peptide synthesis
Su	Succinimide
TAMRA	Tetramethyl-rhodamine
TBAB	Tetra-*n*-butylammoniumbromide
TBAF	Tetra-*n*-butylammoniumbromide
TCEP	tris(2-carboxyethyl)phosphine
TEA	Triethylamine
TEC	Thiol-ene coupling
Tf	Triflyl
TFA	Trifluoroacetic acid
TFE	2,2,2-Trifluoroethanol
THF	Tetrahydrofuran
TIPS	Triisopropylsilane
TLC	Thin layer chromatography
TMPP	Tris(2,4,6-trimethoxyphenyl)phosphonium
TMS	Trimethylsilyl
TMTH	Tetramethylthiacycloheptyne
TMTI	Tetramethylthiacycloheptynium
TNB	5-thio-2-nitrobenzoic acid
TNM	Tetranitromethane
TOF	Time of flight
TRIS	Tris(hydroxymethyl)aminomethane
Trx	Thioredoxine
TrxR	Thioredoxine reductase
Ts	Tosyl
TYC	Thiol-yne coupling
UV	Ultraviolet
VAZO44	2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride
VS	Vinyl sulfone

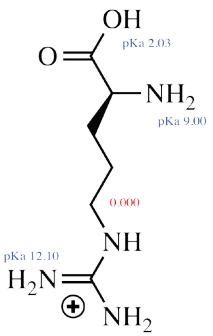
TWENTY-ONE AMINO ACIDS

A. Amino Acids with Electrically Charged Side Chains

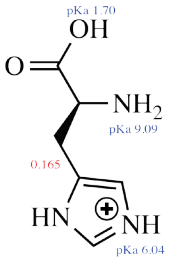
(charge at physiological pH 7.4)

Positive

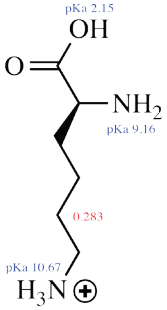
Arginine (Arg) R



Histidine (His) H

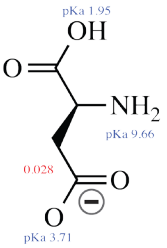


Lysine (Lys) K

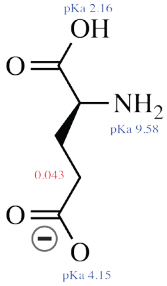


Negative

Aspartic Acid (Asp) D

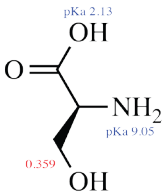


Glutamic Acid (Glu) E

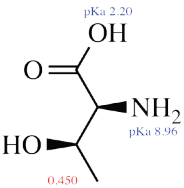


B. Amino Acids with Polar Uncharged Side Chains

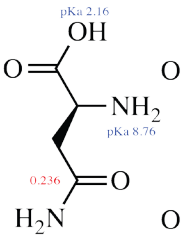
Serine (Ser) S



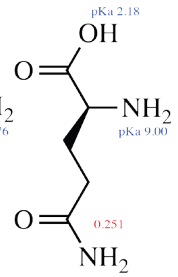
Threonine (Thr) T



Asparagine (Asn) N

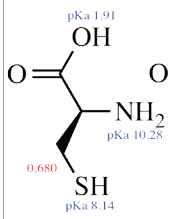


Glutamine (Gln) Q

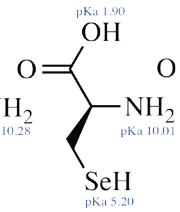


C. Special Cases

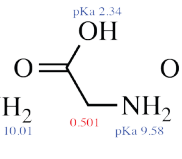
Cysteine (Cys) C



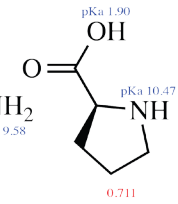
Selenocysteine (Sec) U



Glycine (Gly) G

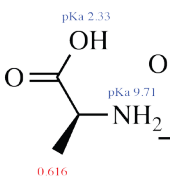


Proline (Pro) P

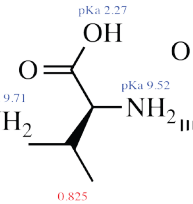


D. Amino Acids with Hydrophobic Side Chain

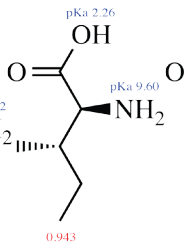
Alanine (Ala) A



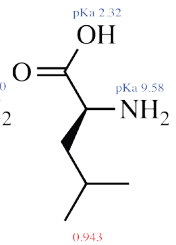
Valine (Val) V



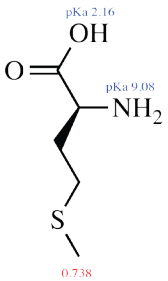
Isoleucine (Ile) I



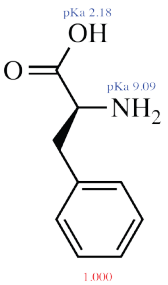
Leucine (Leu) L



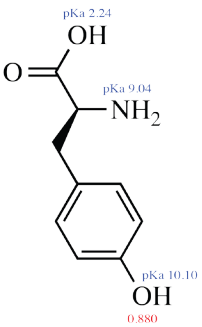
Methionine (Met) M



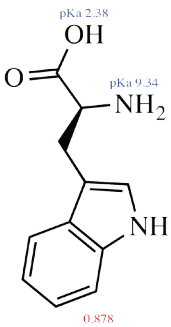
Phenylalanine (Phe) F



Tyrosine (Tyr) Y



Tryptophane (Trp) W



pK_a Data: Lundblad, R. L. Application of Solution Protein Chemistry to Biotechnology. (CRC PressI Llc, 2009).

Hydrophobicity Data: Black S.D., Mould D.R. Anal. Biochem. 193:72-82(1991).

Adapted from Dan Cojocari's "21 Amino Acids" poster **(University of Toronto)**

TABLE OF CONTENTS

FOREWORD

Chemistry is based on evidence and experiment. Its beginnings have arisen from the very emergence of a human being. People always somehow had to deal with chemicals, whether it was first fire experiment, tanning hides or cooking. Every living organism is, after all, a complex chemical assembly. Ceaseless accumulation of knowledge about composition, properties and behavior of matter has finally resulted in what we call today chemical science. This tremendous amount of information has transformed chemistry into a complex, but elegant and omnipresent scientific discipline with numerous independent subdisciplines.

Organic chemistry deals with organic compounds and organic materials, i.e., various forms of matter that contain carbon atoms. The number of described reactions in organic chemistry is striking: only two databases of transformations in organic chemistry - SciFinder® and Reaxys® - contain 53.5 and 31 million of reactions respectively. Even though this amount of information seems exhaustive for any imaginable transformation, none the less, sometimes chemists run short of them. Especially this becomes the case when the reaction is required to take place under physiological conditions, *i.e.* in water within a relatively narrow reaction range of temperature, concentration, and pH. And it's when the discovery of new chemical reactions becomes crucial.

This work was devoted to a discovery and further development of new chemical transformations for a specific field of modern chemistry – bioconjugation. Generally speaking, it consists in the coupling two biomolecules together in covalent manner to form a novel complex having the combined properties of its individual components. Thus, natural or synthetic compounds with their individual activities can be chemically combined to create unique substances possessing carefully engineered characteristics. For instance, an antibody able to target exclusively cancer cells can be crosslinked with a cytotoxic payload to give a highly potent cancer treatment, or a traceable molecule can be attached to a biomolecule of interest making the latter detectable.

The necessity to preserve the intact biological activity of a biomolecule results in numerous severe requirements for a reaction to serve as appropriate bioconjugation methodology. The process should be selective towards given target, give stable conjugation products in high yields at reasonable time scale. It is to an overview of already known and to a discovery of new approaches that we turn in the chapters to follow.

A. INTRODUCTION

In this introduction the current developments in various techniques of bioconjugation on proteins and peptides were highlighted. The utility of each conjugation methodology in the area of biologics and protein science has been surveyed with emphasis on most relevant among reported transformations. Merits and demerits of these methods with reference to selectivity and practical use have been discussed.

Bioconjugation is a set of techniques allowing instalment of moieties into the native structure of biomolecules that endow them with desirable properties. This novel complex having the combined properties of its individual components can serve, for instance, as more stable and efficient therapeutics,[1-4] assemblies for studying proteins in their biological context,[5-11] new protein-based materials,[12-17] microarrays,[18,19] biologics,[20-24] tools for immobilisation,[25] elucidation of the structure[26,27] of proteins.

Large number of reactions exist to modify proteins,[28] however, site-specific conjugation still presents a considerable challenge, considering that the reaction must proceed in aqueous solution within a relatively narrow pH and temperature range in order to preserve protein's tertiary structure. For the purposes of this overview, the focus will remain on selective conjugation of the innate protein residues. Recent developments in the field of protein modifications will be thus highlighted with emphasis on the transformations most relevant to bioconjugation.

Despite the uniquely diverse functions and structures of proteins, only about 20 different amino acids are involved into their composition. Even a smaller number of amino acid functional groups comprise appropriate targets for practical bioconjugation methods. In fact, only one third of all amino acid residues represent chemical targets account for the vast majority of crosslinking.

The reactivity of a specific residue in the protein structure is mainly determined by three parameters: its accessibility, acidity/basicity, and electrophilicity/nucleophilicity. While accessibility of the residue is generally structure-specific, two other parameters are less structure-independent and thus can be used to rationalise the tremendous amount of reactions used for protein conjugation. It is through the prism of these two parameters that bioconjugation techniques are regarded in this manuscript being ranged in a descending order of frequency they are reported in literature.

1 IN-CHAIN CONJUGATION

1.1 LYSINE

The use of chemical groups that react with primary amines is one of the oldest and most versatile techniques for protein conjugation. Virtually all proteins contain primary amino groups in their structure. They can be divided into two groups: α-amino group situated on the *N*-terminus of most polypeptide chains and ε-amino groups of lysine residue (Lys, K). Because these amino groups possess pK_a values of about 8 and 10 (for α- and ε-amines respectively), in a vast majority of cases they are protonated at physiologic pH and, therefore, occur predominantly on the solvent-exposed outside surfaces of protein tertiary structures. As a result, they become easily accessible to conjugation reagents introduced into the aqueous media.

Unprotonated primary amines are among the most nucleophilic from the available functional groups present in a typical protein. However, the protonation drastically decreases their reactivity. As a consequence, despite generally higher intrinsic nucleophilicity of Lys ε-amino groups, it requires higher pH values to be uncovered by unprotonation, which allows distinguishing α- and ε-amino groups by adjusting the pH. That is to say, at the higher pH level, when both types of primary amines are deprotonated, Lys side chain amino groups are generally more reactive towards electrophiles, while at the lower pH it is the opposite because of their prior protonation (Fig. 1). At the acidic pH all amines are protonated and possess no significant nucleophilicity compared to other side chains presented in proteins. In particular, free (non-disulfide-bonded) Cys residues are much stronger nucleophiles and, if accessible, will readily be modified by most amine-reactive reagents.

It is also to be mentioned that like any other parameters, nucleophilicity and basicity, as well as solvent exposure and accessibility of a particular amino group, are influenced by the microenvironment and can vary substantially, regarding the substrate. For instance, Westheimer and Schmidt have found the actual pK_a of the amino group situated in the active site of acetoacetate decarboxylase to be 5.9, which is 4 pK_a units less than that of an "ordinary" ε-amino group of lysine.[29]

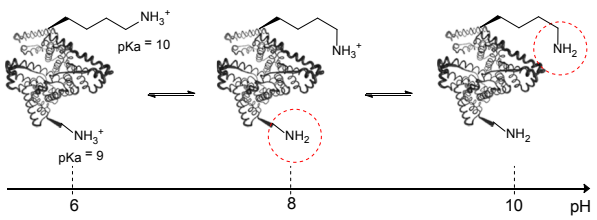


Fig. 1. Unprotonation of different types of amino groups present in protein structure (more nucleophilic amine is encircled in red). Lysine ε-amino groups are more nucleophilic, but also more difficult to unprotonate. Generally, a pH of 8.5-9.5 is optimal for modifying lysine residues, while near neutral pH favours selective modification of *N*-termini.

Depending on reaction conditions, selective modification of either *N*-termini (see Section A 2, page 29) or Lys ε-amino groups can be achieved by using various chemical reagents. They generally belong to one of the following classes (in order of relevancy): activated esters (fluorophenyl esters, NHS

(*N*-hydroxysuccinimides),[30] sulfo-NHS, acyl azides), isothiocyanates, isocyanates,[31] aldehydes, anhydrides, sulfonyl chlorides, carbonates, fluorobenzenes, epoxides and imidoesters. Among this vast variety of reactive functions, NHS esters (and their more soluble sulfo-NHS analogues) and imidoesters represent the most popular amine-specific functional groups that are incorporated into commercially available reagents for protein conjugation and labelling.[28]

Despite their name, amine-reactive reagents are not always entirely selective for amines. Firstly, as it was already mentioned before, they will react with any other stronger nucleophile, if the latter is present and accessible on a protein surface. Particularly, it concerns cysteine, tyrosine, serine and threonine side chains. Secondly, a depletion of these highly activated reagents by hydrolysis is inevitable in aqueous solution. The rate of both side-reactions depends on the particular substrate, conjugation partner, pH, temperature, and buffer composition. Evidently, buffers that contain free amines, such as TRIS (tris(hydroxymethyl)aminomethane), must be avoided when using any amine-reactive probes, since the rate of the reaction with buffer would greatly exceed that with protein amino groups.

1.1.1 Isocyanates and Isothiocyanates

Amines undergo a reaction with isocyanates to readily form stable ureas. However, because of their susceptibility to deterioration during storage,[32] isocyanates are much more difficult to manipulate with and thus are not as well commercially accessible as corresponding isothiocyanates. They can though be easily prepared prior to use from more stable acyl azides by Curtius rearrangement. Using this approach, for instance, Palumbo and colleagues have elaborated a synthesis of a heterobifunctional amine-thiol crosslinker containing an isocyanate group on one end and a thiol reactive maleimide group on the other end (Fig. 2).[33]

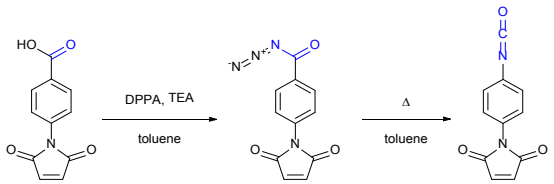


Fig. 2. Synthesis of p-maleimidophenyl isocyanate crosslinker *via* Curtius rearrangement proposed by Palumbo and associates.[33]

Several early studies were devoted to the elaboration of isocyanates conjugation methodology,[34,35] but proven to be especially laborious and complicated. Therefore they are of deferred interest today, being completely displaced by isothiocyanate- and NHS-mediated approaches.

Isothiocyanate-based selective amino group modification was first reported in 1937 by Todrick and Walker,[36] who found that the reaction of allyl isothiocyanate with cysteine in alkaline medium results selectively in obtaining of the thiourea – the product of the amine addition to isothiocyanate. In 1950, exploiting the selectivity of amino-terminal labelling of the peptide with phenylisothiocyanate, Edman has developed a method for peptide sequencing that has changed cardinally the protein science and is known today as Edman degradation.[37] Only 30 years later, Podhradsky *et al*. have examined the reaction of isothiocyanates on complex substrates and demonstrated that the addition of the thiol and phenolate functions of cysteine and tyrosine residues is always prevalent, and that only at pH > 5 amino groups start to manifest themselves in the reaction.[38] While thiol and alcohol additions result in reversible reactions to give dithiocarbamates and *O*-thiocarbamates respectively, amines add themselves irreversibly, thus shifting the reaction equilibrium towards thioureas (Fig. 3). One should however keep in mind that, despite the reversibility of the addition of thiols and alcohols to isocyanates, they can enhance the kinetics of their hydrolysis to unreactive amines or ureas and therefore significantly decrease the yield of the conjugation. Moderately reactive but quite stable in water and most solvents, isothiocyanates represent thus an appropriate alternative to the unstable isocyanates. As a consequence, they are much more popular in bioconjugation.

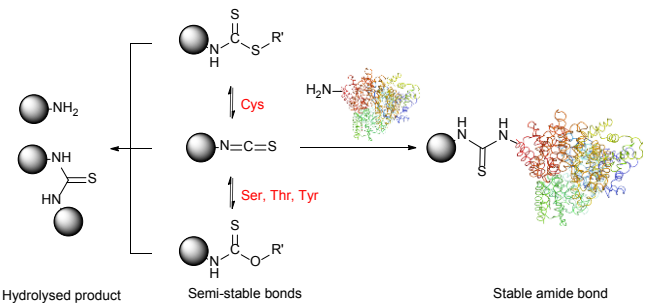


Fig. 3. Reaction of isothiocyanates with nucleophilic amino acid residues present in proteins. Only the reaction of lysine and *N*-terminal residues (considerable at pH > 5) results irreversibly in obtaining the thiourea. Although the reactions of thiol and alcohol groups with isothiocyanates are reversible, they can largely accelerate the rate of isothiocyanate hydrolysis.

Ever since the introduction of fluorescent isothiocyanate dyes as more stable analogues of corresponding isocyanates for fluorescent labelling of antibodies by Riggs *et al*.[39] in 1958, they have found widespread in research laboratories and proved to be an effective means for tagging proteins at specific sites.[40] Fluorescein isothiocyanate (FITC) is arguably one of the most commonly used fluorescent derivatisation reagents for proteins. For instance, it was reported by Tuls *et al*.,[41] that cytochrome P-450 can be selectively labelled by FITC with 75% yield of single-labelled LYS-338 conjugate in TRIS (particularly inappropriate buffer for amine-reactive reagents though) at pH 8.0 and 0 °C. Burtnick[42] has described selective labelling of one out of 34 lysine residues of actin (Fig. 4) in borate buffer with 35-fold excess of the reagent at pH 8.5. Following reports of Miki and collaborators[43,44] further confirmed the selectivity of this labelling, yet without any explanation of such specificity. Bellelli *et al*.[45] were able to covalently label ricin (pH 8.1, 6°C

for 4 h). In fact, the targets of the isothiocyanate-mediated labelling of proteins elaborated over last 60 years are even difficult to enumerate. It was proven to be effective in such diverse applications as tagging of antibodies (usually in carbonate-bicarbonate buffer, pH 9),[46-52] bleaching-based measurement of membrane protein diffusion of FITC-labelled cells (pH 9.5, 24 ˚C),[53] surface topography of the *Escherichia coli* ribosomal subunit,[54] α-actinin distribution in living and fixed fibroblasts,[55] characterisation of a proton pump on lysosomes,[56] and hematopoietic stem cells.[57]

The most stunning examples include ^{125}I labelling by means of isothiocyanates, elaborated by Shapiro and colleagues[58] for regional differentiation of the sperm surface (TRIS, pH 7.7, 12 ˚C for 30 min), and the application of similar methodology by Schirrmacher *et al*.[59] for ^{18}F radioactive labelling of RSA, apotransferrin and bovine IgG (pH 9.0, room temperature for 10-20 min). Conjugation of antibodies with chelating agents for further radiometal labelling of antibodies have been described by several groups[60-62] and are based on the use of phenylisothiocyanate-containing probes. Brechbiel *et al*.[63] went even further by combining the chelating functionality with biotin fragment in a scaffold of trifunctional conjugation reagents. The preparation of silica nanoparticles coated with isothiocyanate groups and used them for apoptosis detection has recently been elaborated.[64]

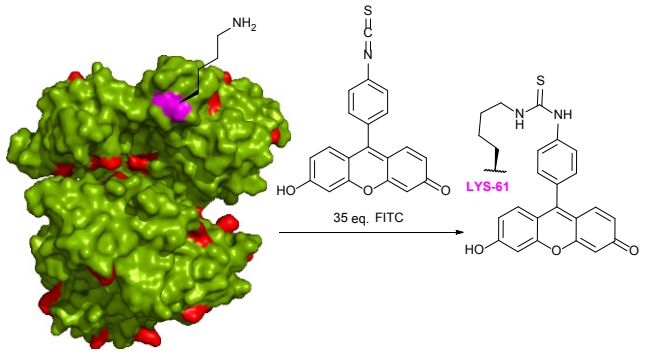


Fig. 4. Selective fluorescent labelling of LYS-61 residue (shown in magenta) of rabbit skeletal muscle G-actin (pdb: 2VYP) reported by Butnick.[42] Such a high level of selectivity towards LYS-61 residue over other 33 lysine residues present in proteins (shown in red) remains unclear, but is hypothesised to be due to anomalously low pK_a value thereof.

The classical protocol of isothiocyanate labelling involves the use of 5-10 equivalents at slightly basic pH from the range of 9.0-9.5.[65,66] Resulting thioureas are reasonably stable in aqueous medium and provide suitable degree of conjugation.[67] For example, Sandmaier and colleagues[68] have recently demonstrated that radiolabelling of anticanine CD45 antibody using isocyanate and isothiocyanate provides a more specific delivery to the targeted CD45-expressing cells than a method exploiting thiol-maleimide conjugation (see Section A 1.3.3, page 13). However, it has been shown by Banks and Paquette[69] that, compared to NHS ester based methodology, antibody conjugates prepared with isothiocyanates are less hydrolytically stable and deteriorate over time. Moreover, the reaction of NHS ester for amine labelling was found to be faster, to give more stable conjugates for both model amino acids and proteins and to proceed readily at lower pH, compared to isothiocyanate. Consequently, NHS esters are preferable over isothiocyanates in many respects for synthesizing bioconjugates.

1.1.2 Activated esters

Because of the poor leaving ability of the alkoxy groups, alkyl esters of carboxylic acids are inert to amines in aqueous media.[70] However, their substitution by good leaving groups activates the carbonyl and renders it susceptible to nucleophilic attack. It is worth remarking, that not only does such activation increase the reactivity of these reagents towards free amino groups, but also often augment their tendency to degrade in the presence of water.[71] Although many activating moieties have appeared over years, only a limited number of them are of significant importance in bioconjugation today. For instance, formerly of a significant importance, especially in the field of peptide synthesis, activated phenyl esters are almost of no use today in bioconjugation because of their lower kinetics and lower stability compared to succimidyl esters.[72,7374] However, they continue to reappear in certain studies.[23]

N-hydroxysuccinimide (NHS) activated esters were introduced in 1963 by Anderson *et al*. as a better alternative for phenyl esters in forming the peptide bond.[30,75,76] Possessing high selectivity towards aliphatic amines, NHS esters are today considered among the most powerful protein-modification reagents. Although several studies drew attention to a certain reactivity of NHS-activated esters with tyrosine,[77-81] serine and threonine (especially when situated in certain locations, see Section A 1.2, page 11),[82-86] these side reactions possess largely decreased rates compared to the reaction with free amines and do not generally hinder the amine-selective derivatisation. High concentrations of nucleophilic thiols should however be avoided because, similarly to isothiocyanates, they may increase the rate of the probe degradation by forming easier hydrolysable intermediates (Fig. 5).

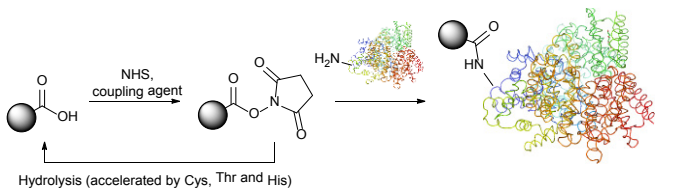


Fig. 5. Preparation of activated NHS-esters and their reaction with nucleophilic amino acid residues present in proteins. Similarly to isothiocyanates, the reversible reactions of thiol and alcohol groups with NHS-esters largely increase the rate of hydrolysis thereof.

The optimum pH for NHS-mediated labelling in aqueous systems was found to be lower than for other amine-selective reagents and is ranging from 7 to 8 units (compare with 9-9.5 for isothiocyanates), which enlarges the prospect of its suitability for modifying alkaline-sensitive proteins. Several elaborated studies of the kinetics[71] and the stoichiometry[87] of the protein

tagging with NHS-containing probes have been recently reported.

Depending on the pH of the reaction solution and temperature, NHS esters are hydrolysed by water (possessing a half-life of 4-5 hours at pH 7, 1 hour at pH 8 and 10 minutes at pH 8.6),[88,89] but are stable to storage if kept well desiccated. Virtually any molecule containing an acid functionality, or a moiety which can give an acid, can be transformed to its *N*-hydroxysuccinimide ester. While the activation with NHS generally decreases water-solubility of the carboxylate molecule, the utilisation of Sulfo-NHS[90] preserves or even increases water-solubility of the modified molecule by virtue of the charged sulfonate group. The development of new reagents based on NHS chemistry can be sometimes challenging,[91] but the derivatives are frequently of a very important usefulness.[92-96] Many NHS derivatives for the preparation of affinity reagents, fluorescent probes and cross-coupling reagents are now commercially available, enabling wide access to investigations.

The formed conjugates are linked by means of a very stable aliphatic amide bond with half-lives in the range of 7 years in water.[97] This excellent stability and biocompatibility of the obtained bonds have provided an exceptional importance of NHS esters in the field of bioconjugation.

NHS ester-mediated covalent conjugation for protein modification has been first accomplished by Becker *et al.*, who studied biotin transport first in yeast[73] and then applied this technique to a covalent attaching of biotin to bacteriophage T4.[98] Since then, the field of NHS-mediated conjugation of proteins has been unceasingly expanding its employability in countless applications.

Cross-linking of proteins often implies using of NHS-containing homobifunctional or heterobifunctional cross-linking reagents. These were used for elucidation of protein-protein[99-103] and protein-drug interactions,[104] protein structural and subunit analysis,[26,105] create protein complex models,[106] and preparation of protein conjugates with enzymes, drugs or other macromolecules.[107-109]

Homobifunctional NHS cross-linkers are generally used in reaction procedures to randomly "fix" or polymerize peptides or proteins through their amino groups. Adding such crosslinker to a cell lysate will result in random conjugation of interacting proteins, protein subunits, and any other polypeptides whose Lys side chains happen to be in the near proximity to each other. This represents a methodology for capturing a "snapshot" of all protein interactions at a certain instant of time. Using this approach, for instance, Sinz and collaborators were able to elucidate binding of calmodulin to mettilin, a polypeptide and principal component of honeybee venom, without chromatographic separation techniques.[110] Cross-linking of the proteins with a number of different length NHS-homobifunctional cross-linkers, and following digestion of obtained products with trypsin and analysing by HPLC enabled the possibility of three-dimensional structure modelling of the calmodulin-mettilin complex (Fig. 6).

Several applications however require the precision of crosslinking which cannot be provided by homobifunctional crosslinkers. For example, the preparation of an antibody-drug conjugate (ADC) implies selective linking of a cytotoxic payload to each molecule of antibody without causing any antibody-to-antibody linkages to form. For such application the combination of different selective approaches in one linker is needed.

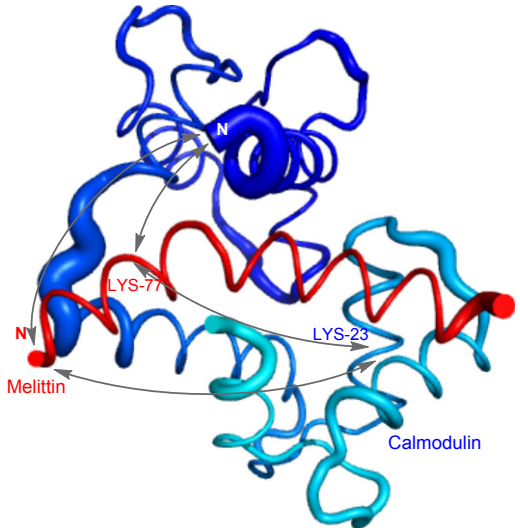


Fig. 6. Mode of binding of melittin in the calmodulin-melittin complex (pdb: 2MLT and 1CDL) calculated from ambiguous distance restraints derived from the cross-linking data by Sinz and associates.[110]

Therefore, heterobifunctional crosslinkers are designed to possess different reactive groups at either end. These reagents allow for sequential conjugations that diminish undesirable self-conjugation and polymerisation. Sequential procedures involve two-step processes, where heterobifunctional reagents (often in excess to ensure high conversion levels) are reacted with one protein using the most labile group of the crosslinker first. After eliminating the excess of nonreacted crosslinker, the second protein is added to a solution containing modified first protein and another reaction occurs with a second reactive group of the crosslinker. According to Pierce website (Rockford, IL, USA), the most popular heterobifunctional crosslinkers are those having amine-reactive NHS esters at one end and thiol-reactive maleimides (see Section A 1.3.3, page 13) on the other end. Because of its less stability in aqueous solution compared to maleimide, NHS-ester group should usually be reacted first. Takeda and co-workers[111] used a bifunctional reagent that contained a NHS function and a benzylthioester function to prepare a DNA–protein hybrid. One of the fastest growing fields requiring heterobifunctional crosslinkers today is targeted drug delivery therapies - ADCs.[22,112-116] They are constituted of three main components: one monoclonal antibody (mAb), targeting specific signs or markers of cancer cells, one cytotoxic agent, and one linker molecule that allows covalent drug binding to the mAb. The composition of trastuzumab emtansine (Kadcyla®, Genen-

tech) - an in clinic ADC for treatment of HER2-positive metastatic breast cancer is depicted in the Fig. 7.

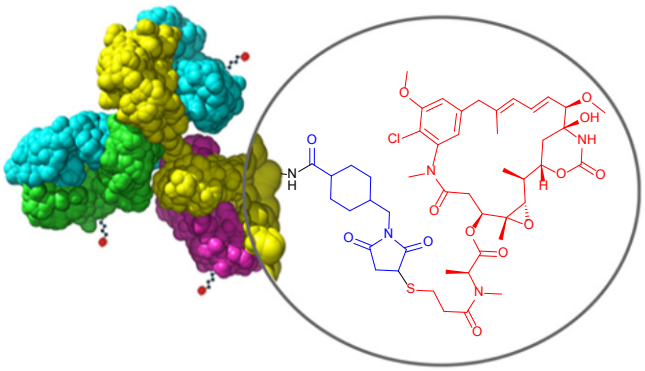


Fig. 7. Structure of an antibody drug conjugate (ADC) Kadcyla®. SMCC linker (shown in blue) serves for conjugation of the antibody and cytotoxic payload (Mertansine, shown in red).

The first example of "cleavable" NHS cross-linking reagent, DSP was reported by Lomant and Fairbanks[89] and allowed effecting reversal of the previously conjugated fragments under mild conditions of disulfide bond reduction. Further advances in the field have resulted in various types of linkers, cleavable under mild nucleophilic conditions (EGS),[117] at basic pH (BSOCOES),[118] in the presence of periodate (DST),[119] or enzymatically.[120] These found their applicability for studies in basic and applied research. The reader is directed to a recent review by Leriche *et al.*[121] that provide an overview of chemical functions that can be used as cleavable and to a publication by Jin Lee[122] for an overview of commercially available cross-linking reagents.

Other combinations of functionalities have been studied over last 20 years and resulted in elaboration of heterotrifunctional[123] linkers usually combining two bioselective reactive groups and a functionality for anchoring of obtained conjugate (*e.g.* biotin moiety).

Many chemical probes widely used in bioconjugation contain NHS-fragment in their structure and are designed to react with free amino groups of proteins. For example, biotinylation,[124-126] as well as PEGylation[2] of proteins are most commonly achieved using NHS-activated probes today. It was recently reported by Anderson and collaborators, that biotinylation of antibodies with NHS-biotin and their following adsorption on the surface of nanocrystal quantum-dots (QD) results in obtaining of highly efficient QD-Antibody conjugates for the detection of protein toxins.[127] Other types of protein immobilisation on matrices have also been reported.[88,128,129] The Bolton–Hunter reagent (SHPP),[130] allowing conjugation of tyrosine-like residues for increasing the yield of subsequent (radio)iodination is also based on *N*-hydroxysuccinimide chemistry.[131-136] Elaborated in 1982 by Ji *et al.*[137] synthesis of structurally similar to SHPP photoactivable heterobifunctional probes for cross-linking experiments have been used in more than 100 studies ever since. NHS ester-based strategy for isobaric, stable isotope labelling of peptides[138-140] has recently found more widespread applications in proteomic studies with a simultaneous developments of enhancing peptide detection by electrospray ionisation mass spectrometry.[77,141] This list can be continued and arguably utilisation of NHS-mediated techniques can be found in all major fields of protein conjugation and represents a gold standard in bioconjugation.

1.1.3 Reductive Amination of Aldehydes

Aliphatic and aromatic amines react under mild aqueous conditions with aldehyde groups to form an imine (known as a Schiff base). This intermediate can then be selectively reduced by a mild reducing agents, such as sodium cyanoborohydride,[142] to give a stable alkylamine bond. Although this approach for amine modification is not used in protein conjugations as frequently as the activated ester or isothiocyanate method, it is to be considered as preferable when the molecule to be attached has an aldehyde group (or can be easily converted to an aldehyde) because of its simplicity and mild reaction conditions.

Historically the conjugation of oligosaccharides to proteins has become the first target for this approach. In 1974, relying on the described three years earlier by Borch[142] exceptional ability of cyanoborohydride anion to reduce selectively Schiff bases generated *in situ* from an amine and an aldehyde, Gray has illustrated the possibility of mild synthesis of carbohydrate coated bovine serum albumin (BSA) and P150 protein (Fig. 8).[143]

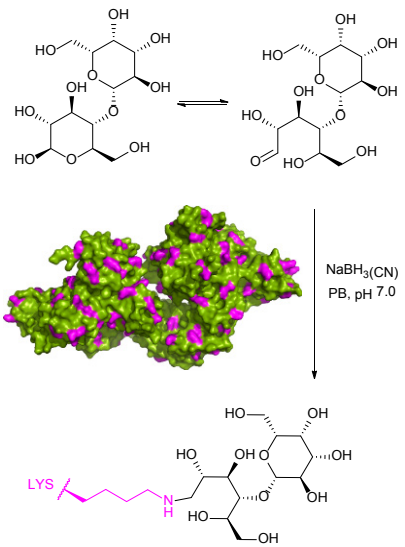


Fig. 8. First example of reductive amination of lactose by bovine serum albumin (BSA, pdb: 3V03) described by Gray.[143] Only about 4 of the 59 lysine residues theoretically available (shown in magenta) were derivatised in 2 weeks.

Reductive amination on proteins proceeds most readily at pH 6.5-8.5 where the reduction of aldehydes and ketones is negligible, and, if feasible, in an alcoholic solution under dehydrating conditions where the rate-limiting formation of the imine is favoured. According to Allred and colleagues,[144] the addition of sodium sulfate (500 mM) may largely improve the coupling efficiency in aqueous media.

To date, reductive amination still plays a central role in the synthesis of carbohydrate-protein conjugates,[20,145-147] that have been used for years to study molecular recognition of carbohydrates.[148] Among these conjugates are FDA approved and used routinely for the prevention of invasive bacterial infections polysaccharide-protein conjugate vaccines such as Menactra, HIBTiter, and Prevnar (Fig. 9),[20,149] potential anti-infective and anti-cancer agents, currently in clinical trials.[23,147,150,151] The reader is directed to a recent comprehensive review of Adamo *et al*.[24] covering current status and future perspectives of carbohydrate-protein conjugates.

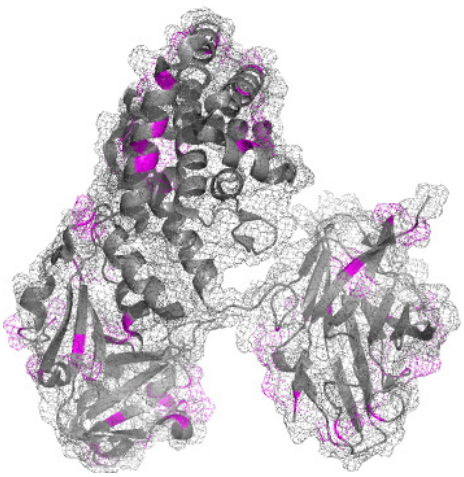


Fig. 9. Structure of Prevnar 13 vaccine. The bacterial capsule sugars, a characteristic of the pathogens, are linked to CRM197, a nontoxic recombinant variant of diphtheria toxin (pdb: 4AE0), by reductive amination at lysine residues and *N*-terminus (shown in magenta).

Another reported applications of reductive amination include the preparation of an organic trialdehyde to be used as a template for the synthesis of three-helix bundle proteins,[152] protein PEGylation[153,154] and immobilisation.[155]

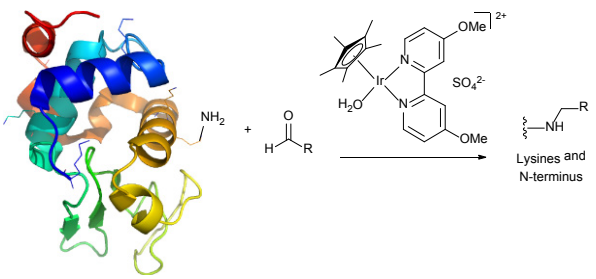


Fig. 10. Modification of lysozyme (pdb: 2LYZ) using reductive alkylation with water-stable iridium catalyst developed by McFarland and Francis.[157]

Reductive amination however possesses several drawbacks preventing it from being generally applicable to protein conjugation.[156] The most important is the necessity to use water-sensitive sodium cyanoborohydride, which has a potential for reducing disulfide bonds within proteins. As an alternative, McFarland and Francis[157] have elaborated a water-stable iridium catalyst (Fig. 10). However, the efficiency of the method is lower than of the classical reduction with cyanoborohydride.

1.1.4 Sulfonyl Halides and Sulfonates

Introduced in 1952 by Weber[158] for fluorescent labelling of macromolecules, dansyl chloride (DNSC) was the first widely used sulfonyl chloride for the modification of proteins. It has gained incontestable popularity for the study of proteins after Hartley and Massey have successfully used it for the determination of the active centre of chymotrypsin.[159] DNS*C*-Edman degradation was proposed by Gray[160] to improve the ease and reproducibility of a classical isothiocyanate-based Edman degradation.[37]

Sulfonyl halides are highly reactive but also very unstable, especially in aqueous media at the pH required for reaction with aliphatic amines. For example, Haugland and collaborators[161] have demonstrated that the rate of hydrolysis of Texas Red (one of the most widely used long-wavelength fluorescent probes)[162] and Lissamine rhodamine B sulfonyl chloride was much higher (complete hydrolysis within 5 minutes in pH 8.3 aqueous solution), than this of corresponding NHS esters (both retained most of their reactivity for more than an hour under the same conditions). Yet, the formed sulfonamide bonds are extremely stable and even survive amino acid hydrolysis,[160,163] which makes sulfonamide conjugates useful for the applications where the stability of the conjugation bond is a crucial feature.

Optimal conditions of proteins modification by sulfonyl chlorides are those under which free amino groups most effectively compete with water for the limited amount of the reagent. It is thus best done at low temperature at pH 8.5-9.5.[164] At lower pH values, the unreactive protonated form of amines slows the labelling reaction compared to the hydrolysis by water, above this range the reagent is hydrolysed too rapidly.[165,166] In practical experiments, a several-fold excess of reagent is usually added, providing the unused probe is hydrolysed to the corresponding unreactive sulfonic acid after the labelling. It must be borne in mind, that unlike other amine-selective reagents, sulfonyl chlorides are unstable in dimethylsulfoxide, classically used for the preparation of stock solutions, and should never be used in this solvent (Fig. 11).[167]

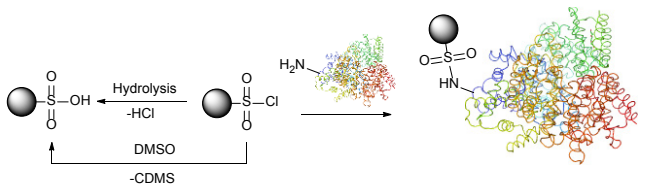


Fig. 11. Reaction of sulfonyl chlorides with amino groups present in proteins. Hydrolysis of the starting material by water or dimethylsulfoxide[167] (chlorodimethylsulfide, CDMS, is a leaving group) results in obtaining unreactive sulfonic acid.

Apart from being reported for fluorescent labelling of proteins,[168] sulfonyl chlorides were used to incorporate chelate moiety into proteins,[169] to study hydrodynamic properties or

introduce long-lived fluorescence labels to macromolecules using tagging with pyrene derivatives[170,171] or as cross-linking reagents.[172]

Because of their very high reactivity towards nucleophiles, sulfonyl halides also form conjugates with tyrosine, cysteine, serine, threonine, and imidazole residues of proteins;[173] therefore, they are less selective than either NHS esters or isothiocyanates. These conjugates are however unstable and can be completely hydrolysed under basic conditions.

Covalent immobilisation of proteins on hydroxyl group containing carrying supports (such as agarose, cellulose, diol-silica, or polylactic acid films) is often accomplished by transforming the latter to corresponding sulfonates: tosylate, mesylate, or tresylate,[174,175] serving as good leaving groups (Fig. 12).[176-178] Albumin, cytokines and other therapeutic proteins and peptides were reported to undergo mild PEGylation by means of PEG tresylates.[179-181] Although rather specific to amino groups, the chemistry of tresylate-mediated conjugation is not unique and well defined. For instance, Gais *et al*. have shown that PEG-tresylate conjugation can produce a product that contains a degradable sulfamate linkage resulting in heterodispersity of the reaction.[182]

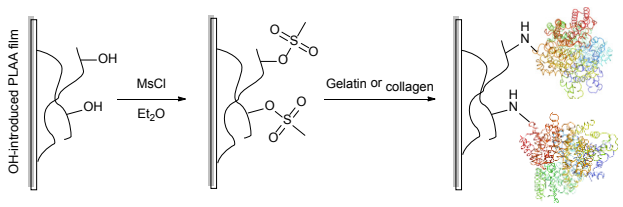


Fig. 12. Schematic representation of the reaction protocol for immobilisation of protein of PLLA film surfaces describe by Ma *et al*.[177]

1.1.5 Fluorobenzenes

Despite their utmost importance for protein modification and amino group quantifying since Sanger and Tuppy's work on the structure of insulin,[183] derivatives of fluoronitrobenzene are of limited usefulness for bioconjugation.

Compared to other aryl halides, fluoro-substituted nitrobenzenes were found to be the most reactive in bimolecular nucleophilic substitution reaction.[184] They are usually regarded as amino-selective reagents, despite their known reactivity towards thiolates, phenolates and imidazoles, for the products obtained in these reactions are either unstable at alkaline pH required for the reaction (Tyr and His), or can be thiolysed by excess β-mercaptoethanol (Cys).[185]

4-fluoro-7-nitro-2,1,3-benzoxadiazole (NBD-F), which has been introduces as a fluorogenic reagent for more than 30 years ago by Imai and Watanabe,[186] still remains important for several applications, mainly pre-column derivatisation and enrichment of peptides. The reader referred to a recent review by Elbashir *et al*.[187] providing an excellent overview of the NBD-F applicability to the analysis of peptides and to a complete overview of NBD-mediated methodologies for the fluorescent labelling of amino acid residues by Imai and associates.[188,189]

An elegant approach for improving protein crystallizability, still remaining a major challenge in protein structure research,[190] was elaborated by Sutton and collaborators[191] and consists in the introduction of a charged ammonium residue. It exploits the amine-selective derivatisation of protein by 1-fluoro-2-nitro-4-trimethylammoniobenzene iodide (Fig. 13) and results in increasing of the hydrophilicity thereof. Using their approach, the authors were able to study the binding site[192] and to obtain crystalline derivatives of modified bovine insulin,[193] which is especially hard to crystallize without inducing a structural changes.[194] Similar protocol was used by Ladd *et al*.[195] for chromophorical PEGylation of proteins with polyethylene glycol fluoronitrobenzene derivatives.

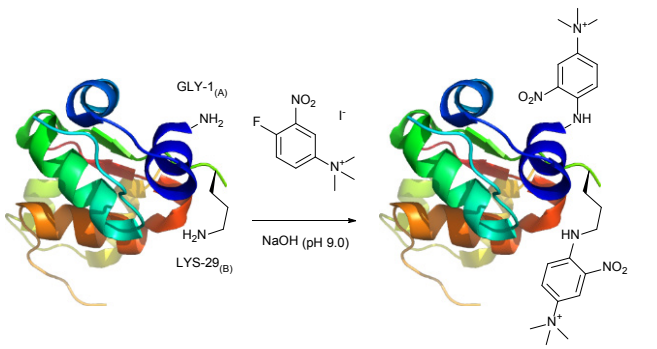


Fig. 13. Derivatisation of bovine insulin with 1-fluoro-2-nitro-4-trimethylammoniobenzene iodide described by Sutton *et al*.[191] Only amine-containing residues of A1 and B1 chains are shown. Two of the four tyrosine residues present in chains A1 and B1 (not shown) also react with the probe under described conditions.

1.1.6 Imidoesters

First investigated by Hunter and Ludwig in 1962,[196] the reaction of imidoesters with peptides reached its climax for protein modification ten years after, when Traut *et al*. have introduced 2-iminothiolane reagent (today carrying Traut's name) for cross-linking.[197] It allowed for producing disulfide-linked dimers of neighbouring proteins on the intact 30s ribosome from *E. coli* using two-step procedure: the reaction of ribosomal amino groups with the imidoester function followed by the mild oxidation of the obtained thiolate-charged ribosomes (Fig. 14). An excellent review of the cross-linking studies for the determination of ribosomal structure was published by Nomura.[198]

Imidoesters react with primary amines to form amidine bonds. However, sufficient levels of specificity towards amines may only be achieved when alkaline conditions (pH 10) and amine-free media, such as borate buffer, are used. Because resulting amidine bonds are protonated at physiological pH, positive charges near modified sites are preserved during the conjugation with lysines and *N*-termini. Consequently, as it was first demonstrated by Wofsy *et al*.[199] such modifications produce little or no significant changes in the conformational properties and biological activities of proteins.

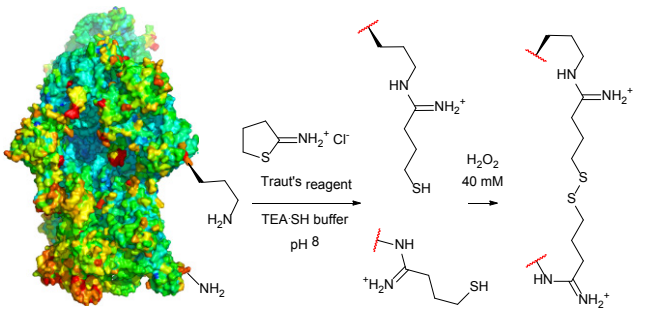


Fig. 14. Procedure of the cleavable crosslinking of the intact 30s ribosomes (pdb: 1J5E) described by Traut *et al*.[197] Lysine residues (LYS-72 and LYS-156) were chosen randomly for simplicity purposes. (TEA·SH stands for triethanolamine buffer adjusted to 3% 3-mercaptoethanol)

Thiolates obtained after the ring opening of Traut's reagent by free amines enable a plentiful thiol-selective chemistry on modified α- or ε-amino groups of proteins (see Section A 1.3, page 11). Although, many imidoesters other than Traut's reagent are today commercially available (for example, see DMA DMP, or DTBP), the amount of described labelling imidoester probes is rather scarce.

Schramm *et al*.[200] have described the synthesis of fluorescent imidoester dyes from corresponding nitriles; the approach was later used by Bozler *et al*.[201] for the preparation of dansyl containing imidoester and selective modification of lysine residues in the active site of glucose dehydrogenase. New readily available reagents for the attaching of sugars to proteins *via* imidoester linkage,[202] hydrophilic spin probes for determining of membrane protein interaction using EPR,[203] immunoreactive probes,[204] tyrosine-like probes for radioactive labelling with ^{125}I,[205] protein PEGylation reagents,[206] the immobilisation of trypsin, yeast alcohol dehydrogenase, and *E. coli* asparaginase onto several types of organic polymer beads[207] were achieved *via* imidoester conjugation and proven to have several advantages compared to other existing methodologies: namely, deprivation from solubility issues and retention of positive charge at the reaction site.

1.1.7 Miscellaneous amine-selective reagents

Several methods of amine-selective modification of proteins were not included in the main chapter, either because of scanty information available or when their applicability is reduced to a specific substrate types and is not general.

Azetidinone chemistry has recently been demonstrated by Barbas and collaborators[208,209] to have potential for selective lysine labelling of a particular IgG framework, containing a very reactive lysine residue with unusually low pK_a of about 6. Some detailed procedures are described for a smooth opening of a β-lactam moiety resulting in a β-alanine peptide bond.[209]

Discovered by Tietze *et al*. as two-step sequential procedures for coupling of amines,[210,211] squaric acid diester amine-amine conjugation is now actively developing by Wurm *et al*., who have recently reported their successful use for the one-pot preparation of poly(glycerol)-protein[212] and glycol-protein conjugates[213] in aqueous media (Fig. 15).

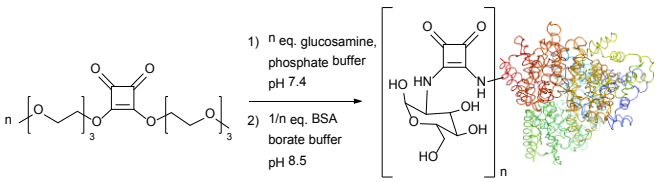


Fig. 15. One-pot, two-step squaric acid diester mediated glycosylation of BSA (pdb: 3V03) described by Wurm *et al*.[213] Up to 22 lysine residues of the 59 present in BSA (30-35 are available for post-modification) could be glycosylated with 25-fold excess of squaric diester and glucosamine in 12 hours.

Dichlorotriazine derivatives were described for amine-selective conjugation mainly as fluorescent dyes[214,215] and PEGylation probes.[216-218] They were shown to possess high reactivity towards protein amines. However, as it was demonstrated by Abuchowski *et al*.,[216] because the hydrolysis of the diclorotriazine occurs readily under slightly basic conditions (pH 9.2) needed for reaction to pass with sufficient selectivity towards amines, a considerable excess of the probe must be used in the coupling reaction. Banks and Paquette[69] have conducted a comparative study of three fluorescent probes, differing only by the moiety responsible for the reactivity with amines: CFSE (NHS ester), DTAF (dichlorotriazine) and FITC (isothiocyanate). It was found, that the rate of conjugation is significantly faster for the NHS ester compared to the diclorotriazine probe, which, in turn, reacts faster than isothiocyanate derivative. Each conjugate provided a satisfactory level of stability in solution over the period of 1 week at 25 ˚C, although the hydrolysis of the remaining, relatively inert, chloro group of DTAF was observed.

Arpicco *et al*.[219] have prepared thioimidoester activated PEG-containing derivatives and shown their superiority over NHS-activated analogue for gelonin modification (the reaction was conducted in PBS at pH 7.4). In this particular case, PEGylation with less active, compared to NHS ester, thioimidoester derivative resulted in gelonin conjugate with higher inhibiting activity. Ikeda and associates[220] have recently described a protocol for the preparation of glutalaldehyde-functionalised PEG reagent, allowing for proteins PEGylation under mild reaction conditions. Similarly, the modified protein exhibited higher biological activity than when reacted with a corresponding NHS-activated PEGylation reagent.

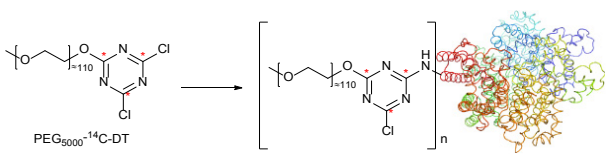


Fig. 16. Modification of BSA (pdb: 3V03) by PEG-[^{14}C]diclorotriazine probe reported by Abuchowski *et al*.[216]

α-Halocarbonyls such as iodoacetamides, can modify lysine residues at pH > 7.0,[221] but the reaction rate is much slower than the reaction with cysteine residues. Another class of reagents usually used in cysteine-selective conjugation – vinylsulfones[222] (see Section A 1.3.4, page 14) – was recently reported to be applicable for lysine labelling at slightly basic pH.[223,224]
Modification of Lys residues with acid anhydrides, including succinic, citraconic, maleic, trimellitic, *cis*-aconitic, and various phthalic anhydride derivatives belongs to a pool of classically used protein modification methodologies[225] and provides for transforming of nucleophilic amines to acids and, as a result, enables carboxylate-selective chemistry thereof.
For more details on the practical aspects of using the above-described methodologies, the reader is referred to a recent review by Brun and Gaury-Lazo[226] on the preparation of antibody-drug conjugates by lysine conjugation.

1.2 Serine and Threonine

With pK_a values > 13, the hydroxyl groups of serine (Ser, S) and threonine (Thr, T) are rather poor nucleophiles close to physiological pH. No examples of direct conjugation of in-chain serine and threonine have been therefore reported to date.
However, highly amine selective *N*-hydroxysuccinimide (NHS) esters have been documented to give occasional side reactions with hydroxyl side chains.[227-230] In a series of experiments, Miller *et al*. have demonstrated that the presence of histidine in sequences of the type His-AA-Ser/Thr or His-AA-AA'-Ser/Thr (where AA, AA' stand for any amino acid) can significantly increase the reactivity of hydroxyl groups toward classical amine labelling agents (Fig. 17).[82,83,85,231]

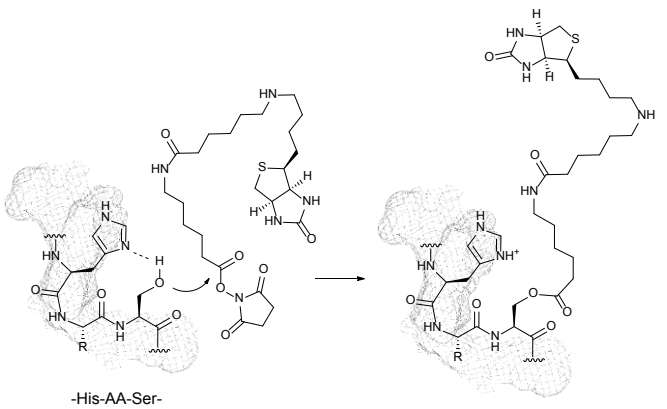


Fig. 17. Nucleophilic attack of serine on the NHS-activated ester gives the stable *O*-acylated derivative. In addition to His-AA-Ser, this stable intermediate can be formed in the presence of linear sequences of His-AA-Thr and His-AA-Tyr, where AA refers to any amino acid.

Similarly, Mädler and Zenobi have reported that the guanidinium group of arginine can contribute to the reactivity of hydroxyl groups toward NHS esters and to catalyse the nucleophilic substitution.[232] In both cases, it is hypothesised that the imidazolyl and guanidine moieties of histidine and arginine respectively catalyse the reaction by stabilizing the transition state by means of hydrogen bonds and electrostatic interactions. This promoting effect is thought to be responsible for side reactions on several substrates while using cross-linking reagents.[229,232]
Despite the fact that methodologies of selective in-chain serine and threonine labelling are rather scarce, these residues represent especial interest for bioconjugation when located on the *N*-terminus (see Section A 2.1, page 29).

1.3 Cysteine

Cysteine (Cys, C) is perhaps the most convenient target for bioconjugation owing to the exceptionally high nucleophilicity of its sulfhydryl (-SH) side chain which, and particularly in its deprotonated thiolate form ($-S^-$), largely exceeds the reactivity of any other nucleophilic function in proteins.[233] Furthermore, its relative rarity in proteins composing living organisms (1-2%)[234,235] and the ease of its introduction into a specific site by site-directed mutagenesis allows access to protein assemblies with a single cysteine at a predetermined position.[236] Even in proteins with multiple cysteines, the multiplicity is usually much smaller compared to lysines, which makes thiol-reactive labelling the preferred approach over amine-reactive methodologies.
In proteins, thiols can also be generated by selectively reducing cystine disulfides with reagents such as dithiothreitol (DTT, D1532),[237] 2-mercaptoethanol (β-mercaptoethanol), or tris[2-carboxyethyl]phosphine (TCEP).[238,239] Generally, all these reagents must be removed before conducting thiol-selective conjugation, for they will compete with target thiols in proteins otherwise.[240] Unfortunately, removal of reducing agents is sometimes accompanied by air oxidation of the thiols back to the disulfides. Although, in contrary to the majority of thiol-reducing agents, TCEP does not contain thiolate group, there have been several reports that it can react with α-halocarbonyls or maleimides and that labelling is inhibited when TCEP is present in the reaction medium.[241,242]
Direct labelling of thiolate group is usually achieved by either a nucleophilic addition or displacement reaction with thiolate anion as the nucleophile. The substantially lesser dissociation energy of sulfhydryl groups compared to the corresponding alcohols provides much higher acidity of the former and, as a consequence, a wider availability of it slightly nucleophilic anionic form at physiological pH.

1.3.1 α-Halocarbonyls

First reports on the use of α-halocarbonyl electrophiles, namely iodoacetamide, go back to 1935, when Goddard and Michaelis[243] have first reported its application for modifying and studying keratin. Even today, almost 80 years later, these electrophiles are still among the most widely used for the modification of cysteine, especially in mass spectral analysis and peptide mapping of cysteine containing proteins.[244] Use of iodo compounds is typical because, as iodide is a better leaving group among other halogens, these render higher reaction rates for conjugation (the relative reactivity is I > Br > Cl > F). Iodoacetyl-containing crosslinkers, biotinylation reagents, immobilisation kits, and mass spectrometry tags are now commercially

available (*e.g.* BIAM, SIAB, UltraLink™ Iodoacetyl Resin and Gel, iodoTMT™). Although corresponding maleimide reagents are more popular because of their even higher reaction rates, the haloacetyl-mediated conjugations are usually preferred for the applications where the elevated stability and compact size of the generated linkage (compared to maleimide) are crucial. Indeed, such bioconjugates degrade to S-alkyl cysteine derivatives only during amino acid hydrolysis.

Typically, the reaction of sulfhydryl groups with haloacetamides is conducted at physiologic to alkaline conditions (pH 7.2-9.0). When iodoacetamides are used, the reaction is preferably carried out under subdued light in order to limit free iodine generation, which has the potential to react with Tyr, His and Trp residues. The reaction is most specific for sulfhydryl groups at pH 8.3. The iodoacetyl group is known to react with other amino acid side chains, especially when there are no cysteine present or if a gross excess of iodoacetyl is used. For instance, free amino groups, the thioester of methionine, and both imidazolyl side chain nitrogens will react with iodoacetyl groups above pH 7 and pH 5, although with much slower kinetics.[245] This, however, can be resolved by the use of less reactive chloroacetamides[246] or cautious control of pH and incubation time.

It is to be noted that local environment has a profound effect on the reactivity of cysteine residues in proteins. If moderately reactive reagents such as iodoacetamide are used for bioconjugation, this difference in reactivities makes it possible to discern different types of Cys moieties present in the protein. Almost half a century ago, Gerwin[247] reported dramatic differences in the reactivities of chloroacetic acid and chloroacetamide in the modification of the active-site cysteine of streptococcal proteinase, which was found to be due to the influence of the neighbouring histidine residue. As a general trend, cysteine residues possessing lower pK_a values are more reactive when reaction is conducted under neutral or slightly acidic conditions, owing to their greater degree of dissociation and, as a consequence, higher concentration of the corresponding thiolate anions in the medium. For instance, Kim *et al.*[248] have described a method for selective biotinylation of low-pK_a cysteine residues in proteins simply by conducting the reaction at slightly acidic pH (Fig. 18).

Davis and Flitsch[249] described a procedure for the selective glycosylation of proteins at one or several sites by reacting the carbohydrate-tethered iodoacetamides with cysteine side chains, which allowed for preparing of homogeneously glycosylated human erythropoietin[250] and dihydrofolate reductase.[251]

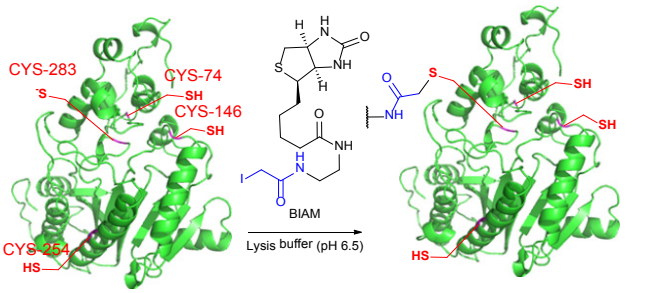


Fig. 18. Biotinylation of low-pK_a cysteine residue of rabbit muscle creatine kinase (CK, pdb: 2CRK) by BIAM.[248] The charge interaction between the negatively charged thiolate and the positively charged amino acid residues nearby results in significantly lower pK_a value of the CYS-283 residue (6.5). Consequently, selective alkylation thereof becomes possible in the presence of three other cysteine residues with higher pK_a values (8.0-9.0).

1.3.2 α-Haloacetophenones

In 1948, Mackworth[252] has published his study upon the reactivity of the biochemical mechanism of the lachrymatory effect of certain war gases and first reported the reactivity of structurally relevant α-bromoacetophenones for the inhibition of several classes of thiol enzymes.

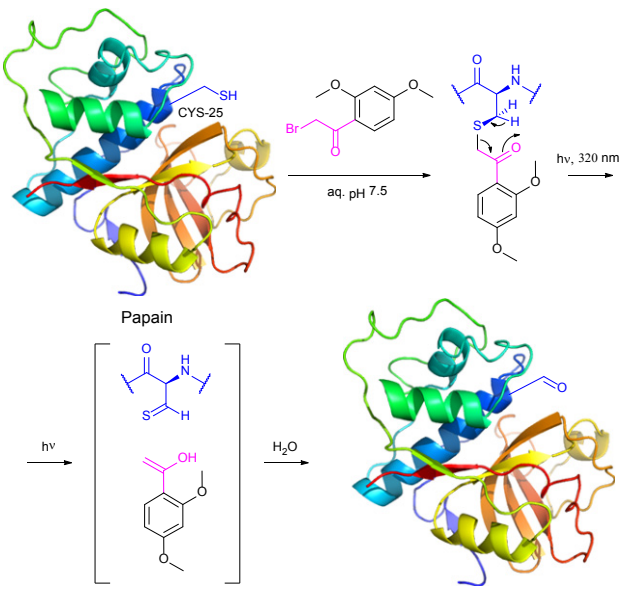


Fig. 19. Photolysis of CYS-25 residue located in papain's active site after its labelling with 2-bromo-2',4'-dimethoxyacetophenone results in the formation of the unstable thioaldehyde, which readily hydrolyses to give the corresponding aldehyde.[253,254]

Despite advances made in the investigation of α-haloacetophenones and related ketoximes for the modification of the active sites of enzymes,[255-258] their utility for the conjugation is very limited because of various side reactions.

An interesting approach that allows photochemical conversion of the cysteine into corresponding thioaldehyde and then to aldehyde thought Norrish type II cleavage was reported by Clark and Lowe.[253,254] Photolysis of the enzyme, alkylated by a bromoacetophenone derivative, results in spontaneous loss of

hydrogen sulfide from the generated thioaldehyde to give the corresponding aldehyde (Fig. 19), which can either be utilised as a locus for aldehyde-selective conjugation, or be transformed into corresponding serine or glycine residue by reduction or transamination respectively

1.3.3 Maleimides

As early as 1949, maleic acid imides (maleimides), products of the reaction of maleic anhydride and amine derivatives, were introduced by Friedmann as cysteine-specific reagents.[259,260] Ever since, persistently gaining in popularity maleimide-mediated methodologies represent today perhaps the most often used functional groups for bioconjugation. This is mainly due to their exceptionally fast kinetics and significantly high selectivity toward cysteine moiety in proteins.

The reason for such remarkable reactivity of maleimide towards thiolates is worth being discussed. In general, the electrophilicity of alkenes is defined by its ability to serve as an acceptor of nucleophile's electron density, and thus interrelated with the energy of electrophile's π* orbital (its lowest unoccupied molecular orbital, LUMO). Generally speaking, the rule is simple: the lower the energy of the alkene's π* orbital – the faster its reaction with nucleophiles. There exist two main approaches for decreasing alkene's LUMO energy: the direct attaching of an electron-withdrawing group (EWG) and the straining of the double bond. Although proceeding *via* two different mechanisms: by decreasing the energy of both orbitals or by diminishing the energetic gap between them, either approach results in the lowering the LUMO energy of the alkene and, as a result, in the increase of its reactivity (Fig. 20). The unique reactivity of maleimide moiety owes to the fact that it exploits these two mechanisms together.[261,262]

To date, a large variety of maleimide-based modifying reagents are available from a number of leading biochemical companies with even more being synthesised in laboratories around the world for specific applications. The applications of these reagents strongly overlap those of iodoacetamides, although maleimides apparently do not react with methionine, histidine or tyrosine.[263,264]

The optimum reaction conditions for maleimide-mediated conjugation, namely conducting the reaction at near neutral conditions (pH 6.5-7.5), allow the reaction of maleimide with amines being avoided, because the latter requires a higher pH to occur. At pH above 8 the hydrolysis of maleimide itself results in obtaining the mixture of isomeric non-strained maleamic acids unreactive toward sulfhydryls and can thus compete with thiol modification.[28,265] Similarly, maleimide-thiol adducts hydrolyse, which either results in complete deconjugation or cause a significant change in the properties of the conjugate.[265] Furthermore, especially at pH above 9, ring-opening by nucleophilic reaction with an adjacent amine may yield crosslinked products.[266]

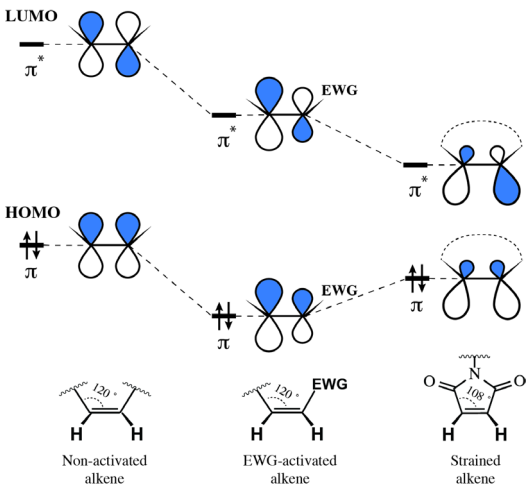


Fig. 20. Influence of an electron-withdrawing group (EWG) and cycle strain on the frontier orbitals of alkenes, σ-orbitals omitted (the form of the orbital is presented approximately, based on the publication of Merchan *et al*[267]). Decreasing the energy of LUMO (lowest unoccupied molecular orbital) results in higher reactivity of the electrophile towards nucleophiles. Although *via* different mechanisms, both EWG and strain of the cycle activate alkene for the attack by nucleophiles.

Schuber and co-workers[268] have found that important kinetic discrimination can be achieved between the maleimide and bromoacetyl functions when the reactions with thiols is conducted at pH 6.5 and 9.0, respectively.

Maleimide-NHS heterobifunctional reagents are especially important for the formation of conjugates. Hydrolysis of both maleimide moiety and generated thioester linkage is considerably dependant on the type of chemical group adjacent to the maleimide. Interestingly, the cyclohexane ring was found to provide increased maleimide stability to hydrolysis due to its steric effects and its lack of aromatic character. For this reason, succinimidyl-4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) and its water-soluble analogue (sulfo-SMCC) are today among the most popular crosslinkers in bioconjugation. They are often used in the synthesis of the protein-protein or protein-probe assemblies such as antibody-enzyme or antibody-drug conjugates respectively. These include the enzyme immunoassays,[269-273] carrier-protein conjugates,[274-276] albumin-binding prodrugs,[277,278] and even approved therapies.[113,115,279,280]

Short homobifunctional maleimide are commonly used to explore and characterize protein structure (*i.e.*, oligomerisation) or protein interactions.[281-287] Maleimide-mediated immobilisation of biomolecules is often achieved either by direct conjugation[13,288,289] or by prior biotinylation of the molecule of interest.[290] The latter approach has been used for protein enrichment,[291] capture,[292,293] and immobilisation on modified supports.[294-297]

Most of the optical thiol-selective fluorescent probes often used as sensors for monitoring of biological processes are represented by maleimide-containing reagents.[298-300] Another testament to maleimide utility is their use for glycosylation,[301] radiolabelling, [302,303] studying of protein interactions,[304-307] and quantitation of cysteine residues.[308,309]

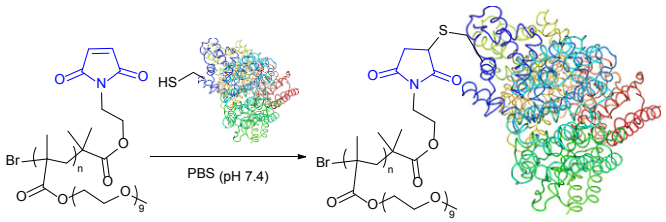


Fig. 21. Conjugation of maleimide-functional poly(PEGMA) to the only free CYS-34 residue of BSA (pdb: 3V03).[310]

Despite of its successful application as a reagent for the chemical modification of proteins, the irreversibility of maleimide's addition makes it impossible to regenerate the unmodified protein by controlled disassembly of the conjugate. Such necessity is however often desirable for *in vitro* or *in vivo* applications. Several studies were devoted to a mild and specific hydrolysis of imido group in maleimide conjugates.[311,312] These approaches turned the originally irreversible maleimide-mediated thiol conjugation into the cleavable methodology. However, the harsh reaction conditions of this cleavage (strong basic conditions or presence of high amount of imidazole) make them incompatible with many fragile protein substrates.

Monobromomaleimide derivatives, introduced by Baker *et al*.[313] in 2009 have expanded the class of reagents for the selective and reversible modification of cysteine (Fig. 22). In contrast to methanethiosulfonates (see below),[314] monobromomaleimides allow much more stable conjugation of thiolates, which are easily cleavable upon reaction with TCEP by addition-elimination sequence (Fig. 22).

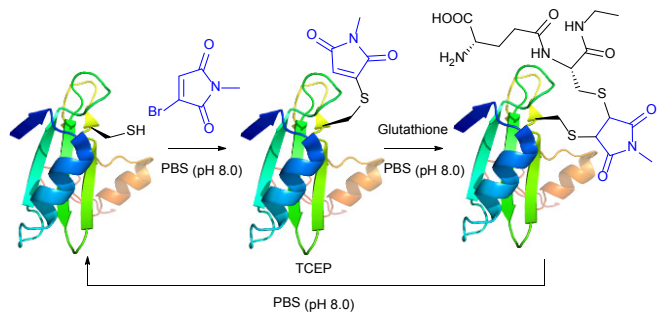


Fig. 22. Conjugate addition of L111C mutant of SH2 domain of Grb2 (wild-type pdb: 1JYU) to bromomaleimide followed by second addition of glutathione resulting in generation of the vicinal bis-cysteine adduct.[315]

Moreover, the initial modification of a protein resulted in obtaining a thiol-maleimide moiety, which was shown to be prone to a second thiol addition and resolved another recognised drawback of maleimide-based methodologies, namely the presence of only two points of attachment.[315] Similarly to a non-substituted maleimide, the hydrolysis of thiol-maleimide linkage results in dramatic decrease in its reactivity towards thiolates, which can be used for "switching off" the linker after the first thiol addition (Fig. 23).[316,317]

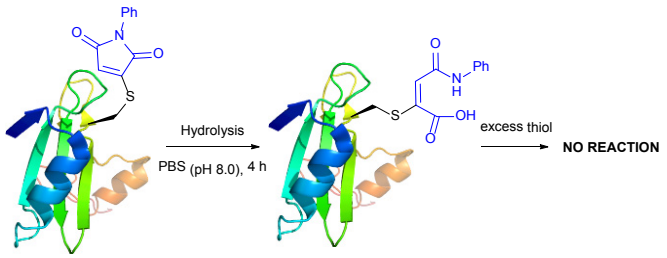


Fig. 23. "Switching off" the linker's reactivity for second thiol addition achieved by hydrolysis of the thiomaleimide linkage.[316]

1.3.4 Vinyl sulfones

The Michael-type addition of vinyl sulfones (VS) is an attractive strategy for protein conjugation, because of the elevated water stability of VS function and almost quantitative yields of their reaction with thiolates.[318-322] The reaction of vinyl sulfones with lysine residues has been reported,[319,321] however, occurring only at high pH values (pH > 9.3).

Initially, VS-mediated approaches have been used almost exclusively for PEGylation of proteins with end-functionalised PEG derivatives.[321,323] Several studies on the immobilisation of macromolecules on solid supports using vinyl sulfones were reported, owing to the elaboration of new methods for the preparation of VS-modified surfaces.[318,319,324-326] Versatile VS-containing probes, namely carbohydrates,[223,327] chelating agent,[328] fluorescent tags,[329,330] and biotinylation reagents[330] were recently developed and applied in bioconjugation of proteins. Ovaa and co-workers[331] used vinyl sulfone handle to conjugate enzymes to a ubiquitin-like protein. The applications of VS-tags in proteomics have recently gained popularity and been reviewed by Lopez-Jaramillo *et al*.[330]

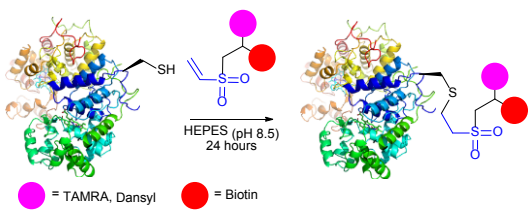


Fig. 24. Bifunctional labelling of horseradish peroxidase (HRP, pdb: 2ATJ) with vinyl sulfone-functionalised tags described by Morales-Sanfrutos *et al*.[329]

Vinyl sulfones react with thiols to form a stable thioether linkage to the protein at slightly basic conditions (pH 7–8).[224] The reaction may proceed faster if the pH is increased, but this usually also increases the amount of side-products. The main advantage of VS-tags is their elevated stability in aqueous solutions, compared to more reactive towards thiols maleimides, which can be subjected to ring opening or addition of water across the double bond. [181]

1.3.5 Thiol-ene coupling

Discovered at the beginning of the last century by Posner,[332] free-radical-based hydrothiolation of terminal alkenes, also called the thiol-ene coupling reaction (TEC), has emerged as a

powerful approach for the chemoselective modification of both peptides and proteins.[333,334] The initial step of the reaction is light- and/or initiator-induced generation of thiyl radical. This adds to alkene in an anti-Markovnikov fashion to yield thioalkyl radical. This leads to the propagation of the radical chain by abstraction of hydrogen from the other thiolate (Fig. 25).

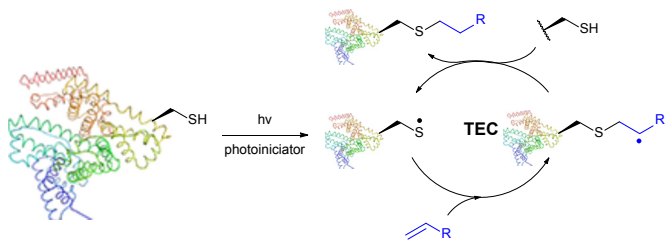


Fig. 25. Mechanism of photoinduced free-radical thiol-ene coupling (TEC).

The TEC conjugations (usually conducted in PBS-DMSO buffer at pH 7.0-7.5) are compatible with oxygen and aqueous media and are usually carried out upon irradiation at λ_{max} 365 nm in the presence of Vazo44 (2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride) as an initiator. The resulting thioether linkage is biologically stable and robust.

The first approach to protein conjugation namely glycosylation *via* TEC was reported by Davis *et al*. in 2009.[335] However, it consisted in photoinduced coupling of various glycosyl thiols with site-specifically introduced unnatural L-homoallylglycine. A complementary approach to peptide and protein glycoconjugaiton by photoinduced coupling on cysteines was first introduced almost at the same time by Dondoni and co-workers.[336] The 66 kDa globular bovine serum albumin (BSA) possessing one free CYS-34 residue was selected as a model protein. Surprisingly, it revealed that not only the one CYS-34 SH group, as expected, but also two more SH groups arising from 75↔91 disulfide bond were modified. It was suggested that such hyperglycosylation was due to well-documented disulfide bond degradation by UV-irradiation,[337] namely to an electron transfer process from photoexcited tryptophan residues. Furthermore, prolonged irradiation of the reaction mixture up to 2 hours induced the introduction of seven glycoside residues into BSA. Despite the necessity for UV-irradiation and ensuing side-reactions result troublesome in some instances, the fact that, in contrast to the majority of thiol-selective methodologies, TEC does not exploits elevated nucleophilicity of the thiolate but its readiness for the generation of radicals, makes it especially tolerant to a wide range of functional groups. For instance, Garber and Carlson[338] have used this feature of TEC for selective capping of thiols in presence of thiophosphorylated groups, free alcohols and amines.

Several approaches involving the combination of cysteine-selective methodologies have been recently reported. Stolz and Northrop[339] studied the reactivity of *N*-allyl maleimides and found this scaffold to be appropriate for consecutive two-step conjugation of thiols: *via* first base-initiated Michael-addition to maleimide moiety and second radical-mediated TEC of allyl-fragment. Scanlan and associates[340] developed the sequential NCL-TEC approach (for more details on NCL, see Section A 2.3.1, page 33) for the functionalisation of the cysteine thiolate generated at the ligation site during native chemical ligation.

1.3.6 Thiol–yne coupling

After the rise of TEC for bioconjugation, its sister reaction of the hydrothiolation of alkynes, also referred to as thiol–yne coupling (TYC), began to receive increased attention.[341] Discovered in 1949 by Jones and collaborators,[342] TYC allows introduction of two thiol fragments across a carbon–carbon triple bond *via* a free-radical mechanism similar to TEC. The first step of the anti-Markovnikov addition of a thiyl radical to the triple bond yields an intermediate vinyl thioether capable of undergoing a second addition of thiyl radical through the same mechanism, leading to the 1,2-dithioether (Fig. 26).

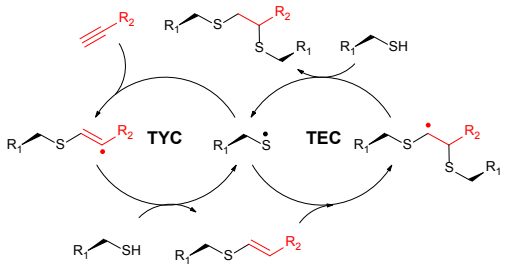


Fig. 26. Mechanism of photoinduced free-radical thiol-yne coupling (TYC).

TYC occurs under the same reaction conditions as TEC and smoothly proceeds at 25 °C in aqueous solutions. First trials on the applicability of TYC on peptides were reported by Dondoni and collaborators in 2010.[343] The authors have demonstrated the possibility of dual glycosylation of a series of peptides (up to 8 residues). Later on, Davis and Dondoni have expanded dual conjugation strategy for achieving sequential glycosylation and fluorescent labelling of BSA (Fig. 27).[344] Just as with TEC, the reaction also occurs at cysteine residues of the 75↔91 disulfide bridge.

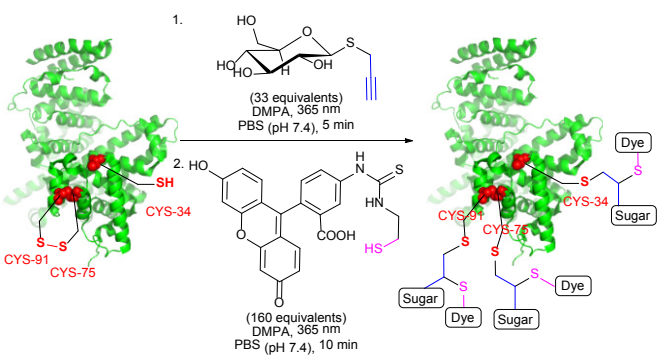


Fig. 27. Glycosylation and fluorescent labelling of bovine serum albumin (BSA, pdb: 3V03) by TYC (first step) followed by TEC (second step) with 2,2-dimethoxy-2-phenylacetophenone (DMPA) as photoiniciator.[344]

The necessity of using a photo- or a chemical iniciator in both TEC and TYC conjugations represent the main drawback of these methodologies, for the presence of free radicals results in a series of side reactions, namely oxidation and crosslinking of proteins.

1.3.7 Disulfide reaction

Simple air oxidation of two thiolates to form a disulfide bond is probably the most straightforward among cysteine-selective conjugation techniques. Very simplistically, it consists in open to air stirring of a protein possessing a free cysteine residue with a thiol-containing probe for several days under basic conditions.[345] Apparently, a large excess of the thiolate-probe is required in order to reduce the likelihood of the protein dimerisation. Treatment with iodine[346] was reported for the activation of cysteine formation of mixed disulfides. However, restricted control of product distribution and long reaction times largely limits the applicability of these methods for bioconjugation.

Diverse disulfides have been extensively used in the past decade for the modification of cysteine by disulfide exchange. This reversible reaction involves attack of cysteine thiolate at the disulfide, breaking the S-S bond, and subsequent formation of a new mixed disulfide. A well-known example of such reaction is colorimetric quantitation of free sulfhydryls with Ellman's Reagent.[347] Several symmetric disulfide-containing fluorescent probes such as BODIPY L-cystine and Fluorescein L-cystine are commercially available. However, because there is no thermodynamic preference for this disulfide exchange to pass one way or another, the labelling with non-activated disulfides generally requires use of a large excess of the probe to achieve sufficient levels of tagging.[348] In contrast, related activated thiols, namely thiosulfates ($R\text{-}S\text{-}SO_3^-$), thiosulfonates ($R\text{-}S\text{-}SO_2\text{-}R'$, MTS), sulfenyl halides (R-S-X),[349] pyridyl disulfides,[350,351] and TNB-thiols (derivatives of 5-thio-2-nitrobenzoic acid) contain good leaving groups in their structures and thus no excess of the reagent is required to drive the equilibrium (Fig. 28).

PEGylation, fluorescent and biotinylation probes containing a thiosulfate (commercialised as TS-link reagents) and pyridyl disulfides motifs are today widely commercially available. Thiosulfonates were the first introduced for bioconjugation by Davis and co-workers[352,353] in their work on the controlled glycosylation[354] and further elaborated by Zhao *et al.*[355] for site-selective PEGylation of proteins. Recent advances resulted in further development of thiol activated methodologies towards selenenylsulfides[356,357] and even methanedithiosulfonates, allowing for synthesizing of trisulfide conjugates.[358] Disulfide-based conjugation was recently reported for the preparation of antibody-drug conjugates and studying the influence of spacer length on their stability.[359] Diselene analogues of disulfide PEGylation reagents were proposed by Jevsĕvar *et al.*[360] as selective and fast alternative for coupling. Although, high conversion yield required the use of a large molar excess of the probe, this elegant approach represents an interesting technology which deserves further investigation.

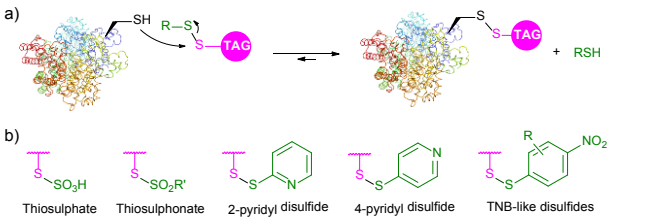


Fig. 28. **a)** General scheme of protein labelling *via* the formation of disulfide bond. **b)** Most relevant examples of activated thiols: thiosulfate, thiosulfonate, 2-pyridyl disulfide, 4-pyridyl disulfide, TNB-like thiol.

The main factor that has gained popularity to methodologies yielding disulfide and selenol-sulfide linkage is the reversibility they afford. Treatment of obtained conjugates with classical reducing agents such as DTT, β-mercaptoethanol or TCEP generally results in their complete cleavage. Yet, in the case of disulfides, the modified protein can be turned to a more stable and resistant to reduction corresponding thioether-linked conjugate by means of HMPT-mediated desulfurisation elaborated by Davis and associates (Fig. 29).[361]

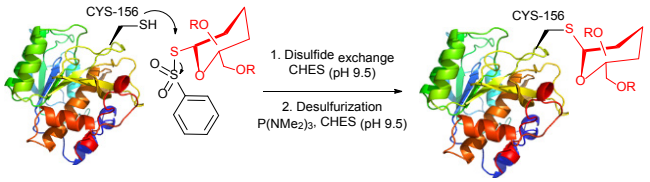


Fig. 29. Two-step protocol for the preparation of thioether linkage *via* disulfide exchange reaction followed by HMPT-mediated desulfurisation ($P(NMe_2)_3$, hexamethylphospotriamide) of glycosylated mutant S156C SBL (wild-type pdb: 1GCI).[361]

1.3.8 Disulfide re-bridging

Use of thiol reactive reagents often requires the necessity for recombinant introduction of a free cysteine into the protein[362] because most proteins do not have a free cysteine.[363,364] This new free unpaired cysteine may cause disulfide scrambling, complicate protein refolding,[362] or lead to aggregation of the protein.[365] In contrast, most of biologically relevant proteins possess at least one disulfide bond in their structure.[366] The direct reduction of disulfide bonds followed by conjugation with thiol-selective reagents is, however, usually inadmissible, since these are responsible for its structure, stability, or function[367,368] and must thus remain bridged after the modification in order not to alter protein tertiary structure.

In an attempt to resolve this problematic, Brocchini *et al.*[369] developed a clever methodology for the PEGylation of protein disulfide bonds with α,β-unsaturated bis-thiol alkylating reagents. Covalent re-bridging of the two thiols derived from the disulfide after its mild reduction allowed obtaining the modified proteins with retained tertiary structure and biological activity. Interferon α-2b (IFN) was used in the initial studies, because it

is representative of four-helical-bundle proteins with accessible disulfide bonds. Following reduction of the disulfide in IFN's, the two free cysteines were re-joined using a three-carbon linked functional PEG.[366,369] The methodology was further expanded towards PEGylation of therapeutic proteins[366,370,371] fragment antigen-binding fragments of Immunoglobulin G,[372] and poly phosphocholine labelling of IFN.[373] Simultaneously with the introduction of previously mentioned monobromomaleimides, Baker and co-workers have introduced a relevant class of reagents containing highly reactive dibromomaleimide or dibromopyridazonedione scaffold, allowing rapid and efficient disulfide re-bridging by installing a rigid two-carbon linker.[315,374] This approach was first applied for equimolar PEGylation of 32-amino acid salmon calcitonin (sCT, Fig. 30)[375] and very recently for the preparation of homogeneous antibody conjugates.[376,377]

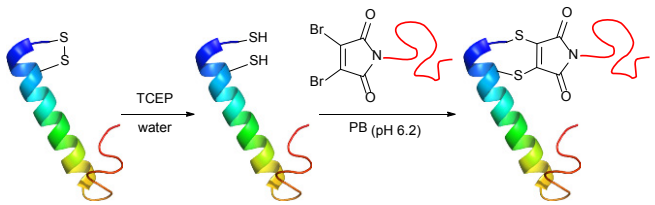


Fig. 30. One-pot reduction-PEGylation of the disulfide bridge of sCT (pdb: 2GLH) followed by rebridging using dibromomaleimide.[375]

Although being very rapid (full conversion is achieved in less than 5 minutes), dibromomaleimide-based conjugation resulted in obtaining of a small amount of the multimers while modifying complex polypeptides.[376] Developed by Baker *et al.*[378] and Haddleton *et al.*[379] more stable and less reactive dithiophenolmaleimides allowed avoiding this apparent drawback of dibromomaleimide probes. In combination with benzeneselenols known for their efficiency in catalysis of disulfide cleavage, dithiophenolmaleimide approach allowed selective antibody fragment conjugation with no detectable formation of multimers and conserving high reaction rate (Fig. 31).[376]

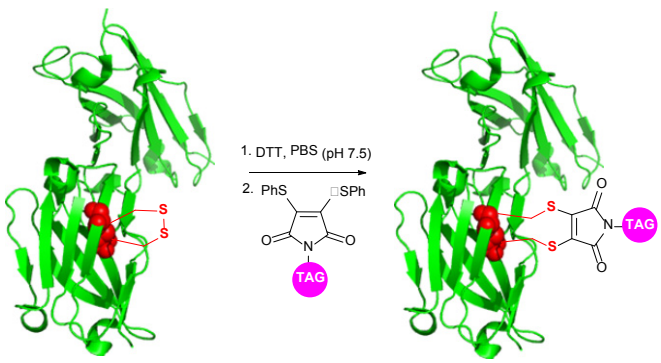


Fig. 31. Dithiophenolmaleimide approach for *in situ* disulfide bridging of antibody fragments.[376]

1.3.9 Transforming to dehydroalanine

β-Elimination of thiolate from cysteine moiety turns this one of the strongest nucleophilic side chain into a dehydroalanine moiety (Dha, Fig. 32) representing an electrophilic centre for reactions with nucleophiles. Such "umpolung" in terms of nucleophilicity-electrophilicity opens an extremely interesting prospect for the transformations, which applicability is generally largely restricted due to a general nucleophilic nature of amino acid side chains present in proteins. Among them such methodologies as Pd and Rh catalysed reactions,[380-383] alkylation, [384,385] hydroboration,[386] and perhaps the most useful serving as a Michael acceptor for thiol nucleophiles.

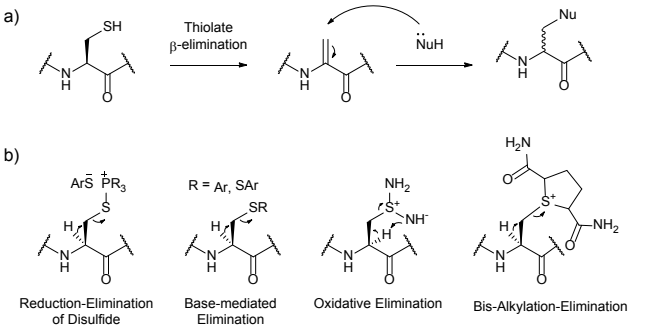


Fig. 32. Utilisation of Cys-Dha transformation for bioconjugation. a) General mechanism of the transformation. b) Representative methods for the generation of dehydroalanine.[387]

Site-selective incorporation of Dha into proteins may be achieved by a considerable number of chemical transformations. Historically, the first example of such reaction was reported by Koshland and collaborators 50 years ago and consisted in transformation of nucleophilic serine of chymotrypsin to dehydroalanine *via* selective sulfonylation followed by base-mediated elimination.[388-390] The method, however, exploited the particularly high nucleophilicity of SER-195 residue located in the active site of chymotrypsin. Logically, more general methods based on the exceptional nucleophilic properties of cysteine received increased attention in the years to follow. These are represented in the Fig. 32 and include: reduction-elimination, representing an often observed undesired side-reaction during reduction of disulfides; base-mediated elimination of activated thiolate, typically requiring temperatures incompatible with protein substrates and thus not being of synthetic interest; oxidative elimination, and bis-alkylation-elimination.[387] Two last approaches seems to us the most promising among available Cys→Dha transformations and it is to these methodologies that we now turn.

1.3.9.1 Oxidative elimination

Oxidative elimination of thiolates is readily achievable, but often required high temperatures and severe reaction conditions yielded the methodology incompatible with fragile protein substances.[391] However, recent efforts focused on finding milder conditions to carry out these desulfurisations, have resulted in developing two promising classes of chemical reagents: *o*-mesitylenesulfonylhydroxylamine (MSH),[384,392] and bromomaleimides (above-mentioned for the generation of bis-cysteine adducts).[313]

Basic conditions are generally required for the reaction to achieve high conversion yields. Other amino acid residues may also react with MSH and bromomaleimide, but the reaction rates are largely inferior to those of thiolates and resulting products are generally unstable in basic conditions and decompose back to their starting unmodified forms.

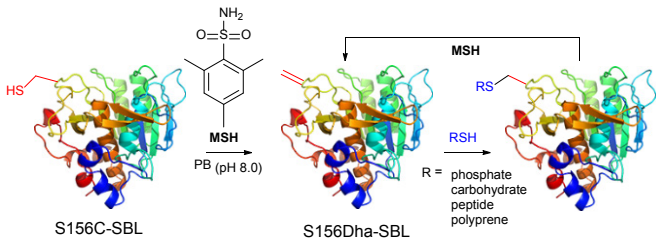


Fig. 33. Conjugation of mutant S156C SBL (wild-type pdb: 1GCI) containing a single, surface-exposed cysteine residue CYS-156 by oxidative elimination followed by conjugation with thiol probes.[384]

1.3.9.2 Bis-alkylation-elimination

Conversion of cysteine to dehydroalanine by means of bis-alkylation-elimination approach was first introduced by Holmes and Lawton.[393] Initially this transformation implied quite strenuous reaction conditions and was only compatible with a restricted number of protein substrates. Only recently have Davis and collaborators[387] reported a more general method for Cys→Dha transformation by means of water-soluble α,α'-dibromoadipyl(bis)amide (DBAA) allowing generation of Dha moiety under mild conditions (37 ˚C, pH 7.0-8.0) at sufficiently high yields. This approach was evaluated on several model proteins, including SBL (see above),[387] the single-domain antibody cAb-Lys3 A104C mutant,[387] histone H3 (Fig. 34),[385] AurA kinase domain,[394] and GFP mutant.[395]

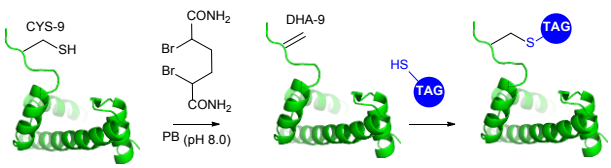


Fig. 34. Dehydroalanine-mediated conjugation on histone 3 mutant H3.[385]

1.3.9.3 Other approaches

Despite recent advances in dehydroalanine-mediated conjugation methodologies, the inherent limitations of these methods preclude their general use for peptide and protein modification. None of these approaches enable general, chemo- and site-selective incorporation of dehydroalanine into proteins without the need for prior incorporation of an accessible Cys residue. Several other approaches continue to appear and are designed to overpass this problematic. These are oxidative elimination of aryl-selenocysteine,[391,396-400] utilising of lacticin synthetase,[401] transformation to selenocysteine thioethers.[402]

1.3.10 Miscellaneous thiol-selective reagents

An example of a simple alkylation reaction that still remains relevant in bioconjugation is aminoethylation. Known for more than half a century,[403,404] it allows transforming cysteine thiolates into lysine mimicking thialysine residues by means of bromoethylamines or aziridines. Obtained thialysines were validated as appropriate synthetic substrates for further amine-selective transformations (see Section A 1.1, page 3). Furthermore, the method was recently demonstrated appropriate for providing the access to more peculiar methylated lysine analogues.[405] The reaction is typically conducted at pH > 8.5 to ensure a high level of cysteine deprotonation.

S_NAr substitution chemistry approaches for cysteine modification in proteins were reported by several research groups.[387,406,407] Davis *et al*.[387] utilised Mukaiyama's reagent (2-chloro-1-methylpyridinium iodide) to generate an arylated cysteine as an intermediate for conversion thereof to dehydroalanine. Pentelute and co-workers have expanded the approach towards perfluoroaryls for protein stapling and conjugation.[406,407] Finally, Barbas and associates[408] have developed a class of Julia–Kocienski-like methylsulfonyl-functionalised reagents, that reacts rapidly and specifically with thiols under biologically relevant pH (5.8-8.0).

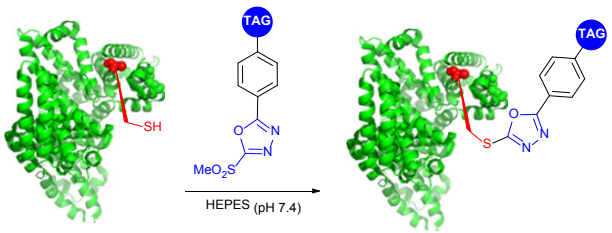


Fig. 35. Labelling of the free cysteine residue of BSA (pdb: 3V03) with Julia–Kocienski-like reagents.[409]

Notably, the resulting conjugates possess superior stability compared to cysteine–maleimide conjugates in human plasma, which makes this methodology appropriate for the preparation of stable protein conjugates and PEGylated proteins (Fig. 35).

An efficient gold-catalysed allene-mediated coupling reaction has been recently developed by Che and colleagues.[410] The method allowed direct thiol-selective functionalisation of model peptides and reduced RNase A (Fig. 36).

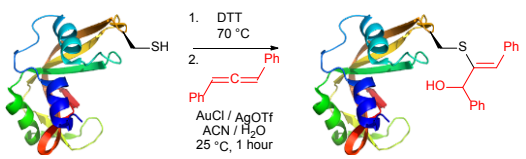


Fig. 36. Cysteine modification of RNase A (pdb: 7RSA) *via* gold-mediated coupling with allenes.[410]

Reactions of thiols with electron deficient acetylenes have been known for decades, being, however, mostly conducted in or-

ganic solvents.[411-415] Several examples of reactions in aqueous media have been recently reported.[409,410,416,417] Che and coworkers[409] have elaborated a versatile method for the selective cysteine labelling of unprotected peptides and proteins in aqueous media with arylalkynone reagents. Notably, modified peptides could be converted back into the unmodified peptides by treatment with thiols under mild reaction conditions (Fig. 37).

Oxanorbornadienedicarboxylates (OND reagents), strained adducts of furans and electron-deficient alkynes, were found to provide better water stability while retaining selective, rapid, and fluorogenic reactivity towards cysteine compared to corresponding alkynes.[418]

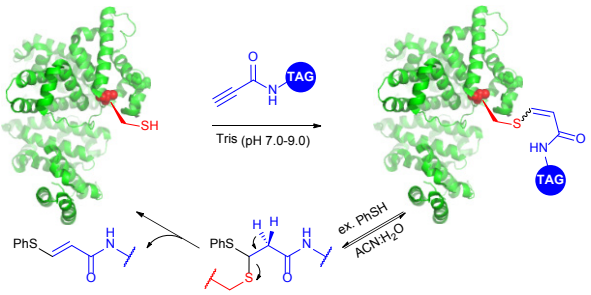


Fig. 37. Cleavable labelling of BSA (pdb: 3V03) with arylalkynone reagents elaborated by Che and associates.[409]

α,β-Unsaturated ketones and amides (typically acrylamides) can undergo Michael-addition.[419,420] However, the rate of addition is not generally high enough to provide it with competitive advantages compared with other approaches. Internal Cys residues were reported to accelerate native chemical ligation (see Section A 2.3.1, page 33), an especially selective approach for *N*-terminal cysteine conjugation, *via* cyclic transition states.[421-426]

1.4 Tryptophan

Tryptophan (Trp, W) is the second (after cysteine) low abundance amino acid with about 1% frequency (depending on the living organism),[234] but approximately 90% of proteins contain at least one Trp residue in their sequence.[427] The specific reactivity of tryptophan in proteins is one of the most challenging problems in bioconjugation. In spite of the variety of reagents introduced over the years for selective modification of tryptophan, only a few can be used for conjugation. For instance, such classically used species as Koshland's reagent (2-hydroxy-5-nitrobenzylbromide)[428] or chlorosulfonium ions[429] present a high degree of cross reactivity with nucleophilic side chains, nonetheless still being used in numerous studies. These are, for example, investigation of the role of tryptophan in active sites of enzymes,[430] estimation of its content in proteins[431], or determination of the surface accessibility of Trp residues in proteins.[432]

1.4.1 Malondialdehydes

In 2007, further exploring the reactivity of dicarbonyl compounds towards tryptophan described by Teuber and colleagues[433] 40 years earlier, Foettinger *et al.*[434] have reported selective reaction of substituted malondialdehydes (MDAs) with the indole nitrogen of the Trp side chain of 8-mer peptide PTHIK**W**GD under acidic conditions (Fig. 38). Obtained substituted acrolein moiety with a remaining reactive aldehyde group, can be further converted to a hydrazone using hydrazide compounds, or using other methodologies for aldehyde conjugations. Hydrazine, phenylhydrazine and secondary amines such as pyrrolidine were reported to act as cleavage reagents and allow releasing of the free tryptophane after conjugation.

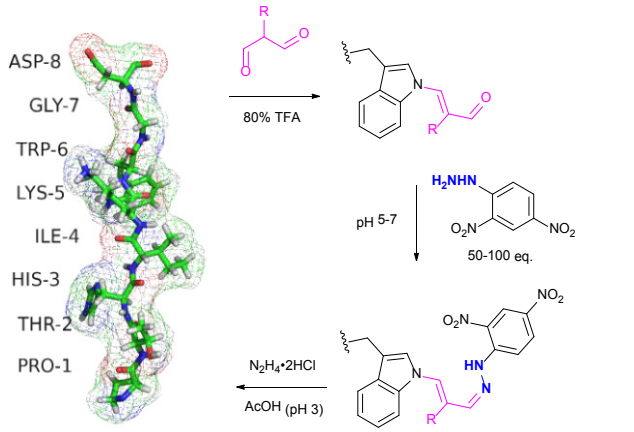


Fig. 38. Selective labelling of Trp side-chain in a 8-mer peptide PTHIK**W**GD with malondialdehyde described by Foettinger *et al.*[434] The peptide structure was simulated using RaptorX web server.[435]

To overpass selectivity issues, namely a known side reaction with Arg side chains,[436] the conditions for the reactions with MDA, the hydrazone formation and the cleavage of the MDA derivative, had to be optimised concerning pH, buffer, temperature, and reagent. No side reactions of MDAs were observed only under strongly acidic conditions, such as aqueous TFA (80%). The following hydrazone formation requires approximately 50-100 fold molar reagent excess at a pH of 5−7 and sometimes increasing the temperature to 50 °C. Although unstable at acidic conditions and when the reagent excess is removed, the hydrazone bond remains firm in alkaline medium (pH > 9). The optimal conditions for the cleavage were found using hydrazine (applied as the dihydrochloride salt) in ammonium acetate solution at a pH ~ 3. Demonstrably, these rather rough reaction conditions prevent this methodology from finding widespread use for sensitive protein targets, yet allowing its application in proteomics on peptide digests.[437]

1.4.2 Metallocarbenoids

The same drawback is shared by another approach involving vinyl metallocarbenoids described in 2004 by Francis and collaborators.[438] The authors have shown, that two Trp residues of horse heart myoglobin can be selectively tagged by a stabilised vinyl diazo compound in the presence of $Rh_2(OAc)_4$ (Fig. 39). Difficulties reminiscent to the known instability of the rhodium carbenoid intermediate in aqueous media[439] were overcome by using an unusual additive - hydroxylamine hydrochloride – that

was found to facilitate the reaction and enhance efficiency of the tryptophan modification pathway relative to hydrolysis of metallocarbene.

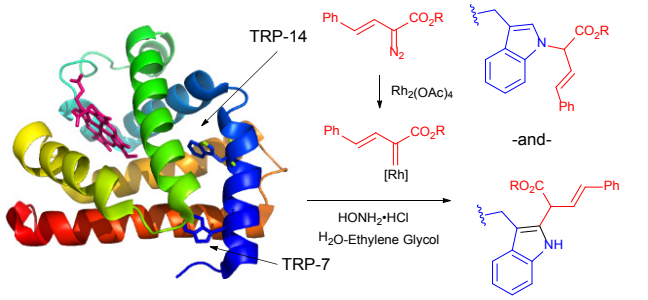


Fig. 39. Modification of horse heart myoglobin (pdb: 1YMB) with rhodium carbenoids described by Francis *et al*.[438] A 100 μM solution of myoglobin was exposed to stabilised vinyl diazo precursor (10 mM) and $Rh_2(OAc)_4$ (100 μM) for 7 h. *N*- and *C*-derivatisation of indole rings of both Trp residues - TRP-14 and TRP-7 - were identified by the mass reconstruction. An excess of hydroxylamine hydrochloride (75 mM) is crucial for the efficiency of the conjugation, although its mode of action was not elucidated.

The initially reported reaction conditions tolerated several aqueous solvent systems and proceeded at 25 °C. Yet, acidic conditions (pH 1.5−3.5) were still necessary for efficient protein labelling and stood out as the main drawback preventing this approach from being generally applicable. For instance, in the same work, authors have stated that myoglobin was denatured and the heme dissociated from the protein due to the high acidity of the medium.

To address these limitations, following efforts of the same group were to improve the pH range of the tryptophan modification methodology.[440] For hydroxylamine was found to be ineffective at generating rhodium carbenoids at pH ≥ 6, a wide screening of commonly used buffers, as well as additives structurally similar to H_2NOH, was conducted in order to identify appropriate conditions. From these studies, *t*BuNHOH was found to be highly effective at promoting carbenoid addition. Despite the precise mode of action for *t*BuNHOH remains unclear, the authors attributed the substantial increase in catalytic activity to a specific interaction between this additive and $Rh_2(OAc)_4$. They speculated that, in contrast to hydroxylamine, *t*BuNHOH binds to $Rh_2(OAc)_4$ through the oxygen, rather than the nitrogen, the latter being disfavoured by the bulky *tert*-butyl substituent (Fig. 40a), and increases both the stability and the reactivity of the complex at neutral and slightly basic pH.

Interestingly, in the same work, the authors have demonstrated the key role of solvent accessibility of residues in determining the outcome of conjugation on tyrosine using rhodium metallocarbenes. Human FK506 binding protein (FKBP) was identified as a suitable substrate for the study. The only Trp residue (TRP-59) of a wild type FDBP (containing an additional *C*-terminal threonine residue) is located at the base of the binding pocket, and therefore is unavailable for modification under nondenaturing conditions (Fig. 40b).

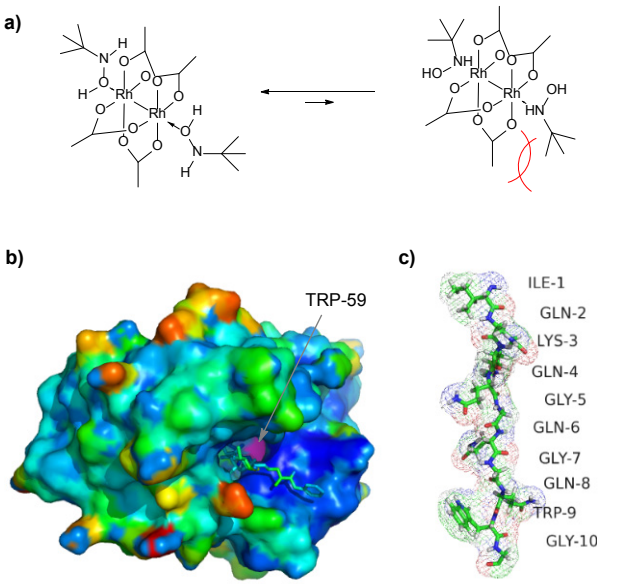


Fig. 40. Optimised metallocarbenoids-based approach, described by Francis and collaborators.[440] **a)** Proposed binding of *t*BuNHOH with $Rh_2(OAc)_4$ at pH 6.0. *O*-coordination is favoured due to a lowed sterical hindrance of *tert*-butyl groups with acetate ligands. **b)** Crystal structure of wild type FKBP (pdb: 1A7X), containing a single, buried tryptophan residue TRP-59 (shown in magenta), which is unavailable for modification under nondenaturating conditions. **c)** A peptide tag (IQKQGQGQ**W**G) incorporated into fusion FKBP protein expressed in *E. coli* as *C*-terminal intein fusion. The total level of modification was estimated to be in excess of 40%.

To overcome these difficulties, a labelling strategy based on tryptophan mutagenesis followed by chemoselective modification with rhodium carbenoids was utilised. Tryptophan-containing FKBP proteins were expressed in *E. coli* with *C*-terminal intein fusions containing a chitin binding domain for affinity purification and short tryptophan-containing peptides (Fig. 40c). Indeed, these newly obtained mutants with solvent accessible Trp residues showed significant level of conjugation (more than 40%) under optimised non-denaturating conditions at 25 °C.

In 2010, further developing the rhodium-carbenoids methodology for selective tryptophan labelling described 6 years before by Francis *et al*.,[438] Popp and Ball have reported structure-selective modification of aromatic side chains (expanding its scope to include Tyr and Phe residues) using proximity-driven approach (see Section A 5, page 42 for details).[441] Structurally similar to $Rh_2(OAc)_4$, metallopeptide complexes with a dirhodium center bounded with two glutamate residues were envisioned to provide delivering of the catalyst to a close proximity of the reactive side chains by exploiting the coiled coil matched peptides,[442] for molecular peptide-peptide recognition.(Fig. 41)

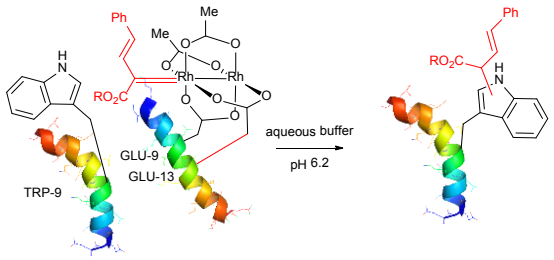


Fig. 41. Selective covalent labelling of TRP-9 residue of the peptide QEISALEK**W**ISALEQEISALEK with its complementary dirhodium metallopeptide KISALQKQK**E**SAL**E**QKISALKQ described by Popp and Ball.[441] Rhodium cluster, chelated with two glutamate residues, is brought to a proximity of the reactive Trp residue by peptides coils self-assembling, resulting in selective peptide modification on TRP-9. Peptide structures were simulated using RaptorX web server.[435]

By combining residue-selective chemistry with secondary-structure recognition, the authors have provided a strategy for selective covalent modification of biomolecules. However, only simple diazo reagents without functional handles were used in controlled environments on model peptide substrates.

In the following year, the same group has extended their initial studies to examine the reactivity of whole proteins in a complex, cell-like environment.[443] For this, the proximity-driven catalysis approach was applied to a recombinant maltose binding protein (MBP), fused with the 21-amino-acid tryptophan-containing coil (almost identical to one used in the initial publication, Fig. 41). Directly after the expression, the lysate was subjected to metallopeptide-catalysed biotinylation. A single band in Western blot analysis indicated highly selective modification of the modified MBP protein with no nonselective modification to be observed.

1.5 Histidine

Histidine (His, H) is the only amino acid with a pK_a in the physiological range, hence often found in active sites of many enzymes and is of crucial importance in mechanisms where abstraction or donation of a proton is needed. Because of its pK_a value, both the acid and base forms are present at physiological pH (Fig. 42). Most of the studies on catalytic activity of enzymes and protein-protein interactions involving histidine-containing active centres were done by measuring the influence of site-specific modifications of His residues on the activity of the macromolecule. Main factors influencing histidine reactivity are (1) the pK_a of the individual His residue, (2) solvent exposure of the residue, and (3) hydrogen bonding of the imidazolium ring. A vast number of studies have been conducted to date, nonetheless, exploiting a small number of chemical functions.

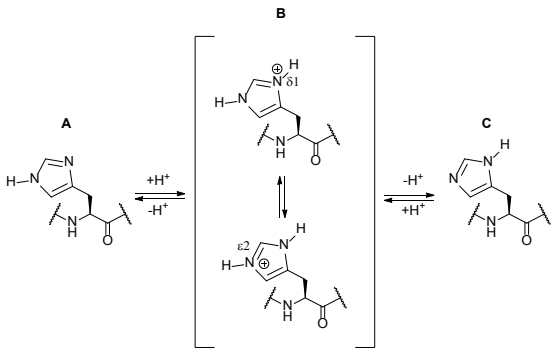


Fig. 42. Tautomerisation equilibrium of the neutral imidazole side chain (base forms **A** and **C**) occurring through the acid form **B**.[444] Form **A** is somewhat favoured over **C** at neutral and acidic pH, while at basic pH the form **C** is preferred.[445]

1.5.1 Epoxides

In 2004, Li *et al*. described two epoxide-containing fluorescent probes[446,447] for selective labelling of histidine. A high selectivity of probes towards free histidine in the presence of a 1000-fold excess of other nucleophilic amino acids[446] and in human serum[447] (Fig. 43) was achieved after 2 to 3 hours of incubation at 80 °C in NaOH solution (pH 11-12).

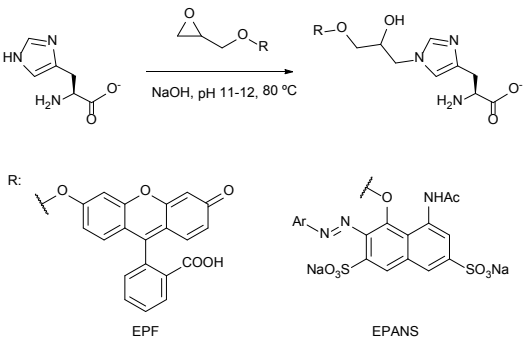


Fig. 43. Fluorescent probes, described by Li *et al.* for selective ligation of histidine.[446,447]

The authors suggest that these probes can be used for specific labelling of His residues in proteins if a mild reaction condition (lower reaction temperature but longer reaction time) was used, but any example of such application is absent.

An affinity-based labelling approach (see Section A 4, page 42) based on the epoxide opening was used by Hamachi and collaborators for selective histidine labelling of bovine carbonic anhydrase II.[448,449] Labelling reagents investigated by the authors must consist of at least three major fragments: (1) a benzenesulfonamide ligand directing specifically to bCA, (2) a reactive electrophilic epoxide for protein labelling, and (3) an exchangeable hydrazone bond between the ligand and the epoxide group for removing the ligand by hydrazone/oxime-exchange and restoring the enzymatic activity (Fig. 44a). Further developing their approach,[449] the authors added an iodophenyl or acetylene handle on the epoxide-containing fragment

to enable the possibility of further derivatisation of the obtained conjugate by Suzuki coupling[450] or Huisgen cycloaddition[451] either after or before removing the ligand from the active site of the enzyme (Fig. 44b).

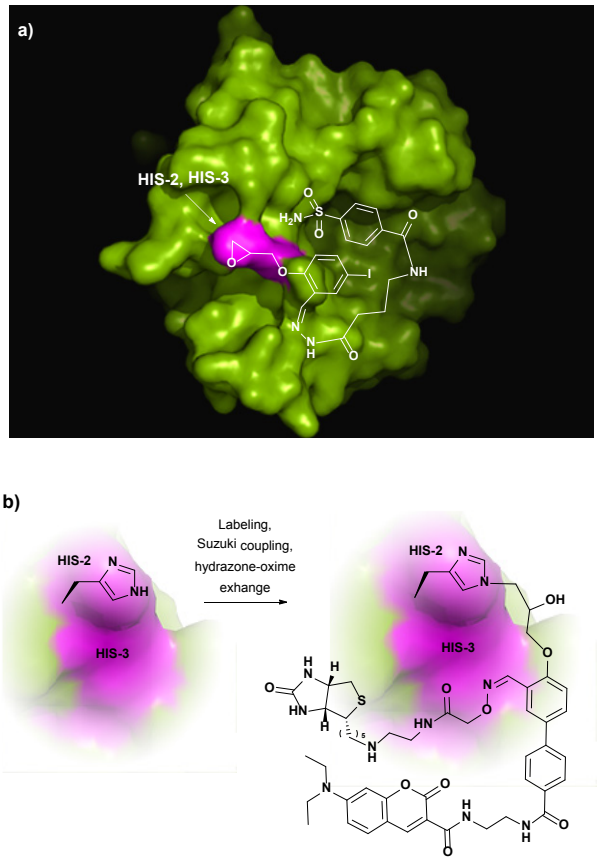


Fig. 44. Affinity labelling, hydrazone-oxime-exchange reaction, and Suzuki coupling reaction on the surface of bCA (pdb: 1V9E) described by Hamachi and collaborators.[449] **a)** Principal structural fragments of the probe: a benzenesulfonamide ligand responsible for targeting bCA; an electrophilic epoxide responsible for reactivity towards vicinal His residues HIS-2 and HIS-3 (shown in magenta), cleavable hydrazone bond, responsible for the recovering of the enzymatic activity; an iodoaryl moiety, utilised for the further bioorthogonal transformation *via* Suzuki coupling. **b)** Dual labelling of bCA *in situ*: fluorescent labelling of the HIS-2 or HIS-3 alkylated intermediate by Suzuki coupling with a coumarine derivative, followed by biotinylation by hydrazone/oxime-exchange.

1.5.2 Complexes with transition metals

The affinity of transition metal ions to histidine in aqueous solutions was known for decades.[452] Copper and nickel ions have the greatest affinity for histidine and this property is the most often used for protein purification by immobilised-metal affinity chromatography (IMAC), exploiting the synergetic coordination effect of oligo-histidine tags (see Section A 4, page 42). Meanwhile, recently reported by Wang *et al,* histidine-specific iridium(III)-probe for peptide labelling demonstrates an excellent example of selectivity based on exceptional coordination properties of only one or two His residues (Fig. 45).[453] Further exploring the advantage of the previously described iridium(III)-complex, used by Wong and colleagues for luminescent labelling of histidine-rich proteins[454], authors showed its applicability for histidine labelling in cell-imaging studies.

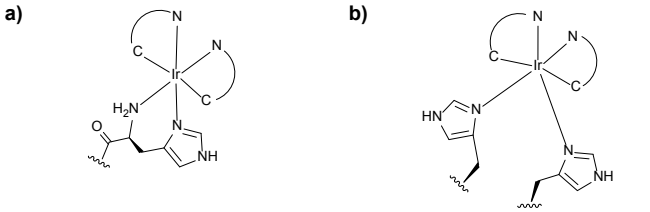


Fig. 45. Luminescent histidine selective peptide tagging.[453] **a)** Labelling of a *N*-terminal histidine-containing HTat peptide (**H**RKKRRQRRR). **b)** Labelling of a dihistidine motif placed in the middle of a P450dHTat peptide (MLAKGLPPKSVLVKGG**HH**GRKKRRQRRR).

Although, reckoning obtaining of the coordination complexes to conjugation techniques would be stretching a point, they are however included in this survey because of histidine liability to complexation and increased stability of obtained complexes to decomposition.

1.5.3 Michael addition

Several examples of histidine-selective Michael addition to carbon double bond of conjugated aldehydes – 2-alkenals – were found during studies of oxidative modification of proteins.[455,456] Even though alkenals are known for the modification of the other basic amino acid residues in the protein,[457] Zamora *et al* succeeded in achieving high selectivity level on histidine ligation in bovine albumin by incubation in PBS buffer (pH 7.4) at 37 ˚C (Fig. 46).[456] Using similar conditions – the incubation in phosphate buffer (pH 7.2) at 37 ˚C - Uchida and Stadtman[458] were able to tag selectively insulin (which contains no sulfhydryl groups) with 4-hydroxynon-2-enal. In both studies, the authors suggest only His residues are modified, but definitive evidence on this point is absent. Obtained conjugates contain active aldehyde residues and represent examples of protein carbonylation, allowing their derivatisation with aldehyde-selective reagents.[459]

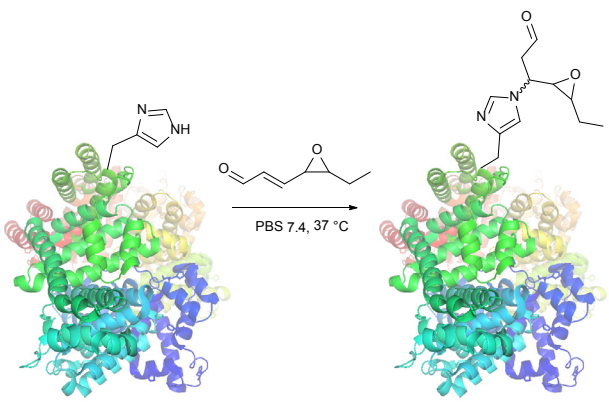


Fig. 46. Michael addition of His residues of bovine albumin (pdb: 3V03) to 4,5-epoxy-2-alkenal described by Zamora *et al*.[456] Only one HYS-536 residue is shown for the purpose of simplicity.

1.5.4 Miscellaneous histidine-selective reagents

Some examples of the selective histidine tagging by reagents that, in general, react more avidly with other nucleophilic residues were reported to date. For instance, Pramatik and colleagues achieved dominant PEGylation rh-IFN-α2B on histidine at mildly acidic levels of pH with a classic amine PEGylation succinimidyl carbonyl precursor.[460] Another group reported selective reaction of His residues of D-amino acid oxidase with dansyl chloride in 0.05 M phosphate buffer at pH 6.6.[461] These reaction conditions resulted in virtually complete inactivation of the enzyme after the reaction and its complete reactivation after the reaction with 0.5 M hydroxylamine (NH_2OH). Such reactivation excludes reaction of primary amino groups, and amino acid analysis suggested that the reaction had not occurred with an oxygen nucleophile such as serine or tyrosine. Even α-halo carbonyl compounds - phenacyl bromides, α-halo carboxylic acids and amides - classically known for their selectivity towards thiols electrophiles as were found to be histidine selective on several substrates and in carefully tuned conditions.[462-466]

Described by Pauly in the beginning of the 20th century,[467,468] the reaction with diazonium salts has been only used for colorimetric determining of histidine and was not futher elaborated for bioconjugation. Other classical examples of chemical modification of histidine are mainly reactions with pyrocarbonate, sulfonyl chlorides, sulfonic esters, phenacyl- and acylbromides and activated esters.[225] These electrophiles react readily with other nucleophilic groups presented in proteins (thiols, amines, alcohols, or guanido groups) and require a careful tuning of the reaction conditions to achieve sufficient selectivity. For instance, at low pH (generally < 6.0) these reactions are quite selective for histidine, as the main side reaction with the ε-amino group of lysine proceeds very slowly.[469]

Reader should, nonetheless, be aware, that these examples do not represent a general rule, but an exception from it. To avoid the reactions of more nucleophilic functions, the His residue must have been located in a unique microenvironment[463,464] or have an enhanced nucleophilic character,[465] but even in this case, prior modifications of highly reactive Cys residues are often inevitable.[466]

1.6 Tyrosine

Tyrosine (Tyr, Y) is one of the important amino acid residues which is known to be the active centre in many enzymes (notably tyrosine-specific protein kinases) and is used in signal transduction and cell signalling.[470,471] Occurring with intermediate to low frequency in native proteins, tyrosine is often considered as an attractive target in bioconjugation, despite being often partially or completely buried due to the amphiphilic nature of the phenolic group.

The reactivity of tyrosyl moiety is easily influenced by its deprotonation, which is a function of the microenvironment inside the protein. All described methodologies take advantage of either the peculiar chemical properties of the electron-rich aromatic ring, or the easiness of the tyrosine hydroxyl group to be transformed into highly reactive phenolate.

1.6.1 *O*-Derivatisation

Although *O*-acetylation of tyrosyl residue with acetic anhydride and *N*-acetylimidazole is arguably the most widely used technique for tyrosine modification,[225] its application for conjugation is rather limited. Mainly, these limitations are due to low selectivity of the acylation in the presence of other nucleophilic amino acid residues and modest stability of obtained conjugates.

An elegant approach – the affinity labelling – allows surpassing the selectivity issues of tyrosine acylation by ligand-tethered directing of the reaction. In the case of tyrosine-selective modification, an acyl transfer catalyst is connected to a ligand with a high affinity to the target protein.[448,472] The acyl group activated by the anchored catalyst is brought to the binging pocket of the protein and transfers an acyl moiety on the nucleophilic Tyr residue in close proximity. Utilizing this methodology, Hamachi *et al*.[472] demonstrated selective tagging of Y51 residue of Congerin II (Fig. 47) using a suitable saccharide as a ligand to the target lectin (carbohydrate-binding protein) and DMAP (4-dimethylaminopyridine) as an acyl-transfer catalyst. In a similar way, Broo and collaborators[473] have demonstrated the possibility of a site-specific acylation of a tyrosine residue situated in an active site of human glutathione transferase (hGST).

Miller and collaborators have shown that biotinylation with NHS esters (see Section A 1.1.2, page 5) may result in preferential *O*-acylation of hydroxyl-containing residues – serine, threonine and tyrosine (though to a greater extent of the first two) – when they are located two positions next to histidine (*i.e.* in sequences His-AA-Tyr, where AA refers to any amino acid).[82,85]

Several approaches for labelling involve the initial modification of tyrosine and successive conjugation of an obtained intermediate. For instance, an *ortho*-nitration of tyrosine with tetranitromethane (TNM)[474] or peroxynitrite[475] results in obtaining of *o*-nitrotyrosine that can be then reduced by sodium dithionite ($Na_2S_2O_4$) to form an *o*-aminotyrosine.

Although much less reactive than aliphatic amines at neutral pH, the aromatic amine of *o*-aminotyrosine can selectively react with amine-reactive reagents at lower pH.[477,478] Namely, Nikov *et al*.[478] have demonstrated that selective labelling of aminotyrosines is achievable in the presence *N*-terminal and *ε*-amino groups of lysines by using NHS-activated ester at particular reaction conditions (acetate buffer, pH 5.0, 2 hours). Exploitation of the pK_a difference between aminotyrosyl residues and other reactive groups in proteins (4.75 for aminotyrosine, whilst much higher values for *N*-terminal and side-chain amino groups, see Section A 1.1, page 3) allows selective labelling thereof. The method was validated on model peptides and then applied to a human serum albumin modification (Fig. 49).

Despite the reaction of TNM and peroxynitrite with proteins being reasonably specific for tyrosine, side reactions with histidine, methionine and tryptophan have been reported, as has

oxidation of sulfhydryl groups. The latter would seem to be the most common side reaction, as it can result in disulfide bond formation and the formation of oxidation products such as sulfone and sulfenic acid derivatives. As a general rule, it is normally assumed that the reaction of nitration reagents with Cys residues proceeds equally well at pH 6 and pH 8, while the reaction with tyrosine occurs at pH 8 and not at pH 6.

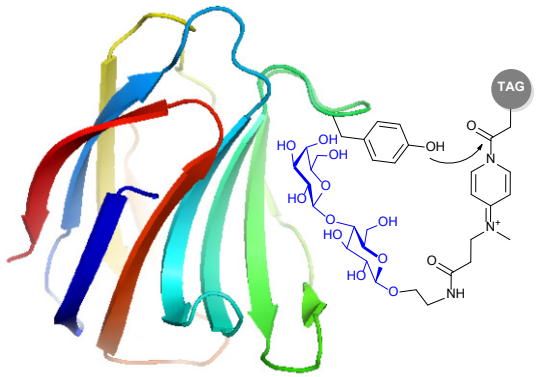


Fig. 47. Sugar-DMAP (4-dimethylaminopyridine) assisted Y51-specific acylation of Congerin II described by Hamachi *et al.*[472] Schematic representation is made using Lactose-ligated Congerin II crystal structure published by Muramoto *et al.*[476] (pdb: 1IS4).

In a like manner, the phenol group in Tyr residues can be initially *ortho*-formylated with chloroform in an alkaline medium to a salicylaldehyde derivative, and then undergo a reaction with *ortho*-phenylenediamine derivatives to form fluorescent benzimidazoles as conjugation products (Fig. 48).[479,480]
Further exploiting the methodology developed by Trost and Toste for selective *O*- and *C*-alkylation of phenols with π-allylpalladium complexes,[481,482] Francis *et al*., have demonstrated the possibility of selective allylic alkylation of surface-exposed tyrosines of several full-size proteins.[483]

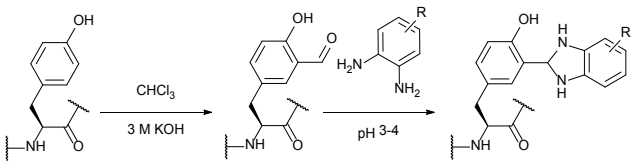


Fig. 48. Two-step tyrosine modification by selective formylation of tyrosyl residue at the ortho position of its phenolic moiety (reaction with chloroform in potassium hydroxide solution) and further derivatisation of the resulting aldehyde described by Kai *et al*. and Ishida *et al.*[479,480]

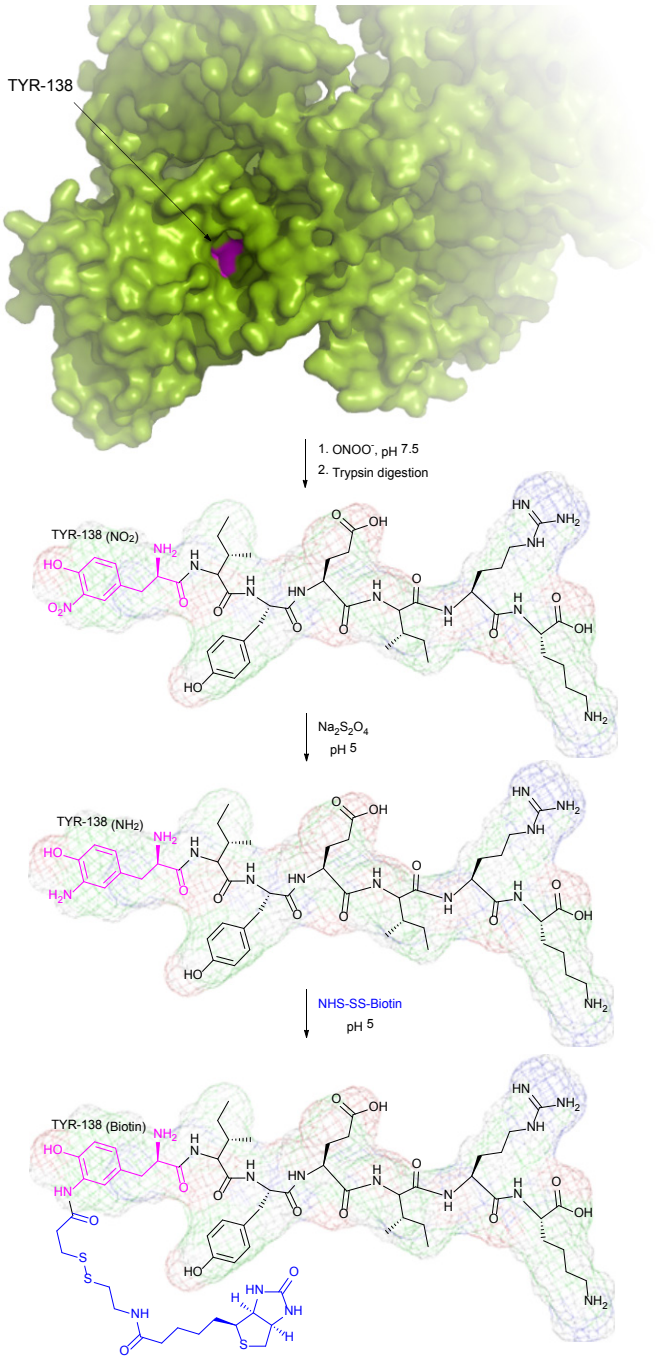


Fig. 49. Two-step biotinylation of HSA (pdb: 1AO6) described by Nikov *et al.*[478] **a)** Preferential nitration of TYR-138 (shown in magenta) and TYR-411 residues of HSA with peroxynitrite was achieved using protocol described by Jiao and colleagues.[475] **b)** Aminotyrosine of the peptide ^{138}**Y**IYEIARK^{144}, obtained after the step of reduction with sodium dithionite and following digestion of nitrated HSA with trypsin, were selectively modified with a cleavable biotin-containing reactant at pH 5.0.

1.6.2 Oxidative coupling

In 1995, Kodadek and co-workers[484] reported that Ni(II) complexed with a Gly–Gly–His (GGH) was chemically activated by oxone ($KHSO_5$) or magnesium monoperoxyphthalate (MMPP). Ni(III) oxo intermediate is hypothesised to promote protein cross-linking.[485] Other peptides, notably His_6[486,487] or the entire ribonuclease A protein,[488] which can be incorporated into proteins of interests at the genetic level, have been shown to be effective ligands for nickel catalysed oxidative cross coupling.

The photoactivable metal-catalysed version of tyrosine oxidation chemistry is significantly faster than the one achieved by Ni(II)-peptide complexes. It has been largely exploited by Kodadek and co-workers for cross-linking of closely associated proteins.[489-491]

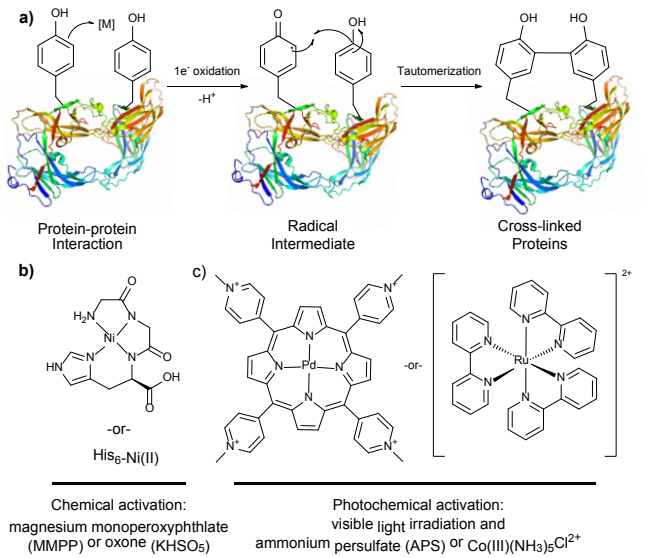


Fig. 50. Transition metal catalysed oxidative cross-linking. **(a)** Schematic representation of the protein-protein cross-linking, described by Burlingame *et al*.[492] The tyrosyl radical generated from ecotin (pdb: 1ECZ) D137H mutant after the abstraction of an electron with Ni(II)-GGH complex reacts with additional tyrosine residue on nearby protein that are in close proximity due to a significant protein–protein interaction to give a dimer. **(b)** Chemically activated by MMPP or oxone as stoichiometric oxidants, Ni(II)–peptide complexes can be used as efficient catalysts of cross-linking. **(c)** Photochemically activated by visible light in the presence of $(NH_4)_2S_2O_8$ or $Co(III)(NH_3)_5Cl_2^+$ as electron acceptors, transition metal complexes are used as catalysts of cross-linking.

These coupling reactions are hypothesised to occur through the addition of tyrosyl radicals to adjacent Tyr residues (Fig. 50). It is worth mentioning, that in some cases nearby tryptophan and other nucleophilic side chains can also participate in oxidation.[489] For more details, the reader is directed to a review by Bonnafous and a publication of Francis that provide an excellent overview of oxidative cross-linking techniques.[493,494]

The use of cerium(IV) ammonium nitrate (CAN) – a classical one-electron oxidant – for chemoselective ligation on tyrosine was demonstrated by Francis *et al*.[495] After the optimisation of reaction conditions, the authors could achieve modification of tyrosine-containing proteins with high yields at neutral pH and low substrate concentration and applied this strategy to modify both native and introduced residues on proteins with polyethylene glycol (PEG) and small peptides, although dealing with the concurrent reaction of Trp residues.[495]

Notwithstanding the issue of specificity, photo-oxidation and oxidation of techniques of tyrosine ligation continue to be of considerable interest for the study of protein-protein interactions,[496] mapping multi-protein complexes,[497] or assembling of macromolecules.[498]

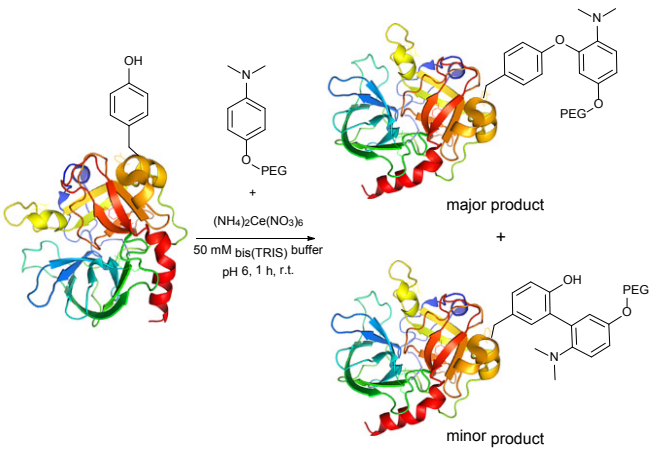


Fig. 51. PEGylation of native tyrosine residues on chymotrypsinogen (pdb: 1EX3) reported by Francis *et al*.[495] (only one Y171 out of four solvent-accessible tyrosine residues is shown). Intermediate tyrosyl radical, generated in the presence of cerium(IV) ammonium nitrate (CAN), gives two addition products with electron-rich aniline derivative (*O*- is preferred over *C*-arylation).

1.6.3 Diazonium reagents

Diazonium reaction of tyrosine has been of special interest ever since its introduction by Pauly in 1915.[468] In 1959, inspired by these pioneering efforts, Higgins and Harrington[499] advanced the use of this methodology and tried applied it to complex proteins. The authors concluded that the reaction was not confined to the tyrosine and emphasised its competitive nature and strong dependence on the relative concentration of protein and diazonium salt. Moreover, strongly acidic conditions generally required for the preparation of diazonium salts from anilines,[500] are not compatible with pH-sensitive proteins. Together with a relative instability of diazonium salts and the prerequisite of their preparation just prior to use, these drawbacks prevented this methodology from the widespread use.

The optimised conditions have nonetheless allowed its application for selective modification of tyrosines on the surface of bacteriophage MS2,[501,502] the modification of the tobacco mosaic virus,[503] and the direct conjugation on proteins.[504] Recently described by Barbas *et al*., formylbenzene diazonium hexafluorophosphate reagent[505] represents an elegant example of a stable ready-to-use reagent for tyrosine labelling and introduction of an aldehyde bioorthogonal tag, capable for future bioorthogonal modifications (Fig. 52).

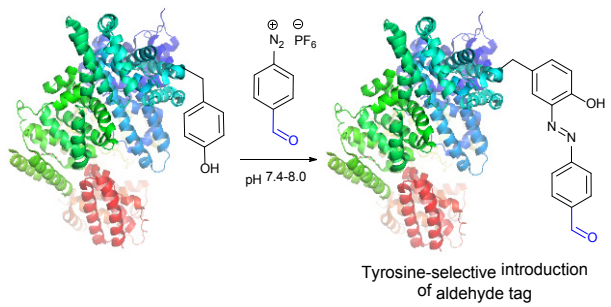


Fig. 52. Selective tyrosine labelling of BSA (pdb: 3V03) by 4-formylbenzene diazonium hexafluorophosphate (FBDP) described by Barbas *et al.*[505] for installing a bioorthogonal aldehyde functionality.

1.6.4 Mannich-type reaction

Albeit with no control of selectivity, tyrosine conjugation *via* Mannich-type cross-linking reaction have been first reported by Fraenkel-Conrat and Olcott[506] and proceeded through the conjugation of tyrosines with imines formed *in situ* by condensation of lysine amino groups and formaldehyde. The reaction conditions, namely the necessity of using high concentrations of formaldehyde and significant heating, limit the utility of this approach for the vast majority of biological applications.

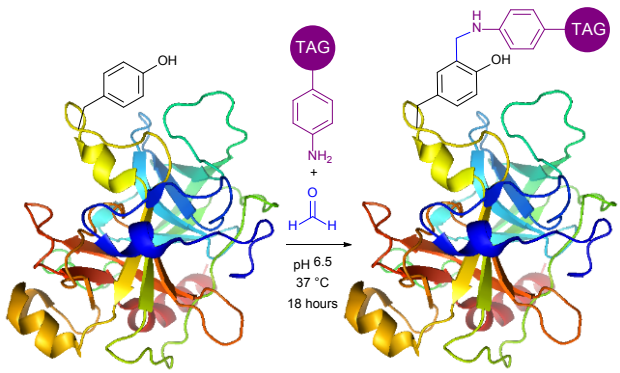


Fig. 53. Three-component Mannich-type selective modification of Y154 Tyr residue of α-chymotrypsinogen A (pdb: 1EX3) by a rhodamine dye, described by Francis *et al.*[507]

The three-component Mannich-type methodology – involving the *in situ* reaction between a Tyr residue, an amine and formaldehyde – was reincarnated more than 50 years later by Francis *et al.*[507] The authors demonstrated the possibility of selective modification of tyrosine residues of α-chymotrypsinogen A under mild conditions (pH 6.5, 25-37 ˚C) and at low concentration of the protein (20-200 μM). However, 18 hours of incubation were needed to reach a reasonable level of tagging (66% in the case of a fluorescent labelling, Fig. 53). The same group then used this to incorporate synthetic peptides into full-sized proteins.[508]

Despite recognised selectivity issues of a three-component Mannich-type approach for tyrosine labelling,[509] its main advantage is the possibility to easily vary the participating partners: an aldehyde (Fig. 53, shown in blue) and an aniline residues (Fig. 53, shown in violet). In the following publication on the subject, Francis *et al.* have demonstrated the *via*bility of NMR-based characterisation of the conjugate isotopically enriched by incorporation of ^{13}C-formaldehyde into the coupling reaction.[509] Interestingly, while a reaction by-product arising from tryptophan indole ring was revealed, Cys moiety was found to not participate in the reaction, except in the case of a reduced disulfide, which formed a dithioacetal.

Using similar precursors – electron-rich aniline derivatives – Tanaka *et al.*[510] could demonstrate the potential of *in situ* obtained imines as fluorogenic probes for tyrosine labelling. While the educts, as well as the imine derivatives, exhibited weak or no fluorescence, the addition products had a significantly higher (more than 100-fold) level of fluorescence.

In the successive study, the same group has expanded this approach to presynthesised cyclic imines completely excluding the need for using an excess of highly reactive formaldehyde.[511] Although, the authors have clearly demonstrated the applicability of their methodology in water at 25 ˚C over a wide pH range (pH 2 – 10) on a set of model phenols, no example of peptide or protein conjugation has been given.

1.6.5 Dicarboxylates and dicarboxamides

As early as 1969, the reaction of electron-rich arenes with acyclic diazodicarboxylates was reported by Schroeter.[512] Numerous examples of electron-deficient diazodicarboxylates were established in further studies being mainly focused on their synthetic usefulness for electrophilic amination in organic solvents in the presence of activating protic or Lewis acid additives.[513-520]

However, these highly reactive reagents decompose rapidly in aqueous media, which makes them not suitable for bioconjugation.[521] On the other hand, corresponding diazodicarboxyamide reagents are too stabilised and do not react with phenols in aqueous media.[521] Cyclic diazodicarboxyamides like 4-phenyl-*3H*-1,2,4-triazoline-3,5(*4H*)-dione (PTAD) were recently reported by Barbas and collaborators and represent a good compromise between reactivity and stability of diazodicarboxyl-containing reagents.[522,523] A versatile class of stable PTAD precursors, possessing different functional groups, was developed and applied for a selective tyrosine conjugation (Fig. 54). Their utilisation implies prior to use oxidation with 1,3-dibromo-5,5-dimethylhydantoin and the addition of a small amount of TRIS (2-amino-2-hydroxymethyl-propane-1,3-diol) during the step of conjugation. The latter is of crucial importance for the coupling selectivity, for it is hypothesised to serve as a scavenger of a putative isocyanate by-product of the PTAD decomposition, which is promiscuous in labelling.

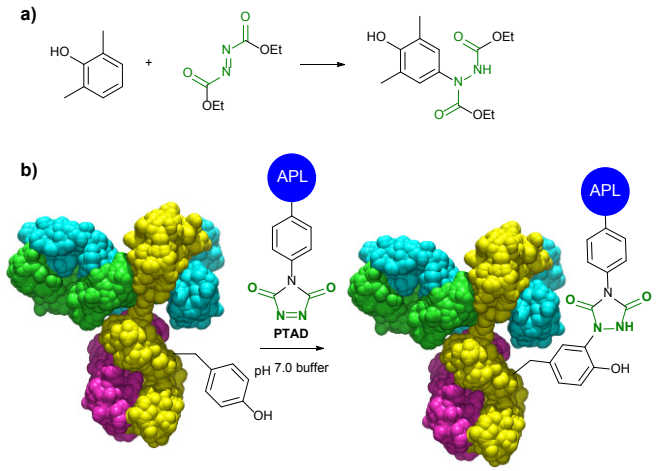


Fig. 54. Reaction of electron-rich arenes with azodicarboxyl compounds (shown in green). **a)** First example described by Schroeter in 1969. **b)** Conjugation of an HIV entry inhibitor aplaviroc (APL) with cyclic diazodicarboxamide derivative - 4-phenyl-*3H*-1,2,4-triazoline-3,5(*4H*)-dione (PTAD) – on the antibody trastuzumab demonstrated by Barbas.[523]

The non-selective labelling of other aromatic side chains of proteins is the Achilles' heel of the vast majority of approaches described for tyrosine labelling. Careful tuning of reaction conditions is important for achieving appropriate levels of selectivity. In some cases where purely chemical distinction of reactivity of amino acid moieties is not feasible, catalysis on the basis of molecular shape rather than local environment can be used to induce selectivity. This concept is routinely exploited by enzymes and allows enabling reactivity that would otherwise be kinetically impossible. In 2010, Popp and Ball used dirhodium metallopeptide catalysts for selective conjugation on tyrosine and tryptophan using the concept of the proximity-driven mechanism (see Section A 5, page 42).[441] In the following year, Silverman *et al*. have demonstrated a DNA-catalysed approach for selective labelling of tyrosine, although only on small peptide substrates.[524]

1.7 Arginine

With a pK_a value above 12, arginine (Arg, R) is mainly presented in its protonated form in acidic, neutral, and even most basic environments. Effective delocalisation of a positive charge between nitrogen lone pairs and the double bond favours the formation of hydrogen bonds[525,526] and makes the guanidinium side chain of arginine the least acidic cationic group among all 20 natural amino acids (Fig. 55).[527]

However, the pK_a value of arginine was found to vary significantly in the microenvironments within certain proteins,[528,529] allowing, in terms of Leitner and Lindner,[530] the grouping of arginines in "exposed" or "partially buried" residues, basing on the difference of their reactivities.

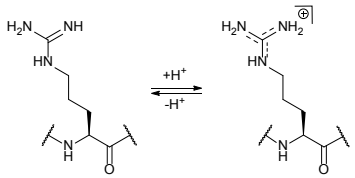


Fig. 55. Protonation of arginine side chain at neutral pH.

Most of the described approaches for arginine labelling and modification exploit the chemistry of α-dicarbonyl compounds. For instance, introduced by Takahashi[531] as an arginyl reagent, phenylglyoxal has since been applied for the study of complex systems in the past decade.[532-536] The reaction occurs under mild conditions and consists of two steps: first addition of phenylglyoxal resulting in the formation of hydrolytically instable imidazolidine diol, and the second step results in a relatively stable addition product (Fig. 56).

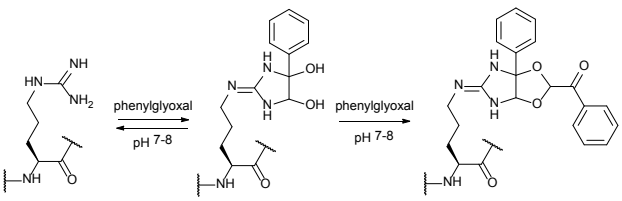


Fig. 56. Reaction of phenylglyoxal with arginine moiety.

Substituted phenylglioxal analogs, such as p-hydroxyphenylglyoxal, p-nitrophenylglyoxal and 4-hydroxy-3-nitrophenylglyoxal, were reported for spectrophotometric reagents for the study of the modification of arginine in proteins.[537-540] Because phenylglyoxal, like glyoxal, reacts with ε-amino groups at a significant rate,[531] many efforts were made to increase its selectivity towards guanidinium residue. Cheung and Fonda have studied the effect of buffers and pH on the reaction rate[541] and found that the reaction of arginine is greatly accelerated in bicarbonate-carbonate buffer systems, possibly due to the stabilisation of the obtained diol.

Geminal diones - namely 2,3-butanedione (introduced by Yankeelov)[542,543] and 1,2-cyclohexanedione (introduced by Itano)[544] - are another well-characterised reagents for the modification of Arg residues. The reaction progresses through the pathway which is similar to the phenylglyoxal addition. However, it was not until the observation that borate had a significant effect on the selectivity of the reaction that the use of this reagents became practical.[545,546] The presence of borate in the solution allows shifting of the equilibrium during the addition to a guanidine moiety through the stabilisation of reversibly obtained diol (Fig. 57).

In 2005, using this approach, Lindler and colleagues have described a method for the selection of arginine-containing peptides from a tryptic digest of the model proteins (BSA, lysozyme, ovalbumin) by a solid phase capture and release.[547] First, arginine containing peptides presented in the digest were cova-

lently modified on the guanidine moiety with 2,3-butanedione and phenylboronic acid under alkaline conditions.

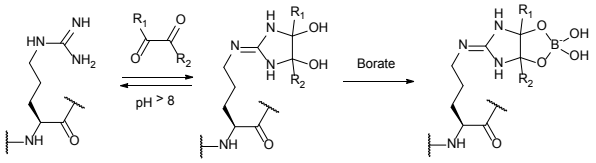


Fig. 57. Reaction of 2,3-butanedione with arginine residues of Carboxypeptidaze A (pdb: 1HDU) in the presence of borate described by Riodan.[545]

Polymeric materials allowing the immobilisation of phenylboronic acid were then used to capture the arginine-peptides on a solid support while washing away all not covalently bonded arginine-free peptides. Finally, the arginine-peptides were cleaved again from the boronic acid beads due to the reversibility of the reaction. Photoactivable bifunctional reagents for cross-linking of arginine moieties have been elaborated by Ngo *et al*. and Politz *et al*. to study enzymes with an arginine at their active sites. [548,549]

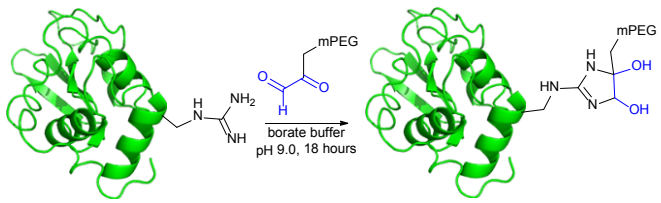


Fig. 58. Selective PEGylation of lysozyme's arginine side chains (pdb: 2LYZ) described by Gauthier and Klok.[550]

Arginine-specific PEGylation of lysozyme using polyethylene glycols containing an α-oxo-aldehyde motif in borate buffer was recently reported by Gauthier and Klok[550] and represents mild and selective method for protein modification.

Other methods described to date[551,552] possess selectivities, which are not sufficient (especially in the presence of Lys moieties) to consider them as suitable for bioconjugation.

1.8 Aspartic and Glutamic Acids

Carboxylic acid groups can be found in protein structure either on its *C*-terminus, or as side chains of Asp and Glu. Due to the low reactivity of carboxylate in water, it is usually difficult to selectively conjugate proteins at these moieties. Carboxylic acid should thus generally be converted to a more reactive ester by means of so called activating reagents.

For more than a half century, carbodiimide-mediated activation is the most extensively used methodology for the modification of free-acids in protein.[553,554] The reaction of carbodiimides with protonated carboxyl groups yields activated acylisoureas, which then react smoothly with a variety of nucleophiles, namely amines (Fig. 59).[555] It is important to utilise weakly basic amines, that remain unprotonated and thus reactive at pH below 8.0, to avoid protein cross-linking occurring at higher pH values. For this reason, weakly basic hydrazides are often reagents of choice in coupling reactions with activated carboxylic acids.[556] Although water-insoluble carbodiimides (DCC, DIC) still continue to be useful for acid-selective protein conjugation,[557,558] most current reports exploit water-soluble carbodiimides such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC). Developed by Stehan and Hlayka[559,560] these carbodiimides first proved their especial usefulness as zero-length cross-linking reagents to study proteins.[560] Subsequent studies were devoted for the application of carbodiimides for quantitation of accessible carboxyl groups in proteins,[553,554,561] preparation of antigenic conjugates,[562] and protein immobilisation.[563]

As mentioned previously, the upper limit for the optimal pH of carboxylate conjugation is defined by the reactivity of free amino groups present in protein. The lower limit is mainly determined by aqueous stability of the activating reagent. Border and co-workers[564] studied the stability of EDC in aqueous solution. It was found that EDC has a $T_{1/2}$ of 37 hours (pH 7.0), 20 hours (pH 6.0), and 3.9 hours (pH 5.0) in 50mM MES buffer at 25°C; in the presence of 100 mM glycine, the $T_{1/2}$ values were 15.8 hours, 6.7 hours, and 0.73 hours respectively. This supports the optimal pH for acid-selective conjugation to be in a range from 6.0 to 7.0. NHS (or its water-soluble analogue sulfo-NHS) is often included in coupling protocols to improve efficiency or to create a more stable intermediate. Possible side-reactions involving activating reagents were recently reviewed by Valeur and Brandley.[565]

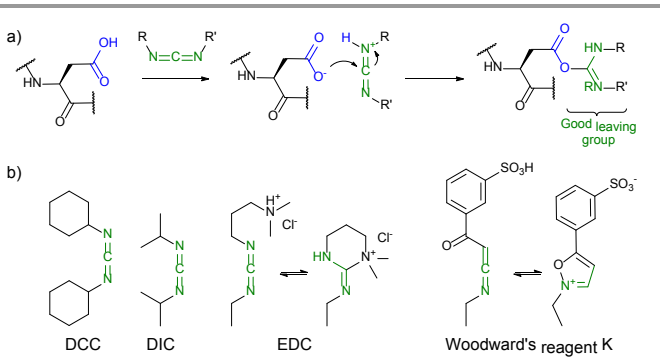


Fig. 59. **a)** Carbodiimide-mediated activation of carboxylic acid side chain of glutamate. **b)** Relevant examples of activating reagents.

Woodward's reagent K (*N*-ethyl-5-phenylisoxazolium-3'-sulfonate)[566] and analogous substrates were used as activating reagents of carboxyl groups for synthetic purposes. Bodlaender *et al*.[567] used *N*-ethyl-5-phenylisoxazolium-3-sulfonate, the *N*-alkyl derivatives of 5-phenylisoxazolium fluoroborate, to activate carboxyl groups on trypsin for subsequent modification with methylamine or ethylamine.

Lastly, several studies revealed unexpected examples of carboxyl group modification with reagents usually reacting far more effectively with other nucleophiles. For instance, *p*-bromophenacyls and iodoacetamides have been found to selectively alkylate carboxylic acid moieties of pepsin and ribo-

nuclease T1 respectively.[568-570] However, the applicability of these reagents is not general and is appropriate on specific substrates only.

1.9 METHIONINE

In spite being often considered a rather simple target for chemical modification (mainly through the oxidation and the reaction with α-halo acetic acids and their derivatives),[571,572] only a handful of conjugation methodologies involving methionine (Met, M) were described up to the present.

All approaches described in literature exploit alkylation of Met residues in acidic media. Although many nucleophilic functional groups present in proteins can react with alkylating reagents, at low pH all of them except methionine exist in protonated forms, which greatly decreases their reactivity.[573] Consequently, alkylations of other nucleophilic functional groups, such as thiols, are commonly conducted at high pH,[392] while methionine is the only functional group in proteins able to react with alkylating reagents at low pH.

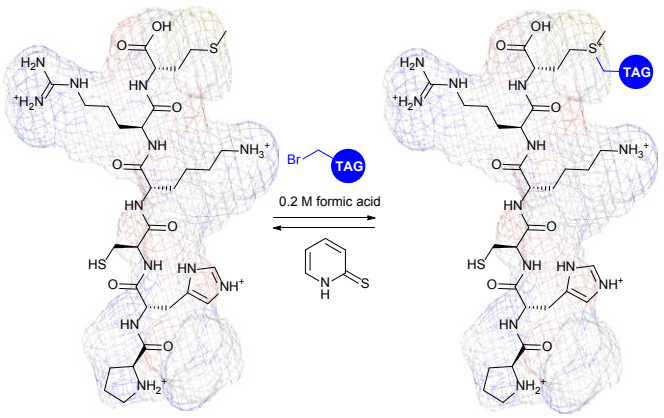


Fig. 60. Selective reversible modification of a 6-amino-acid peptide (PHCRK**M**) *via* methionine alkylation reported by Kramer and Deming.[574] No reaction with other amino acids were detected.

Basing their research on pioneering studies being done by Toennies in the 1940s,[575,576] Kramer and Deming have recently reported a reversible chemoselective labelling of methionine in peptides and polypeptides.[574] Treatment of the model peptide (PHCRK**M**) with alkylating reagents of different structures in 0.2 M aqueous formic acid (pH 2.4) gave a single product, where only the Met residue was alkylated. This resulting sulfonium salts can readily be dealkylated by addition of pyridine-2(1*H*)-thione (PyS) to give the starting peptide as the sole product along with the alkylated PyS byproduct. The removal reaction was found to be selective and allows selective dealkylating using concentrations of PyS that do not react with the disulfide bond in cystine under identical conditions (Fig. 60).

2 *N*-TERMINAL CONJUGATION

2.1 α-AMINO GROUPS

The most important thing to note about *N*-terminal amino groups is that these are the only primary amines present in the protein structure that possess an adjacent amide bond in α-position, which slightly influences their reactivity. Consequently, the majority of classical methodologies exploit this peculiarity of the *N*-terminus.

2.1.1 Classical approaches

As rather unpolarizable nucleophiles, amines react preferentially with hard electrophiles like acid anhydrides and acyl halogenides.[577] The anchimeric influence of the adjacent amide bond consists in the lowering the pK value of the amine by electron withdrawal. Consequently, this makes the discerning of *N*-terminal amino groups from the ε-amino groups of Lys residues achievable by working at a pH close to their pK values, *i.e.* under slightly acidic conditions. That is, virtually all methodologies described in the Section A 1.1 (page 3) of this manuscript are, to some extent, applicable for selective labelling of *N*-terminal amino groups of the proteins.

In the classical work on acetylation of the growth hormone, Reid[578] has for the first time demonstrated the possibility of selective modification of *N*-termini, if acetylation is performed with a relatively small amount of acetic anhydride. Further development of this approach has in several protocols for selective labelling of α-amino groups of proteins,[3,579] peptides,[580,581] and proteomes.[582-584] Under optimal reaction conditions, the use of a 5-fold excess of amine-reactive reagent in PBS (pH 6.5) at 4 °C, high levels of selectivity can be achieve after 2-24 hours of reaction. However, the preference for terminal amino groups achieved by control of pH is rather limited, mainly owing to the fickleness of the pK_a of the amino group depending on the microenvironment and reaction conditions. Consequently, more proficient methods that rely upon direct participation of the adjacent side chains or peptide bond were developed and represent to date preferential approaches for bioconjugation.

2.1.2 Ketene-mediated conjugation

A method for selective *N*-terminal modification of proteins by ketenes was introduces by Che and co-workers[585] and consisted in *N*-terminal ligation of peptides through oxidative amide bond formation using the "[Mn(2,6-Cl_2TPP)Cl]/alkyne/H_2O_2" systems. Initially, the method was tested on a set of six peptides and demonstrated its applicability. However, inevitable oxidation at Cys and Met residues hindered the application of the protocol in the field of bioconjugation. Only after having conducted the mechanistic studies of this approach, the authors have realised that ketenes generated *in situ* were the key intermediates accounting for the reactivity; therefore prior preparation thereof would allow for refraining from the need to use an oxidant. In the publication to follow, Che[586] has introduced a general approach for the modification of *N*-terminal α-amino groups of a series of proteins and peptides using an isolated

alkyne-functionalised ketene (Fig. 61). Interestingly, in contrast to classical approaches, increasing the pH of the reaction mixture did not significantly affect *N*-terminal selectivity of conjugation. For comparison, the ketene reagent was side-by-side benchmarked with a corresponding NHS-ester. Remarkably, much poorer *N*-terminal selectivity was obtained for all studied peptide substrates when the NHS ester was used. The reason for this impressive specificity of ketenes, however, remains unelucidated.

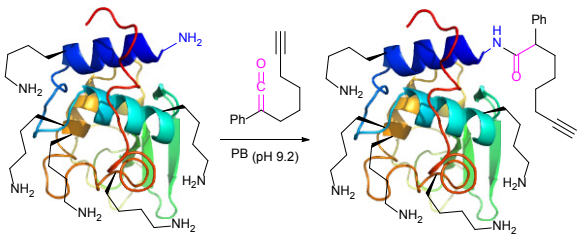


Fig. 61. Selective labelling of lysozyme (pdb: 2LYZ) described by Che.[586] Moderate-to-high level of selectivity towards *N*-terminal residue (K1) was achieved in the presence of 5 other in-chain lysines (LYS-13, LYS-33, LYS-96, LYS-97 and LYS-116). The ketene was synthesised from the corresponding acid by a two-step protocol: transformation into a mixed anhydride (oxalyl chloride, DCM) followed by the transformation thereof into ketene upon the reaction with a base (TEA, THF).

2.1.3 Transamination

It was after the success of the conversion of glyoxyloyl groups into glycyl groups (see Fig. 69, page 32)[587] that Dixon and co-workers realised that if a terminal Gly residue can be made by transamination then a terminal residue of any kind might be transformed into a corresponding carbonyl-containing residue by transamination. These introduced carbonyl groups are not naturally occurring functionalities in proteins and can therefore be used as unique loci of attachment for synthetic groups through the formation of hydrazone or stable oxime bonds.[588,589] Inspired by the pioneering works of Metzler and Snell, [590] and Cennamo and collaborators[591,592] on the transamination of simple amino acids and peptides under harsh conditions (heating at 100°C and pH 5.0), Dixon and Moret[593,594] developed a method for mild transamination in the presence of copper(II) salts, which allowed the reaction to pass at 25 °C (Fig. 62). The isomerisation of the imine generated *in situ* by catalysed 1,3-proton shift transfer is the key step of the transformation which defines both its direction and the reaction rate. It is clearly the activation of α-protons of *N*-terminus by the adjacent peptide bond and the metal ions that makes this reaction specific to α-amino groups. Interestingly, Dixon reports that no traces of the reaction of lysine side chains were ever observed.[595]

Quite a wide range of reaction conditions has been tested ever since. The discovery that pyridine[596] and acetate[597] greatly accelerated the transamination of amino acids led to a slightly milder reaction conditions, which however were still too harsh to maintain the folded structure of most proteins, and were therefore more appropriate for sequence-analysis applications.[598]

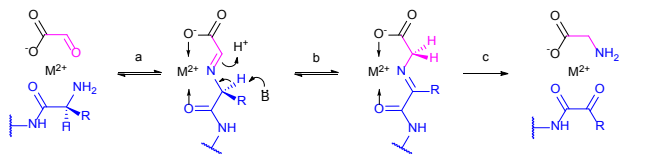


Fig. 62. General scheme of the transamination reaction activated by metal ions (typically Cu^{2+} and Ni^{2+}).[593] Principal steps of the reaction mechanism: a) generation of the imine; b) isomerisation of the obtained imine by proton removal; c) hydrolysis of isomeric imine to generate transaminated reaction partners.

Only recently Francis and co-workers[599] have re-examined Cennamo's approach[591] of amino acid transamination in the presence of pyridoxal-5-phosphate (PLP, vitamin B_6) at 100 °C and found much milder the reaction conditions when modifying the *N*-terminal residues of peptides (65 °C for reaction to be over in 2 hours; 25 °C to achieve a complete conversion in 24 hours). Screening experiments on different *N*-terminal amino acids of the peptides indicated that the aldehyde structure strongly influenced the reaction efficiency. Amusingly, known for more than 50 years PLP emerged as the most effective aldehyde among dozens being screened, affording the highest yields at milder reaction conditions.

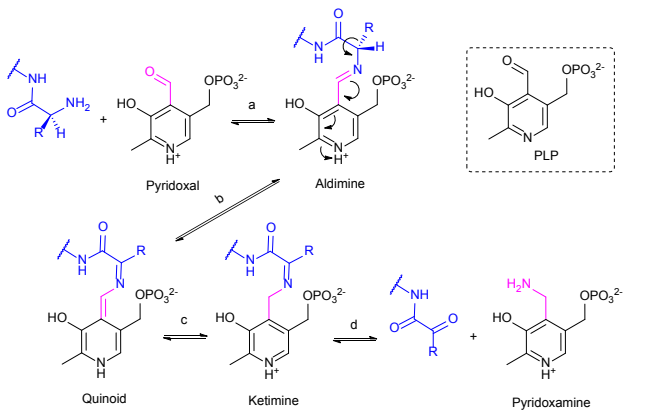


Fig. 63. Structure of PLP and mechanism of PLP-mediated transamination. Reaction pathway consists of (a) condensation reaction between pyridoxal and the amine; (b-c) tautomerisation of the obtained aldimine being favourable because of a much lower intrinsic pKa values of α-proton (shown in blue); (d) hydrolysis of the resulting ketimine, accompanied by decarboxylation in the case of aspartic acid (R = -CH_2COOH).[599]

The mechanism of PLP-catalysed transamination is depicted in Fig. 63. The reaction of PLP and *N*-terminal amine results in forming of a Schiff base aldimine (a). Then the α-proton the amino acid transferred to the 4' position of the pyridoxal unit (b-c). Finally, hydrolysis of the obtained ketimine leads to the desired α-ketoacid and pyridoxamine phosphate (d). Quinoid is an important intermediate for the transformation of aldimine to ketimine and can be found in all transamination reagents described to date.

Under optimal reaction conditions: 10-50 μM protein and 10-50 mM PLP at 37 °C in PB (pH 6-7), a complete conversion is generally achieved after 2-24 hours.[600] An example of transamination-conjugation methodology was demonstrated by Francis

in the initial publication on selective labelling of an *N*-terminal glycine residue of horse heart myoglobin and enhanced green fluorescent protein (EGFP, Fig. 64).[599]

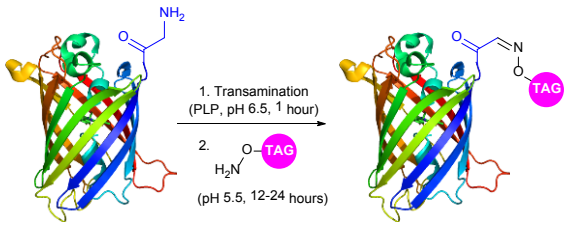


Fig. 64. Site-specific *N*-terminal labelling of EGFP (pdb: 2Y0G). Proteins possessing *N*-terminal carbonyl groups obtained by in the first step PLP-mediated transamination were labelled with hydroxylamine probes in the second step.[599]

Although the side chain of the *N*-terminus does not participate directly in the transamination mechanism, the reaction rates were found to vary significantly depending on the amino acid in the *N*-terminal position.[601] Generally, the majority *N*-terminal amino acids provide high yields of the desired transaminated products; however, some residues (His, Trp, Lys, and Pro) generate adducts with PLP itself, while other are incompatible with the technique because of known side reactions (Ser, Thr, Cys and Trp) or complete inertness (Pro).

In the attempts for the investigation of transamination reaction scope and limitations, Francis and collaborators have prepared an 8000-member one-bead-one-sequence combinatorial peptide library in which the three *N*-terminal residues were varied.[602] Interestingly, Ala-Lys (AK) motif was found especially favouring the transamination yields (Lys residue is hypothesised to accelerate the isomerisation step of 1,3-proton shift acting as a general base). To demonstrate this, labelling of the Type III "Antifreeze" Protein and its mutant presenting the AKT sequence on the *N*-terminus were side-by-side benchmarked. At every time point analysed, AKT terminus outperformed the wild-type one (GNQ) at different concentrations of PLP.

Although mild reaction conditions of PLP-mediated transamination render it amenable for the modification of intact proteins,[603] the yields are generally not high and elevated temperatures are usually required, which largely limits the practical applicability of the approach. Given this situation, Francis and co-workers[604] have utilised above-described combinatorial approach to identify another transaminative agents. As a result, *N*-methylpyridinium-4-carboxaldehyde benzenesulfonate salt (Rapoport's salt, RS, Fig. 65) was identified as a highly effective alternative to PLP. Furthermore, this was found to be particularly efficient for glutamate-rich sequences. The fact that several antibody isotypes dispose at least one glutamate-terminal chain makes RS particularly amenable for their selective conjugation. Remarkably, the difference of transamination reaction rates on Glu- and non-Glu polypeptides was significant enough for selective labelling of only the heavy chain of Immunoglobulin G1 (containing the *N*-terminal Glu residue), while leaving unmodified the light chain. This was assigned to be due to the higher steric hindrance of the already less-reactive substrate (*N*-terminal Asp). To ensure these results an IgG1 mutant possessing Asp-Asp-Ser sequence on both chains was prepared. Indeed, this underwent the modification of both sites when exposed to RS. Facing another recognised drawback of Francis methodology, namely low efficiency for bulky amino acid termini (Leu, Ile, Val), Zhang *et al.*[605] have elaborated an efficient PLP analogue, FHMDP (Fig. 65), that demonstrated much higher efficiency in transamination thereof.

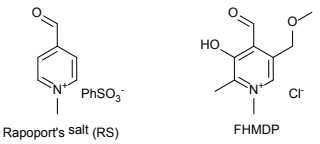


Fig. 65. PLP-analogues used as transamination agents: Rapoport's salt[604] and FHMDP.[605]

Above-mentioned transformations provide just a few examples of the rapidly growing field of transaminative modification of proteins. Recent advances have also resulted in elaboration of general approaches for protein immobilisation (Fig. 66),[15,606] dual fluorescent modification of periplasmic solute binding proteins,[607] protein PEGylation and PEG-like conjugation (*e.g.* OEGMAtion),[608] preparation of phage conjugates,[609,610] *N*-terminus proteomics,[611,612] enabling Wittig[613] and Pictet-Spengler ligation on transaminated proteins.[614,615]

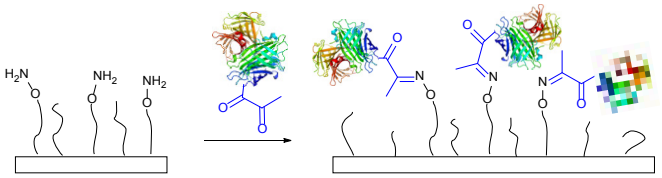


Fig. 66. Patterning of PLP-transaminated streptavidin (pdb: 4BX5) on aminooxy terminated polymer films.[15]

2.2 Serine and Threonine

In 1960 Waller and Dixon have described one the first procedure for selective *N*-terminal modification of peptides. It consisted of preparation of corticotrophin selectively acetylated on its terminal serine.[616] Although being possible only under highly denaturing conditions of alkali exposure, the approach allowed for a spontaneous intramolecular O → N acyl transfer[617] on the *N*-terminal Ser residues, while *O*-acetyl groups of in-chain Ser, Thr and Tyr residues were hydrolysed. The general idea of overcoming the entropy barrier of an otherwise slow intermolecular process by bringing two reacting partners together through a covalent linkage to initialize an intramolecular reaction has been later transformed into a large variety of selective modification reactions.

2.2.1 O → N shift of oxazolidines

35 years after the early publication of Waller and Dixon,[616] Liu and Tam[618-620] have extended the applicability of O → N acyl transfer mediated methodologies by developing an approach to the chemical ligation of *N*-terminal serine, threonine, and cysteine unprotected peptide segments with no need of using protecting groups. The key step of the process is the formation of the peptide bond through an intramolecular rearrangement between the two closely neighbouring carboxyl and secondary amino groups of formed of oxazolidine (in the case when Ser and Thr residues are involved) or thiazolidine (in the case of cysteine, see Section A 2.3.2, page 38) (Fig. 67). When conducted at the optimal pH of 4-5 for the condensation step and slightly higher for the rearrangement, the reaction was found to be very clean, and no side products were observed.[619]

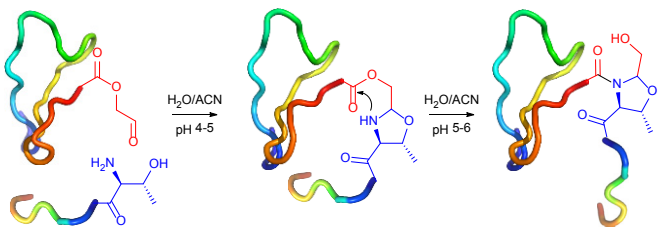


Fig. 67. General scheme of the ligation strategy proposed by Liu and Tam.[618] Conjugation with *N*-terminal Thr is depicted; analogous transformation occurs with Cys and Ser, but not with His, Tyr and Trp.

Although this methodology have demonstrated its high potential for the ligation of unprotected peptides,[618,621,622] the generating of a "non-native" heterocyclic fragment at the site of ligating the two peptides made it extruded almost completely by other "native ligation" approaches; notably, by a similar mechanistically native chemical ligation (NCL) (see Section A 2.3.1, page 33).

2.2.2 Periodate oxidation

A mild version of serine and threonine modification implies their prior conversion into glyoxyloyl derivatives *via* periodate oxidation first described by Fields and Dixon[623] and later transformed into a general method for site-directed modification of proteins with *N*-terminal Ser or Thr by Geoghegan and Stroh.[624] Based on the periodic acid mediated oxidation,[625] the reaction occurs only when there exists the target site for the periodate to form a cyclic intermediate, that is to say, when *N*-terminal residue is represented by a serine or a threonine, or when hydroxylysine is present (rarely occurring in proteins). Possible side reactions include oxidation of the side chains of Met, Trp, and His. However, the potential for side reactions can be diminished by using very low periodate-to-protein molar ratios, as it was demonstrated by Geoghegan and Stroh in their experiments on two model peptides, **S**IGSLAK and **S**YSMEHFRWG, and with recombinant murine interleukin-1α (an 18-kDa cytokine with *N*-terminal Ser, Fig. 68),[624] or by conducting the oxidation at neutral pH.[626]

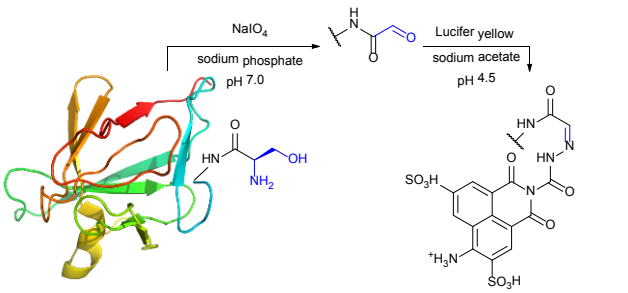


Fig. 68. *N*-terminal serine labelling of recombinant murine interleukin-1α (pdb: 2KKI) with Lucifer yellow dye described by Georghegan and Stroh.[624] The method is also applicable if *N*-terminal threonine is present.

As in the initial publication of Geoghegan and Stroh, obtained glyoxyloyl can serve as the locus for further chemical modification involving aldehyde-selective reactions (*e.g.* through the formation of stable oxime, hydrazone or previously described oxazolidine moieties).[588,589] Robin and colleagues have demonstrated the possibility of using this two-step methodology for assembling two unprotected protein fragments: oxidised to glyoxyloyl-containing and hydrazide peptide derivatives.[627] Rose and co-workers[628] exploited the reactivity of generated glyoxyloyls towards *O*-alkyl hydroxylamine derivatives to synthesize a pentameric form of the cholera toxin subunit B.

Further investigation of periodate oxidation allowed its promoting for site-selective tagging, PEGylation, preparation of protein conjugates, protein capture and synthesis of large protein dendrimers[626,629-634]. It is worth mentioning that periodate oxidation is incompatible with a number of protein classes. For instance, glycoproteins will undergo periodate-based cleavage of polysaccharide chains as a side reaction pathway.[635]

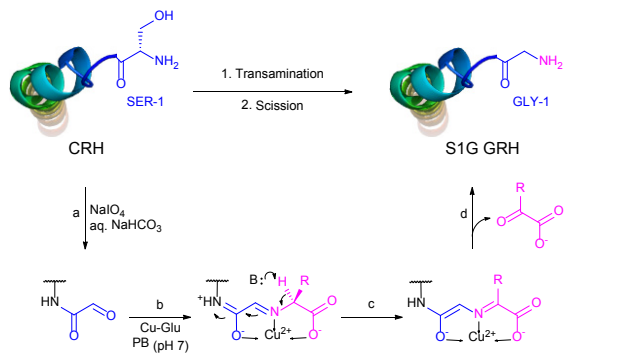


Fig. 69. Preparation of S1G mutant of corticotropin-releasing hormone (GRH) by regioselective transformation of *N*-terminal serine to glycine.[587] Reaction steps: (a) selective serine oxidation by periodate; (b) formation of imine with amino group of glutamic acid precomplexated with Cu^{2+} (Cu-Glu); (c) base-assisted isomerisation of imine; (d) hydrolysis of imine, completed when aspartate and copper ions are removed.

Lastly, glyoxyloyles can easily be transformed into corresponding amines *via* transamination reaction in the presence of copper(II) or nickel(II) salts.[587,593,636] The reaction mechanism as well as need for both essential components of the system: the acceptor of the glyoxyloyl (usually aspartic acid or glycine) and

the cation of a heavy metal are explained in Fig. 69. Despite being of moderate interest for bioconjugation by itself, this approach has initiated the development of a more general methodology for selective *N*-terminus modification – transaminative conjugation (Section A 2.1, page 29). The reader is directed to a recent review by El-Mahdi and Melnyk[637] for a complete overview of the glyoxyloyl transformations in bioconjugation.

2.2.3 Phosphate-assisted ligation

Conceptually catching phosphate-assisted ligation at serine and threonine was recently reported by Payne and Thomas.[638]

The inherent reactivity of an *N*-terminal phosphorylated Ser or Thr residues was demonstrated to significantly facilitate the amide bond formation with a range of *C*-terminal peptide thioesters. Although it is not yet clear what exact intermediate is formed during ligation, the authors have hypothesised that rapid acyl migration to the *N*-terminal amine of a peptide occurs through the formation of unstable acyl phosphate (Fig. 70).

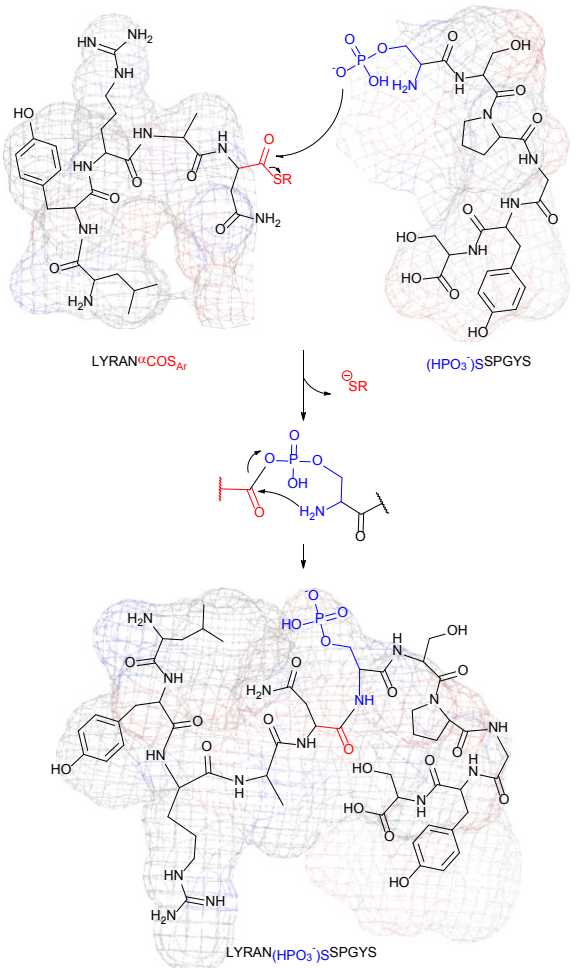


Fig. 70. Hypothesised mechanism of the phosphate-assisted ligation reported by Thomas *et al.*[638]

2.2.4 Indirect approaches

Ligation at Ser/Thr can also be achieved considering two distinct indirect approaches. Firstly, as it was demonstrated by Danishefsky and co-workers,[639] NCL-desulfurisation methodology can be used to access threonine at ligation sites (see Section A 2.3.1.4, page 37). Secondly, as an alternative to the NCL-desulfurisation sequence, cysteine obtained during NCL can be chemically transformed into serine by methylation followed by activation of the obtained S-methylcysteine by cyanogen bromide (BrCN, Fig. 71).[640]

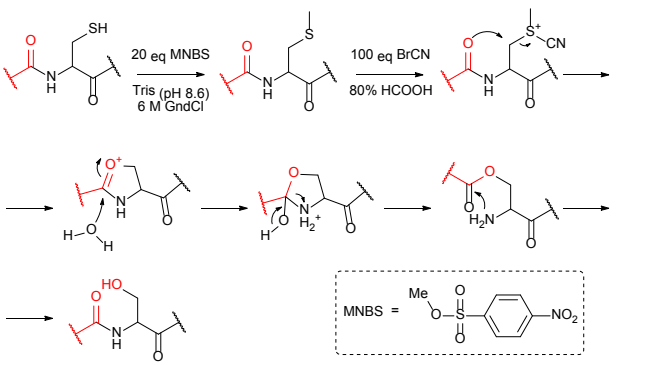


Fig. 71. Conversion of a Cys into a Ser residue by transforming the former into a corresponding methyl cysteine and following thereof by CNBr.[640] Methionine must be protected by transforming into corresponding sulfoxide form, because this is inactive for the CNBr reaction.[641]

2.3 Cysteine

Generation of *N*-terminal Cys residue for native chemical ligation can be accomplished using solid-phase peptide synthesis,[642] proteolytic processing,[643] or by the spontaneous hydrolysis of intein fusion protein.[644] Moreover, genetically directed, site-specific incorporation of 1,2-aminothiols handle into proteins has been recently reported by Chin and associates.[645]

2.3.1 Native Chemical Ligation

The very principle of "chemical ligation" was coined by Kent in the early 1990s and consisted in an approach for covalent condensation of unprotected peptide segments by the means of "unique, mutually reactive functionalities designed to react only with each other and not with any of the functional groups found in peptides".[646] That is, a general method that would enable the application of chemical tools to the world of the proteins. However, the original ligation chemistries exploited the reciprocal reactivity of chemical functions which are not present in native proteins and thus their prior introducing onto reacting partners is require and often associated with synthetic difficulties. Moreover, a non-native linkage is generated at the ligation site; therefore, many scientists remained sceptical about the validity of using such "analogous" proteins as tools for understanding the molecular basis of protein function.

In 1994, confronted with this criticism, Kent and co-workers[647] introduced a versatile approach to the linkage of peptide fragments using a native peptide bond – native chemical ligation (NCL). Based on the original principles of the chemical ligation

methodology[646] and the ability of thioesters to undergo S → N acyl shift discovered by Wieland *et al.*[648], NCL allowed to achieve chemoselective formation of the amide bond in the presence of unprotected nucleophilic amino acid side chains as alcohols (Ser/Thr), phenolates (Tyr), free amines (Lys), carboxylates (Glu/Asp), or other thiolates (Cys) presented in the macromolecule.

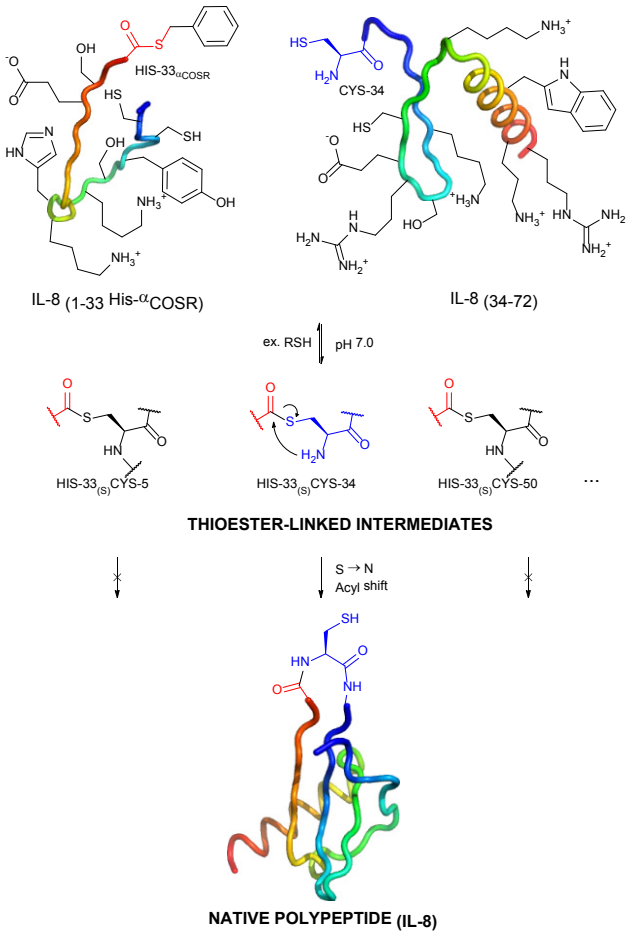


Fig. 72. Synthesis of the human interleukin 8 by native chemical ligation elaborated by Kent *et al.*[647] The ligation reaction occurs between an unprotected peptide thioester fragment – IL-8(1-33 His-αCOSR), and a second unprotected peptide possessing a *N*-terminal cysteine – IL-8 (34-72). First step of thioester exchange results in different thioester-linked intermediates, among which only the peptide obtained from the corresponding *N*-terminal cysteine can undergo following irreversible step of an S–N acyl shift resulting in obtaining of a native amide bond at the linkage site.

By analogy with previously developed O → N acyl shift (see Section A 2.1, page 29), the reversibility of the thioester-thiol exchange in the presence of an exogenous thiol additive coupled with the capture of the acyl segment by S → N acyl shift, being possible only in the case when the latter is brought in the close proximity to an amine in a *N*-terminal cysteine thioester intermediate, result in an exquisite regioselectivity of this methodology. The product resulting from this S → N acyl shift represents a peptide, consisting of two fragments linked by a native peptide bond through a cysteine residue (Fig. 72).

Typically, the reaction performed in PS or PBS buffer (pH 7.0-8.5) at 37°C is complete in less than an hour and with high yields.[649,650] Solubilizing agents such as guanidine hydrochloride or urea do not interfere with the ligation and are usually used to enhance the concentration of peptide segments, and thus increase reaction rate. It is important to prevent the thiolate of *N*-terminal cysteine from the oxidation resulting in a disulfide-linked dimer, which is unreactive in the ligation. A reductant (*e.g.* TCEP) or an excess of thiol corresponding to the thioester leaving group (4-5%, vol/vol) is generally added to keep the Cys residues in reduced form. Moreover, the latter largely increases the overall rate of NCL by reversing the first step of transesterification for in-chain intermediate adducts deprived from the possibility to undergo S → N acyl shift and to generate a stable amide bond.

The first step in synthesizing a protein by NCL generally consists in defining the fragments to be used in the ligation reactions. Preferentially, naturally occurring AA-Cys motifs in the native sequence should be chosen as the ligation sites (AA stands for any amino acid). Val, Ile, Asp, Asn, Glu, Gln and Pro represent less favourable choices, because of lower ligation rates and possible side reactions,[649] which, however, can be accelerated either by transformation of the corresponding thioesters into selenoesters,[651] or by tuning the reaction pH.[652]

Higher reaction rates were reported to be achievable while using good thiol-containing living groups, *i.e.* mildly acidic thiols such as thiophenol, 4-(carboxymethyl)thiophenol (MPAA), or 5-thio-2-nitrobenzoic acid (TNB, the reduced form of Elman's reagent). These are generally generated *in situ* by thiol-thioester exchange from the relatively unreactive peptide-($^{\alpha}$COSCH$_2$CH$_2$CO)-Leu alkylthioester by adding an excess thereof (1-5%, vol/vol).[649]

2.3.1.1 Sequential NCL

The complexity of proteins that can be obtained by NCL technique is limited by the maximum size of the accessible synthetic peptide segments. Two main approaches used today for the preparation thereof are: solid phase peptide synthesis pioneered by Merrifield (allowing the synthesis of proteins containing up to 50 residues),[642] and recombinant DNA expression elaborated by Lobban *et al.*[653] and Jackson *et al.*[654] These allow for the preparation of peptide fragments containing up to 50, and up to 150 amino acid residues respectively.

Sequential native chemical ligation allows further extending of this limit by means of *N*-terminal cysteine protected peptides. Three polypeptide fragments: a peptide$_1$-COSR$_{Ar}$ (*N*-terminal fragment), a protected PG-Cys-peptide$_2$-COSR$_{Ar}$ (middle fragment), and a Cys-peptide$_3$ (*C*-terminal fragment), are thus assembled in a one-pot three step synthesis. Firstly, the middle fragment and the *C*-terminal fragment are ligated under the classical reaction conditions of NCL. Then protecting group is removed, uncovering *N*-terminal cysteine of the obtained poly-

peptide (central plus *C*-terminal fragment), which undergoes the second NCL with the *N*-terminal fragment to give the target protein (Fig. 73).

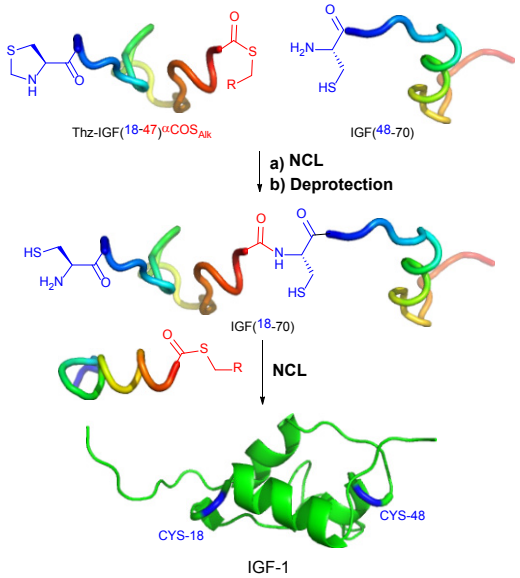
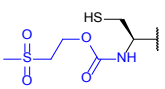
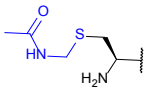
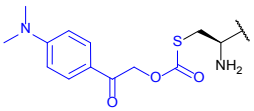


Fig. 73. Synthesis of insulin-like growth factor 1 (IGF-1) *via* sequential native chemical ligation described by Sohma *et al.*[655] The reversible protection of the α-amino group of the central peptide fragment IGF(18-47) prevents its self-reaction with the α-thioester moiety present in the same molecule. Thiazolidine protecting group can be easily removed by brief treatment with $NH_2OMe{\cdot}HCl$ at pH 4. NCL reaction condition used: PB (pH 6.7), 6M GndCl, 10 mM MPPA, 20 mM TCEP.

Since its introduction, sequential native chemical ligation has demonstrated its general applicability to the preparation of various complex assemblies. Consequently, several methodologies compatible with NCL for the protection of *N*-terminal Cys residues were elaborated. The most relevant among them are depicted in the Table 1.

Table 1. Chemical functions generally utilised for the protection of *N*-terminal cysteine during sequential native chemical ligation.

	Protecting group		Deprotection	Examples
1	Thz		NH_2OMe pH 4	IGF,[655] HIV-1 Tat,[657] Crambin,[658,659] EPO,[660] PYP.[661]
2	Msc[662]		pH 12	SOD,[663] Abl-SH3.[664]
3[a]	Acm[665]		AgOAc, DTT	Crambin,[666] DAGK.[667]
4	Mapoc[668]		hν > 300 nm	hBNP-32[668]

[a] Require a preparative-HPLC step before removal, gives lower overall yield compared with Thz.[666]

The reader is referred to a recent review by Melnyk and collaborators[656] for a more complete overview of the sequential ligation strategies on proteins.

2.3.1.2 Kinetically controlled ligation

The simple observation that the reaction rate of NCL largely depends on the thioester nature was exploited by Bang *et al.*[658] to introduce a convergent strategy for the synthesis of native peptides – kinetically controlled ligation (KCL).

The fact that the kinetics of NCL with alkylthioesters is significantly inferior of those with arylthioester makes it possible to control the intrinsic dual reactivity of a bifunctional Cys-peptide$_2$-$COSR_{Alk}$ so that it would selectively react with a peptide$_1$-$COSR_{Ar}$ and then undergo a classical NCL (*i.e.* in the presence of exogenous aryl thiol) with a third Cys-peptide$_3$ to yield an assembled peptide1-Cys-peptide$_2$-Cys-peptide$_3$ with no need to use protecting groups. Bang and collaborators have applied this methodology to assemble a 46-residue protein crambin from six peptide fragments (Fig. 74).

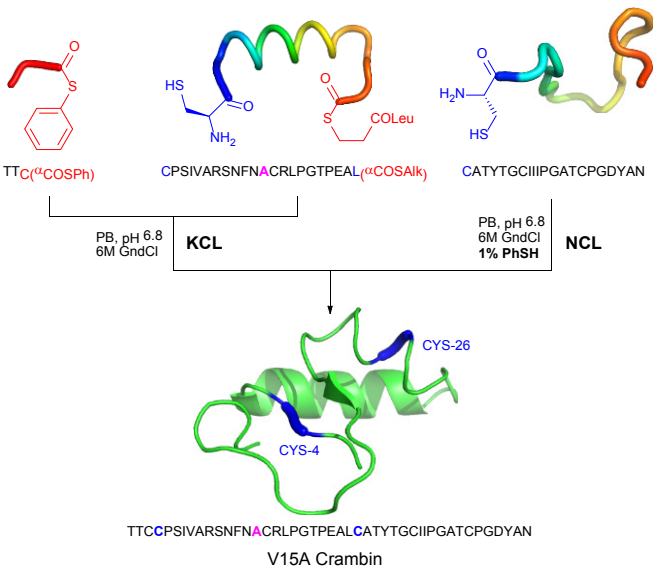


Fig. 74. Two final steps of the synthesis of V15A crambin (pdb: 3NIR) described by Bang and associated[658]. The mutation was introduced for simplifying the prior KCL step of assembling the second peptide. Kinetically controlled ligation of S_{Ar} thioester spontaneously occurs in aqueous media in the absence of exogenous thiophenol, while native chemical ligation of S_{Alk} thioester must be accelerated by the addition of 1% PhSH.

Further advancing pioneering works done by Botti *et al.*[669] on *in situ* acyl migration, KCL methodology has been recently extended from alkylthioesters to a full class of *O*-esters undergoing a spontaneous transformation to produce a thioester when exposed to a reducing agent through disulfide bond reduction followed by O → S acyl shift (Fig. 75).[670]

Through a thorough investigation Zheng *et al.*[670] have defined that structures of the *O*-esters have an important effect on their reactivity. The authors have side-by-side benchmarked their methodology with previously described KCL by synthesizing the same V15A crambin (Fig. 74) by a one-pot one-step condensation of peptide segments and found its applicability to this

system. Readily available by Fmoc solid-phase synthesis, these *O*-ester scaffolds can expand the applicability of NCL to substrates with hardly accessible thioester peptide fragments.

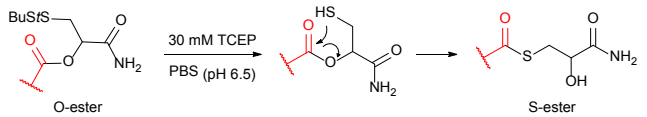


Fig. 75. Schematic representation of the O → S acyl shift of *O*-esters containing a disulfide bond described by Zheng *et al*.[670]

2.3.1.3 *Access to C-terminal peptide thioesters*

Classical approaches

The preparation of *C*-terminal α-thioesters involved in native chemical ligation is often associated with synthetic difficulties. Being especially reactive species, they either have to be introduced at the end of the synthetic pathway or be kept in a hidden form of thioester surrogates possessing higher stability.

Despite its recognised drawbacks due to hazardous acid treatment often leading to undesired side-reactions, the protocol of *in situ* neutralisation for Boc-based solid-phase peptide synthesis represents the most effective approach for the preparation of peptidyl thioesters.[649,671-673] Alternatively, the Fmoc synthesis approach was investigated[674-676] and found to be favoured when synthesizing phospho- and glycopeptides.

Expressed protein ligation (EPL)

Introduced by Muir *et al*. in 1998,[677] EPL represents another approach for the preparation of α-thioesters. It allows for obtaining the recombinant protein thioester by thiolysis of an intein fusion protein and thus enables a large pool of elaborated recombinant protein techniques for NCL. The reader is directed to several recent reviews of this area on chemical research for more details.[678,679]

X → S acyl-transfer

An elegant approach for keeping thioester in a relatively inert "hidden" form ready to be uncovered when required was first introduced by Danishefsky and colleagues in 2004 (Fig. 76).[680] Several years later, the name "crypted thioesters" was coined by Otaka[681] as a general term for such compounds. This approach is of especial interest, because it enables the assembly of peptide segments in *N*-to-C direction, which is rather rare and often difficult to achieve.[656] Indeed, all above-described methodologies rely on the assembling of polypeptide chain in *C*-to-N direction: for instance, the sequential native chemical ligation *per se* consists of iterated cycle of ligation-deprotection-ligation…, *i.e.* adding new peptide fragments onto *N*-termini after deprotection thereof (see Section A 2.3.1.1, page 35).

Synthesis of thioesters *in situ* from stable amides *via* N → S acyl-transfer was demonstrated by Ohta *et al*.[682] who studied acylated oxazolidinones derived from S-protected cysteine. These possess a distorted amide planarity provoking so called ground-state destabilisation[683] and, as a consequence, favour the acyl migration to a deprotected thiolate. Nakahara and collaborators have studied and elaborated two classes of secondary amides amenable to the N → S acyl shift at low pH values: 5-mercaptomethyl prolines[684] and *N*-alkyl cysteamides (Fig. 77a).[685] Oxazolidinones[682] and *N*-sulfanylethylamilides (SEAlides)[686] described by Otaka and collaborators were found to possess similar aptitude towards N → S acyl-transfer shift at low pH (for a complete overview of N → S acyl-transfer systems described before 2010 see review by Kang and Macmillan).[687] Erlich *et al*.[688] have recently applied *N*-alkyl cysteamide-based approach for the synthesis of 76-residue ubiquitin thioester, while Otaka and collaborators have demonstrated high potential of SEAlides by conducting the chemical synthesis of 162-redue active glycosylated GM2-activator protein.[689]

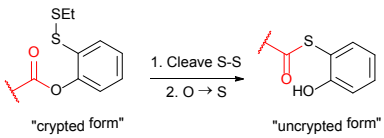


Fig. 76. Principle of "crypted thioesters" demonstrated by Danishefsky and co-workers.[680] Intramolecular O → S migration of acyl residue resulting in generation of highly active thioester (its "uncrypted form") occurs only upon reductive cleavage of disulfide bond of reasonably stable "crypted form".

Almost simultaneously have two research groups reported a general approach, based on the application of bis(2-sulfanylethyl)amides (SAM) as precursors for NCL.[690-693] An interesting extension of the methodology enabling the possibility of triggering the reactivity of SAM – so called SEA$^{on/off}$ system – has been further elaborated by Melnyk and collaborators.[694] The transition between reactive (SEA^{on}) and unreactive (SEA^{off}) states is simply triggered by mild oxidation/reduction procedures (Fig. 77b). SEA^{off} can be easily switched on *via* TCEP reduction, while the reverse switching off is achieved by mild oxidation with iodine. After few seconds, the excess of iodine is decomposed by the addition of dithiothreitol (DTT).

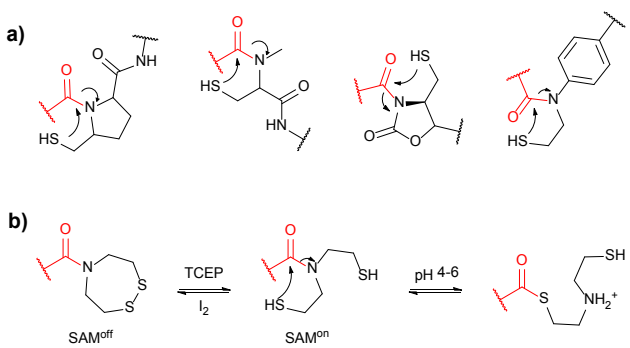


Fig. 77. N → S migration of acyl residue. The thioester-amide equilibrium is shifted towards the thioester form at low pH due to the protonation of the secondary amides. In the presence of 3-mercaptopropionic acid (MPA) all intermediates produce MPA-thioester. **a)** From left to right: reduced form of Danishefsky's ester,[680] 5-mercaptomethyl prolines[684], *N*-alkyl cysteamides,[685] oxazolidinones,[682] and *N*-sulfanylethylamilides.[686] **b)** SAM$^{on/off}$ approach by Melnyk and collaborators.[694]

Other amino acid residues susceptible to oxidation such as methionine or tryptophan are not affected because of this very

short exposure to oxidant. However, cysteine residues must be protected with *tert*-butylsulfenyl groups to remain unaffected. At low pH values these are not reducible by DTT, allowing thus reliable protection of cysteines during the cycles of oxidative-reductive SAM triggering.

To demonstrate the potential of this newly established and optimised SAM$^{on/off}$ system, Melnyk and collaborators have synthesised 85-residue domain of the hepotocyte growth factor (HGF 125-209).[694]

2.3.1.4 Indirect ligation on other amino acid using NCL

In 2001, Yan and Dawson[695] have introduced an approach that, for the first time, allowed virtually conducting NCL at alanine. While the ligation still occurred at *N*-terminal cysteine, its subsequent desulfurisation with freshly prepared Raney nickel produced the native target sequence containing alanine residue at the ligation site (Fig. 78). This strategy has inspired the development of a large pool of various ligation junctions that includes phenylalanine,[696,697] glycine,[698] valine,[699,700] leucine,[657,701] threonine,[639] serine,[640] proline,[702-704] aspartate,[705] glutamine,[706] homocysteine,[707,708] methionine (by subsequent S-methylation of the ligated homocysteinyl product by *p*-nitrobenzenesulfonate)[707] arginine,[709] and lysine (α- or ε-selective) ligations.[710,711]

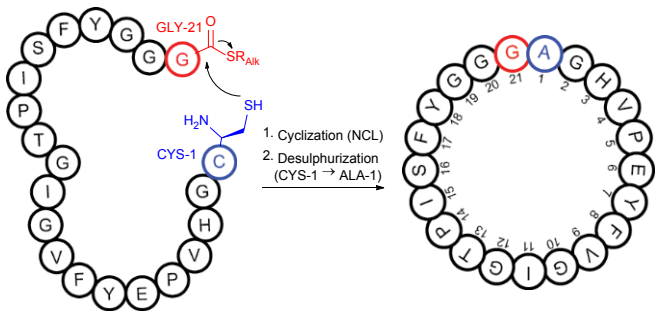


Fig. 78. Synthesis of 21-residue cyclic antibiotic peptide Microcin J25 *via* NCL-desulfurisation strategy resulting in alanyl-linked polypeptide chain developed by Yan and Dawson.[695] Reaction conditions: cyclisation – Tris-HCl, 6 M GndCl; desulfurisation – NaOAc, pH 4.5, 6 M GndCl, H_2 (Pd/Al_2O_3).

In order to extend native chemical ligation-desulfurisation approach to these amino acids, a SH-group should be attached to the carbon atom situated in the β-position to amino group (in some cases in the γ-position). The new building block containing a β-mercapto-α-amino or γ-mercapto-α-amino fragment is then introduced at the *N*-terminal position during solid-phase peptide synthesis or by means of DNA-recombinant technologies.[697] Following ligation is conducted under classic NCL conditions: at pH 7.5-8.0 in the presence of TCEP as reducing agent and 1% of exogenous thiol additive. Finally, the desulfurisation of cysteine gives a nascent residue of interest (Table 2). Besides aforementioned reduction on Raney nickel, various milder conditions such as nickel boride,[695] Pd/Al_2O_3,[712] or metal-free conditions[713,714] were developed to achieve efficient desulfurisation. More recently, *in situ* ligation-desulfurisation approach was also reported.[715,716]

Desulfurisation-based methodology of the NCL expanding towards other amino acid junctions have contributed in many ways to prepare proteins and posttranslationally modified analogues for biochemical and structural analyses ever since its introduction. However, despite this broad utility, carrying out desulfurisation of the linkage site in the presence of other Cys residues in the protein sequence usually requires using of protecting groups.[712,717] Not only does this necessity represent an undesirable step to protein synthesis, but it also implies some limitations on the applicability of the approach, mainly due to the solubility issues.

Developed by Dawson and collaborators[718] the protocol for selective reduction of selenocysteines (Sec, U) in the presence of cysteine allowed overpassing the need for protection of thiolates and expanded the already established field of Sec-mediated native chemical ligations (see Section A 2.6, page 41).[719-722]

The sensitivity of Sec peptides to reduction was noted in several works on selenocysteine ligations,[720,721,723] During their preceding work on the synthesis of seleno-glutaredoxin 3 analogues (Se-Grx3), Dawson and associates[724] have observed this incompatibility of Sec-containing proteins and peptides with TCEP-assisted native chemical ligation due to the generation of significant levels of a deselenised side product. In the publication to follow, the authors have successively applied the ligation-deselenisation strategy on a model peptide system. Accordingly, *N*-terminal Sec-peptide$_1$ (**U**GLEFRSI-amide) prepared in the form of a diselenide dimer was ligated to the thioester peptide$_2$ (Ac-LYRAG-SR) (Fig. 79) to produce the deselenised alanyl-peptide (Ac-LYRAG**A**GLEFRSI-amide) after the treatment with 50-fold excess TCEP at pH 5.5. Importantly, an excess of 200 mM 4-mercaptophenylacetic acid (MPAA) was needed for the ligation step. MPAA is both served as a catalyst to activate the alkyl thioester and as mild reducing agent to generate a small pool of free selenol to facilitate the ligation reaction.

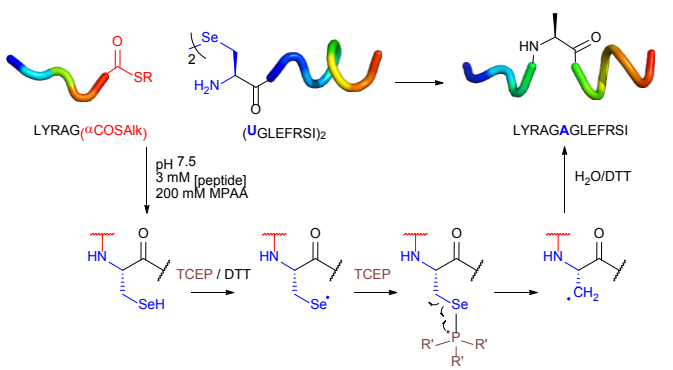
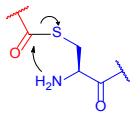
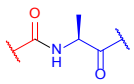
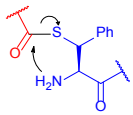
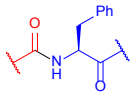
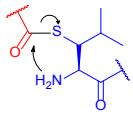
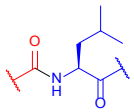
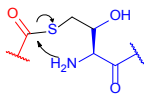
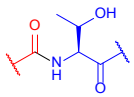
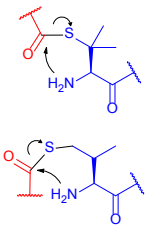
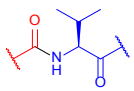
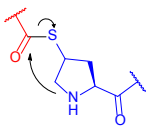
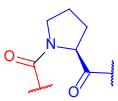
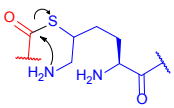
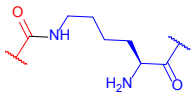
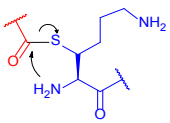
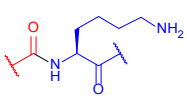
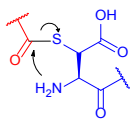
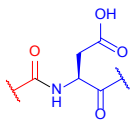
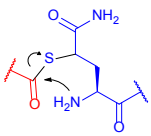
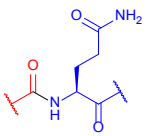
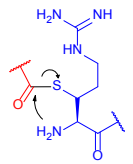
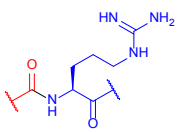
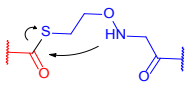
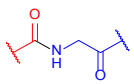


Fig. 79. Traceless ligation of peptides using selective deselenisation described by Dawson and collaborators.[718]

Table 2. Desulfurisation-based NCL methodologies.

Entry	AA	Intermediate	Product
1	Ala		
2	Phe		
3	Leu		
4	Thr		
5	Val		
6	Pro		
7	Lys_{ε}		
8	Lys_{α}		
9	Asp		
10	Gln		
11	Arg		
12	Gly		

This selectivity of reduction with TCEP is hypothesised to be due to the weaker Se – C bond compared with the S – C bond coupled with higher propensity of selenols to form radicals.[725] It should be mention, however, that albeit upon heating, cysteine can be desulfurised in the absence of a radical initiator when treated with excess phosphine.[726]

Selenocysteine-mediated NCL deselenisation procedure have been recently exploited for selective ligation of selenolphenylalanine,[727] and γ-selenolproline,[703] easily transformable into peptides phenylalanine and proline respectively at the ligation site by treating Se-containing intermediates with TCEP or DTT. Analogously to N → S acyl transfer (Section A 2.3.1.3, page 36), N → Se acyl shift was recently observed by Adams and Macmillan and allowed NCL to take place at lower temperatures and on shorter time scales.[728] Corresponding selenoesters can be readily accessible by direct solid phase synthesis.[729]

NCL principles of S → N acyl transfer found their application in ligation assisted by proximity effect (see Section A 5, page 42),[730] allowing for conjugation of *N*-terminal residues other that cysteine by auxiliary-mediated acyl transfer. Furthermore, the Cys side chain thiolate introduced during NCL can also provide a synthetic handle for further functionalisation using cysteine-selective methodologies (see Section A 1.2, page 11). It was recently demonstrated by Fang and co-workers that more accessible than thioesters simple phenyl esters could undergo native chemical ligation smoothly under the promotion of imidazole.[731] Lastly, recently reviewed by Monbaliu and Katrizky[732] Kemp's template-mediated thiol ligation,[733-741] Tam's ligation by thiol/disulfide exchange,[742-744] and other auxiliary-driven extensions of native chemical ligation,[745,746] represent significant importance in the field of protein synthesis and can be considered as appropriate for bioconjugation.

2.3.2 Thiazolidine formation

Further expanding the applicability of the reaction of cysteine with formaldehyde described by Ratner and Clarke,[747] Tam and collaborators have elaborated a method for the selective conjugation of *N*-terminal cysteines with aldehydes, resulting in obtaining of stable thiazolidines.[619,626,748-750] The reaction of 1,2-aminothiols readily occurs at slightly acidic pH of 4-5, while the concurrent reaction of free amines with aldehydes results in obtaining Schiff bases reversibly under the same conditions.

The thiazolidine-mediated ligation was first applied by Tam and associates to the preparation of peptide dendrimers[748,750] by attaching unprotected peptide dendrones containing Cys residues at their *N*-termini to a branched core matrix with aldehyde functions. Botti *et al.*[749] have transformed this approach to a general method for the preparation of cyclic peptides. Villain *et al.* have demonstrated that obtained thiazolidines (sometimes referred to as pseudo-prolines)[751] can be selectively cleaved by adding hydroxylamine derivatives, which react with aldehyde functions protected under the form of thiazolidine to form oximes. The authors applied this methodology for the covalent

capture of proteins possessing *N*-terminal Cys or Thr residues (Fig. 80).[752] Interestingly, under the same conditions *N*-terminal Ser residues reacted only poorly.

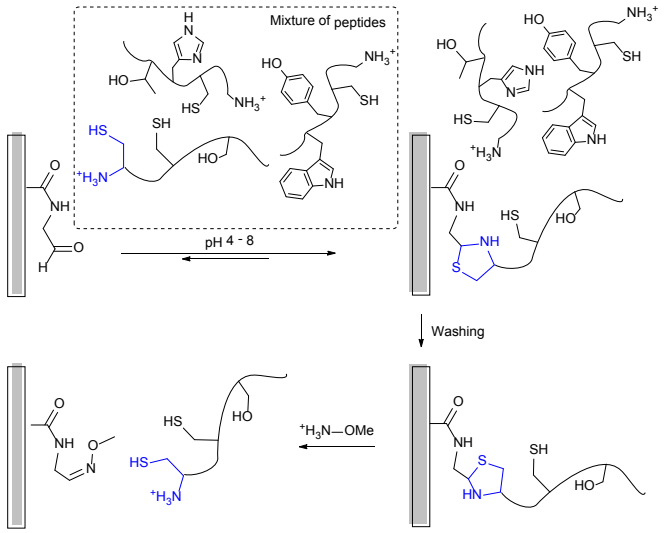


Fig. 80. General scheme of covalent capture purification of *N*-terminal cysteine containing proteins developed by Villain *et al*.[752]

Because of the recent advances in the semisynthesis of proteins and the encoding of 1,2-aminothiols into recombinant proteins,[645,753,754] thiazolidine-mediated conjugation is now experiencing a reappraisal of its potential for bioconjugation.[755,756] For instance, Casi *et al*.[755] have exploited thiazolidine formation for the preparation of antibody-drug conjugates by site-specific incorporation of a potent drug, containing an aldehyde moiety, to engineered recombinant antibodies displaying a Cys residue at their *N*-terminus, or a 1,2-aminothiol at their *C*-terminus.[755] Lastly, thiazolidines represent one of the most often used *N*-terminal cysteine protecting groups for sequential NCL (see Table 1, Section A 2.3.1.1, page 34).

2.3.3 2-Cyanobenzothiazoles (CBT)

The reaction of 2-cyanobenzothiazole (CBT) with D-cysteine was first conducted by Field and collaborators[757] for the preparation of synthetic luciferin: a compound found in various living organisms and responsible for emitting light after being oxidised by a specific enzyme luciferase. Ever since its isolation (9 mg from 15000 firefly lanterns!),[758] luciferin enzymatic oxidation has been studied for the last 50 years.[759]

The regeneration pathway of luciferin in firefly was found to be consisted of the condensation of CBT with D-cysteine (Fig. 81).[757] The reaction mechanism underlying this addition include first attack of cyano group of CBT by cysteine thiolate. This results in the formation of the electrophilic imidothiolate, subjected to the second attack by cysteine amino group to form thiazole structure after the yielding of ammonia gas.

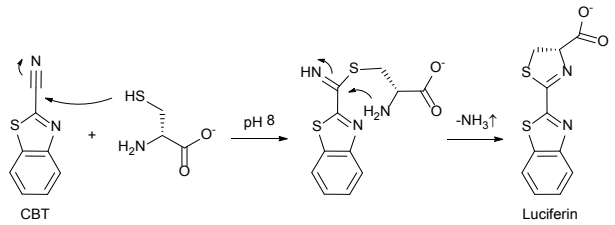


Fig. 81. Synthesis of lucifering by the reaction between cyanobenzothiazole (CBT) and D-cysteine.[757]

Inspired by these early works, Rao and co-workers[760] have further investigated the reaction of cyano-substituted aromatic compounds with amino-thiol substrates. They have demonstrated that benzotriazole motif plays an important role for the activation of the cysteine addition to a nitrile group. For instance, under optimal reaction conditions (PBS, pH 7.0-7.5) its replacement by other aromatic fragment such as picolinonitrile or benzonitrile largely decreases the reaction yield. All naturally occurring amino acids are unreactive towards CBT, except for cysteine owing the highest second-order rate constant among six other tested amino-thiol substrates (9.19 M^{-1} s^{-1}, which is significantly higher than these of the majority of biocompatible click reactions).[761] Finally, the efficiency and specificity of CBT-based labelling of terminal cysteine residues was demonstrated on proteins *in vitro* (Fig. 82) as well as on cell surfaces.

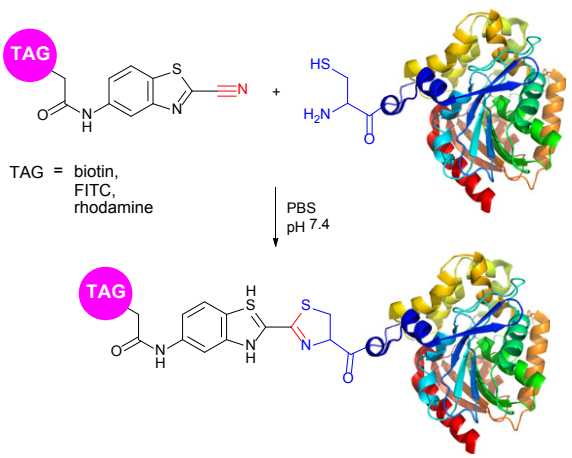


Fig. 82. Labelling of *Renilla* luciferase (pdb: 2PSD) by CBT-based probes.[760] *N*-terminal cysteine was generated by protease processing of the fusion protein.

In the publications to follow, Rao and colleagues have extended the applicability of CBT towards biocompatible condensations to create polymer assemblies *in vitro* and in living cells under the control of either pH, disulfide reduction or enzymatic cleavage.[762,763] Yuan *et al*. have taken advantage of this approach and developed a method for the determination of glutathione (GSH) concentration *in vitro* and in HepG2 human liver cancer cells.[764] Jeon *et al*.[765] have elaborated a CBT-based ^{18}F-probe radiolabelling of *N*-terminal cysteine-bearing peptides and proteins. Two labelled substrates: a dimeric RGD-peptide - [^{18}F]CBTRGD$_2$, and *Renilla* lucifierase bearing a cysteine at *N*-terminus, have been synthesised with excellent radiochemical

yields and shown good *in vivo* molecular PET imaging efficiency. Proceeding efficiently at physiological conditions, CBT-mediated *N*-terminal Cys conjugation represents a useful alternative to existing approaches for protein labelling.[766]

2.4 TRYPTOPHAN

2.4.1 Sulfenylation-coupling

Encouraged by early reports from Scoffone and colleagues, who examined the site-selective modification of the nucleophilic 2-position of the tryptophan indole ring through electrophilic sulfenylation with various sulfenyl chlorides,[767,768] Payne and collaborators have recently brought back into life a classical reagent for Trp selective modification - 2,4-dinitrophenylsulfenyl chloride (DNPS-Cl).[769] The authors have demonstrated that, in acidic conditions, all nucleophilic amino acid side-chains except tryptophane either remained unmodified, as in the case of serine, threonine and the ε-amino groups of lysine, or were reversibly modified in the case of cysteine, which forms an easily reducible asymmetric disulfide. Further thiolytic cleavage of the resulting 2-Trp thioether derivatives with an external thiol nucleophile gives the corresponding 2-thiol Trp derivatives (2SH-Trp, Fig. 83a). Interestingly, being placed on the *N*-term of the peptides, 2SH-Trp scaffolds were found to enhance the kinetics of the native ligation with peptide thioesters, and could thus serve as Nα acyl transfer auxiliaries (see Section A 2.3.1, page 33).[770]

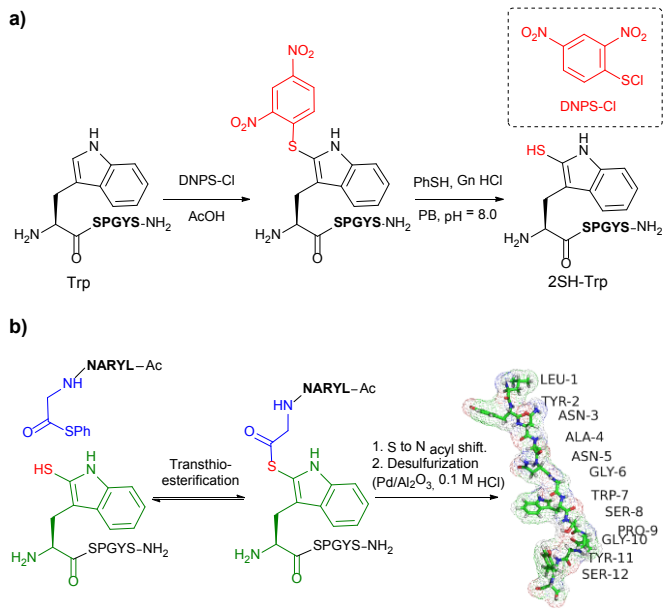


Fig. 83. Ligation–desulfurisation at tryptophan, reported by Payne and collaborators.[769] **a)** Electrophilic sulfenylation of Trp in acidic conditions with DNPS-Cl and the obtention of the corresponding 2SH-Trp derivative. **b)** Auxiliary-assisted native ligation of *N*-terminal 2SH-Trp peptide (2SH-**W**SPGYS-NH_2) with a thioester of the peptide Ac-LYRAN**G-SPh** resulting in 12-mer peptide Ac-LYRAN**GW**SPGYS after the desulfurisation step.

The proposed mechanism for this approach is mechanistically similar to NCL. It was hypothesised that the reaction would proceed *via* an initial step of the peptide thioester transthioesterification with an indole 2-thiol functionality followed by an S- to *N*-acyl shift through a 7-membered ring transition state to generate a native amide bond. The last step of 2-thiol Trp desulfurisation results in obtention of a ligated product with only naturally occurring amino acid residues (Fig. 83b). Although this methodology represents a clever chemoselective approach for the ligation of completely unprotected peptide fragments through Trp moiety, the harshness of the reaction conditions of sulfenylation and desulfurisation limit it only to peptide substrated and don't allow its application on complex biomolecules.

2.4.2 Pictet-Spengler reaction

Another approach for the ligation of unprotected peptides was proposed by Li *et al.*[771] and exploited the Pictet-Spengler reaction: an acid-catalysed intramolecular condensation between an iminium ion and an aromatic *C*-nucleophile described in 1911.[772] This non-natural ligation (peptides are linked by non-natural bonds) involves reaction of two peptide partners in acetic acid: one containing a Trp residue at its *N*-term and another with a *C*-terminal aldehyde function. The latter should generally be introduced by means of solid-phase peptide synthesis on acetal resin (described by the same authors).

2.4.3 *N*-Acyl tryptophan isopeptides

Lastly, an interesting example of native *N*-terminal Trp ligation, mechanistically very similar to NCL (see Section A 2.3.1, page 33), was recently reported by Popov *et al.*[773] The key intermediates of this methodology – *N*-acyl tryptophan isopeptides – undergo selective acyl transfer to yield natural peptides. These are, however, not accessible directly by methods reported so far, which substantially restricts the applicability of the methodology.

2.5 HISTIDINE

Despite being known for its particular importance for the acyl transfer in many enzymatic processes,[648] histidine has been rarely used for bioconjugation. The only study describing the ligation of *N*-terminal His peptides with activated thioester was reported by Zhang and Tam.[774]

Ellman's regent[347] was used to activate the *C*-terminal peptide thiocarboxylic acid by forming acyldisulfide derivative, which is then nucleophilically attacked by *N*-terminal histidine. Captured by the imidazole of the *N*-terminal histidine, the obtained N_{im}-acyl intermediate is hypothesised to undergo $N_{im} \rightarrow N_{\alpha}$ shift to form histidine at the ligation site (Fig. 84). However, the N_{im}-acyl intermediate has not been isolated and it is quite possible that regioselectivity is obtained simply because of anchimeric assistance of the proximal imidazole moiety at the ligation site.

Interestingly, no sign of coupling reaction has occurred when a corresponding non-activated *C*-terminal thiocarboxylic acid is participating in the reaction instead of the acyldisulfide. The reaction pH plays an important role on the effectiveness of the

reaction. Only when maintained at slightly acidic values (pH 5-6) and in the absence of the thiol nucleophiles, would the imidazolyl moiety of histidine be the sole nucleophile present in the polypeptide. This methodology has been applied to generate histidine-containing peptides with yields up to 75 %.

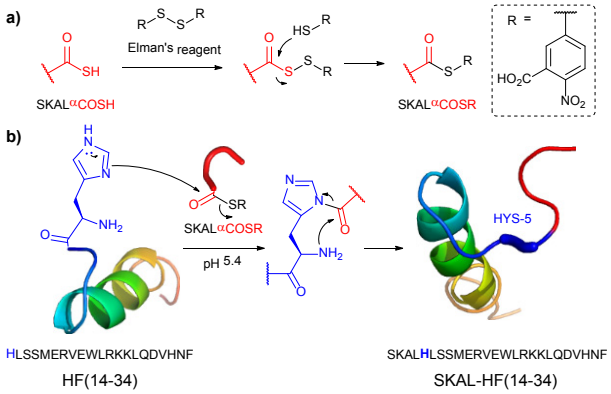


Fig. 84. *N*-terminal histidine labelling described by Zhang and Tam.[774] **a)** Mechanism of thiocarboxylic acid activation by Ellman's reagent.**b)** Labelling of bovine parathyroid hormone fragment (14-34, pdb: 1ZWC) with activated tetrapeptide thioacid.

2.6 SELENOCYSTEINE

Selenium and sulphur belong to the same main group of elements; therefore, 21st proteinogenic amino selenocysteine and Cys residue exhibit rather similar properties in terms of reactivity for bioconjugation.[775-777] For instance, both Hivert's[721] and Raines's group[719] demonstrated that *C*-terminal peptide thioesters react smoothly with peptide fragments containing a *N*-terminal selenocysteine in exactly the same manner as with corresponding cysteine analogues. Presumably proceeding through the same mechanism as NCL (Section A 2.3.1, page 33), the first step of the ligation process consists in the nucleophilic attack on the thioester by selenolate to give a selenoester intermediate that subsequently rearranges to give a native chemical bond.

Sec ligation can be chemoselective when conducted at slightly acidic pH. Low intrinsic pK_a values of selenocysteine (5.2)[778] and consequently its higher dissociation level at low pH, endows this amino acid with unique biochemical properties, allowing regiospecific covalent conjugation with electrophilic compounds in the presence of the side chains of all other natural amino acids including the thiol group of Cys (pK_a 8.3). For instance, the reaction rate with selenocysteine was found to be 1000 fold faster than with cysteine at pH 5.0.[719] Moreover, the lower pH generally suppresses β-elimination of the selenol group from selenocysteine resulting in the obtaining of unreactive dehydroalanine.[384]

Initially, considerable efforts were made to show the applicability of selenocysteine NCL for the preparation of selenium-containing derivatives of enzymes and benchmarking activities thereof. Hilvert and associates synthesised a C38U analogue of bovine pancreatic trypsin inhibitor (BPTI). Amusingly, the wild-type BPTI and its artificial analogue folded into alike conformations and demonstrated similar inhibiting affinity of trypsin and chymotrypsin.[721] Raines *et al*. selected 124-residue ribonuclease A (RNase A) as a model protein for the investigations.[719] DNA recombinant technology was utilised to prepare a *C*-terminal thioester fragment corresponding to residues 1-109, while standard solid phase peptide synthesis methodology was used to obtain a *N*-terminal Cys and Sec peptides corresponding to residues 110-124 (Fig. 85). Just as in the case of BPTI, the semisynthetic wild-type RNase A and C110U RNase A presented equivalent ribonucleolytic activities. Further advances in the field of Sec-NCL have resulted in synthesis and investigations of other different proteins such as seleno-glutaredoxin,[779] azurin,[780,781] and thioredoxine reductase

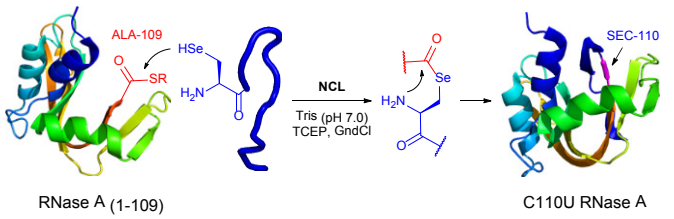


Fig. 85. Selenocysteine native chemical ligation applied for the preparation of C110U mutant of RNase A (wild-type - pdb: 7RSA).[719]

Ease of the post-ligation transformation of selenocysteine to alanine (by deselenisation), dehydroalanine (by β-elimination) or non-natural amino acids (by addition reaction to dehydroalanine, see Section A 1.3.9, page 17) became a spur to a further propagation of Sec-mediated methodologies as very effective tools in rational design of peptides and proteins. Quaderer and Hilvert[782] exploited such transmutations of selenocysteine to access a series 16-residue cyclic peptides (Fig. 86).

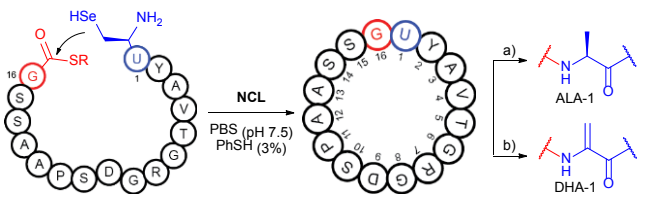


Fig. 86. Synthesis of cyclic 16-resiudes selenopeptides by native chemical ligation and following transformation thereof to a) alanine (Raney Ni, H_2, TCEP, 20% AcOH); b) dehydroalanines (H_2O_2, 10% ACN).

In this initial report, the deselenisation step was conducted rather harshly (Raney Ni, H_2), implying that all Cys residues (if there were any) would have been reduced as well. Recently, however, Dawson and collaborators[718] have demonstrated that selenocysteine can be chemoselectively deselenated with TCEP in the presence of cysteines. This allowed overpassing the main limitation of the NCL-desulfurisation strategy (see Section A 2.3.1.4, page 37), namely the inability to control regioselectivity of desulfurisations if several cysteines are present in the peptide or protein, and yielded in a pool of NCL-deselenisation strategies for mild incorporation of alanine, phenylalanine and proline into the ligation site by classic NCL approaches (see Table 2, page 38). Finally, selenocysteine peptides were found

to efficiently undergo reverse NCL at acidic pH and thus to be of particular interest for the generation of thioesters by sequential N → Se acyl-transfer and substitution of the obtained selenoester by exogenous thiol (see Section A 2.3.1.3, page 36).[728]

Because the incorporations of selenocysteine by the cell translational machinery are generally very laborious,[783,784] selenopeptides are largely obtained by SPPS.[719,720,777]

3 *C*-TERMINAL CONJUGATION

Chemical methods for *C*-termini conjugation are rather scarce. Definitely, previously mentioned approaches for the conjugation of side-chain carboxylates (see Section A 1.5, page 21) can also be applied effectively to modify the terminal carboxylates. The selectivity, however, would be the main issue, for up to date there is no described method for somewhat selective activation of *N*-terminal carboxylates by any activating reagent.

Because of the impetuous development of NCL (see Section A 2.3.1, page 33), requiring *C*-terminal thiolates as reacting partners for *N*-terminal Cys proteins, those became widely accessible, namely by means of SPPS. A perspective approach, exploiting these advances was proposed by Goody[785] and collaborators, who developed a protocol for selective transformation of *C*-terminal thioesters to corresponding hydroxylamines, enabling thus the application of aldehyde- and ketone-selective methodologies on the *C*-terminus.

On the other hand, the unique position of protein *C*-termini has stimulated numerous efforts to target this location, which resulted in numerous enzymatic and intein-based approaches for *C*-terminal-selective protein modification.[786-791] These methodologies are, however, not covered by this review devoted to chemical methods of bioconjugation.

4 SEQUENCE-SELECTIVE APPROACHES

Several especially useful methodologies in bioconjugation exploit not a specific property of the residue, or its peculiar position in protein, but rather the synergetic effect of a batch of neighbouring amino acid residues. For instance, an example of such selectivity is aforementioned selective modification of His-AA-Ser and His-AA-Thr peptide motifs (see Section A 1.2, page 11) by usually promiscuous activated esters. In this case, the His imidazolyl side chain located in close proximity to Ser/Thr side chains increased the reactivity thereof towards electrophiles.

Metal-chelation methodologies are perhaps the most elaborated among sequence-selective approaches (Fig. 87).[792] Oligohistidine sequence (usually H_6), called His-tag is known to interact robustly with transition-metal complexes, including nitrilotriacetic acid (NTA) complex of Ni^{2+}, thereby the sequence is widely used for protein purification by affinity chromatography. Similarly, oligo-aspartates (most often $(\mathbf{D}_4)_n$, n = 1-3) were developed for selective labelling with multinuclear Zn^{2+} complexes.[793,794] Tetracysteine motif (**CCPGCC**) was reported to especially effectively chelate biarsenical dyes such as 4',5'-bis(1,3,2-dithioarsolan-2-yl)fluorescein (FlAsH),[5,795] while similar structurally tetraserine sequence (**SSPGSS**) was found to be selective for diborate-scaffolds.

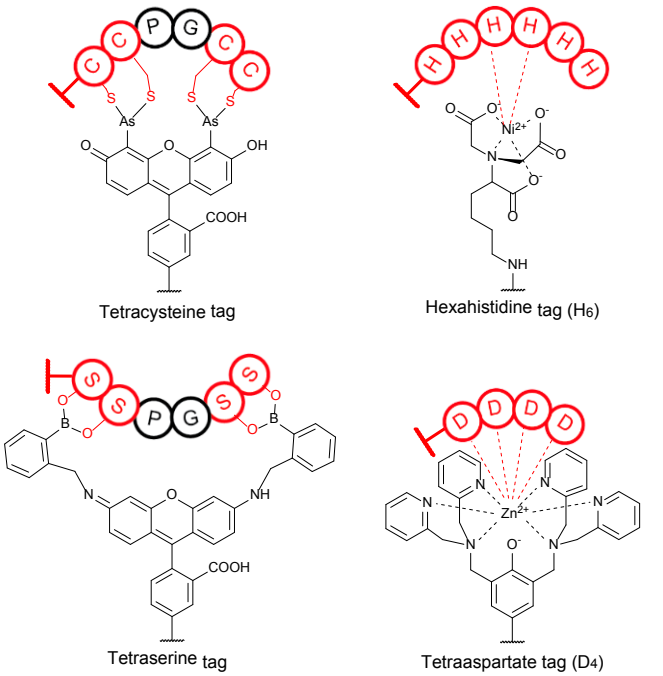


Fig. 87. Strategies for the selective conjugation of proteins based on metal-chelation: tetracysteine/biarsenical system, oligohistidine/nickel-complex system, tetraserine-borate system, oligo-aspartate/zinc-complex system.[27]

5 PROXIMITY-DRIVEN MODIFICATIONS

All above-mentioned methodologies are mainly residue-specific. That is to say, they exploit specific reactivity of the functional group of interest or of an assembly of residues. As a result, bioconjugations of highly nucleophilic Cys, Lys, and Tyr residues with electrophilic reagents are definitely prevalent among described methodologies.

The inherent reactivity is rare to be overcome, which largely limits of the scope of known methods for bioconjugation of the amino acids possessing low nucleophilicity. However, bringing the reaction partners into a close proximity can accelerate a reaction thereof, which would not otherwise be possible because of the presence of other more reactive species. Routinely exploited by enzymes, this approach enables selectivity on the basis of molecular shape rather than reactivity or local environment (Fig. 88).

Developed by Hamachi and collaborators in 2006,[448] the post-affinity-labelling approach can be used for the selective tethering a functional molecule at the proximity of the active site of enzymes. This method was then used on numerous substrates and enlarged the scope of selective conjugation especially on histidine and tyrosine (see Section A 1.5.1, page 21).[448,449,472,796]

Despite these successes, a prevalent limitation of the applicabil-

ity of the method due to the possible modification of the residues situated only in the vicinity of ligand-binding pocket of the target protein prevents this methodology from becoming a general approach for protein labelling.

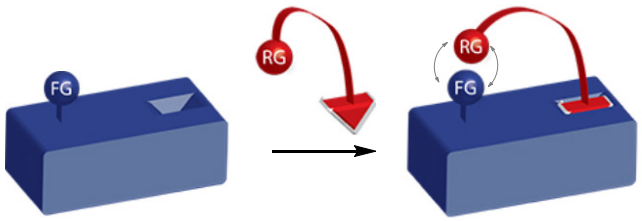


Fig. 88. Schematic representation of the proximity-driven approach for selective modification of macromolecules. Combining of promiscuous reactivity of the reactive group (RG, shown in red) with specific molecular recognition results in specific modification of the functional group of interest (FG, shown in blue).

In 2010, exploiting similar idea, Popp and Ball[441] envisioned the combination of two previously described techniques: the coiled-coil based molecular recognition of complementary peptides[442,797-799] and high catalytic activity of dirhodium complexes on carbene C—H insertion, previously reported by Francis and collaborators[440] for the selective modification of tryptophan (see Section A 1.4.2, page 19). Two complementary peptides: one, containing a dirhodium catalytic centre (precomplexated through two glutamate side chains), and another, containing a side chain to be modified are thus involved in this methodology (Fig. 89a). Because of the compelled proximity of the side chain of interest and the active catalytic centre in the obtained supramolecular assembly, the reaction of rhodium-catalysed C—H insertion is largely accelerated (more than 10^3 times).[800] As a result, the conjugation of amino acid residues with lower reactivities becomes possible. This allowed to expand the scope of originally tryptophan-selective dirhodium carbene methodology first to the other aromatic residues, phenylalanine and tyrosine,[441] and then to over half of the naturally occurring amino acid residues (Fig. 89b).[800] To date, dirhodium metallopeptide represent the only reported method for selective modification of Gln, Asn, and Phe side chains. The authors have also demonstrated the possibility to apply their methodology on chimeric proteins, containing fused coils,[443] as well as on full-sised natural proteins possessing coiled-coils in their structures.[800]

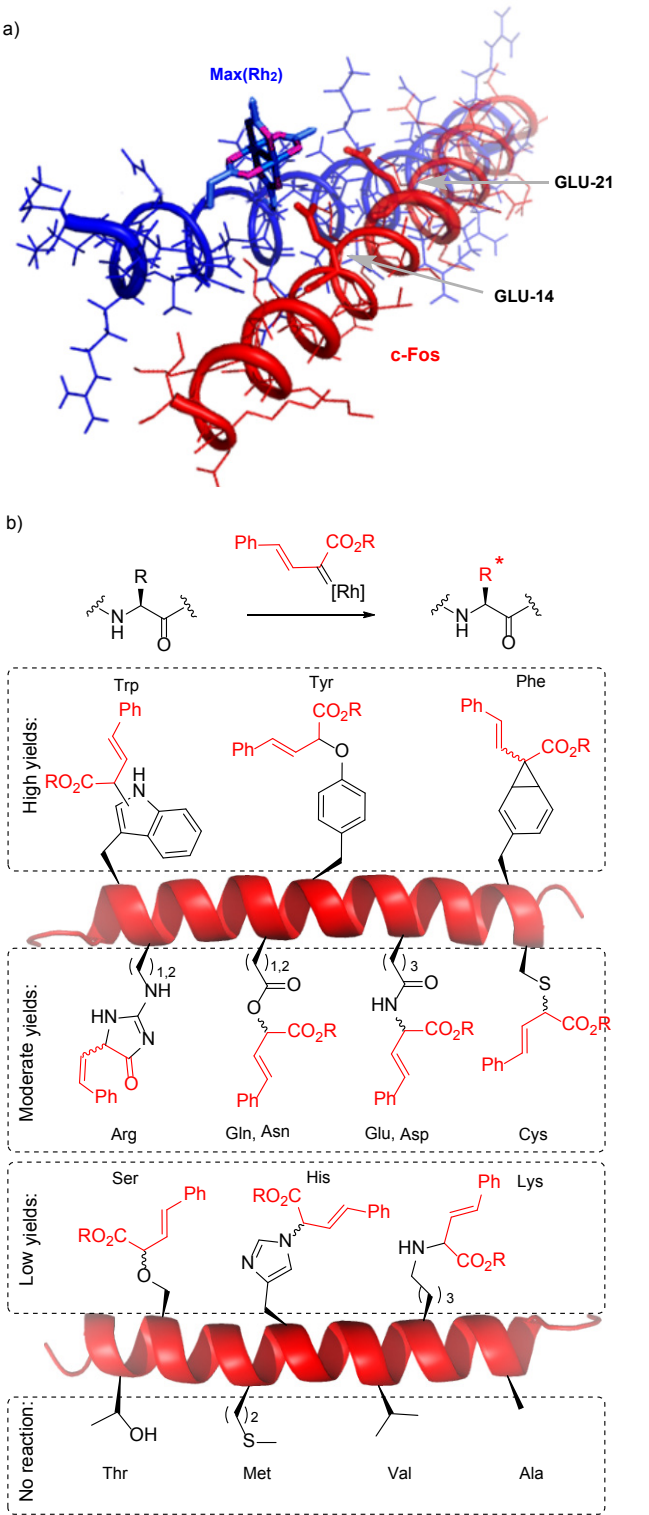


Fig. 89. **a)** Modification of c-Fos (shown in red) catalysed by Max(Rh_2) (shown in blue) metallopeptide described by Popp and Ball.[800] Two possible coiled-coil alignments result in modification of GLU-14 and GLU-21 residues, located in close vicinity to dirhodium catalytic center. b) Proposed product bond connectivity and

Despite its important potential, the metallopeptides methodology is however not devoid of drawbacks. Because binding to dirhodium is nonselective and thus cannot be performed in the presence of other carboxylate-containing peptides, rhodium-peptide complexes must be synthesised beforehand, which is often challenging mainly due to their poor solubility.[801] Moreover, the method is restricted to proteins containing coiled coil fragments in their structures, which in case of the vast majority of targets would mean the need for resource- and time-consuming expression of fused proteins.

Another approach, developed by Silverman and colleagues, exploits self-assembling of complementary DNA to bring into proximity two reacting fragments and allowed, although only on simple substrates, selective phosphorylation of tyrosine and serine, otherwise not feasible.[524]

Beyond coiled coils and DNA-based preorganisation, the principles of proximity-driven selectivity should be extended to other helix-binding protein domains and to biological molecular recognition generally. A significant broadening of the applicability of this elegant approach for protein modification, biochemistry and biomaterials engineering is anticipated in the nearest future.

6 CONCLUSIONS

The field of bioconjugation has expanded in the last 100 years and passed from a blind-eye modification of proteins one could found in nature to a well-established independent domain full of approaches allowing precise and reliable introduction of various tags into proteins' structure.

Many, if not most of these methods, however, often possess drawbacks limiting their general applicability. This fact has become of especial consideration with a rise of novel exigent applications for bioconjugation, namely the preparation of new therapeutic conjugates, vaccines, and biomaterials.

Several parameters of the mode of conjugation, previously completely neglected, were recently revealed to be of paramount importance. For instance, the stability of the generated linkage and the distribution of products generated upon conjugation can be determining for the overall efficiency of the conjugate.

Considering the state of the art in bioconjugation, we decided to start a project which aimed at the development of general approaches allowing discovery of new bioselective chemical functionalities (both residue-selective and sequence selective). After having defined the main requirements of these screening systems to be simple, reproducible, and effective, and conducted a thorough literature survey we could finally get a clear project in hand and this is to its first part that we now turn.

B. DISCOVERY OF RESIDUE-SELECTIVE REACTIONS

This chapter is devoted to the design and applications of the screening system for the discovery of new residue-selective ligation reactions. The approach was envisioned to make the screening more efficient and less laborious. First, reactive probes of interest, which were selected based on literature survey, are tested on a series of UV-detectable amino acid derivatives using simple HPLC technique. The most promising among tested reagents are then subjected to structure-reactivity study in order to ameliorate their properties and to define the appropriate anchor points for further insertion of the functional tags. The tagged reactive probes are then benchmarked on a model peptide mixtures obtained by trypsin digestion of lysozyme.

As it was previously mentioned, the screening for new bioselective reactions is an important challenge in modern chemistry. The complexity arises from the fact that, in order to be appropriate for bioconjugation, the reactive probe of interest must meet several severe requirements; namely it should possess:

- High selectivity towards given target: a specific amino acid residue, a sequence of amino acids, or a position in a protein (*N*-terminus, *C*-terminus, in-chain).
- Applicability at physiologic conditions (aqueous media, 4 - 37 ˚C, physiological pH).
- High conjugation efficacy: the reaction should be fast and exquisite.
- Biocompatibility and high enough stability of the starting material as well as the obtained conjugates.

In order to be able to easily assign tested chemical probes with relative ranks according to these requirements, we decided to start with elaborating a general screening system.

1 DESIGN OF THE SCREENING SYSTEM

The three main steps of the screening were envisioned in order to minimise time and work expenditures, which include (1) a simple HPLC-based screening for the profiling of the stability and reactivity of a reactive probe in the reaction with a series of representative UV-detectable amino acid derivatives; (2) structure-reactivity study of the most promising among tested probes; (3) validation of the probes' selectivity in close to biological conditions on model peptide mixtures. The current section is devoted only to the theoretical basis for each step including design, optimisation and probable difficulties one can encounter during the screening. The practical application on a set of selected reactive probes is given in the section to follow (Section B 3, page 50).

1.1 UV-TRACEABLE AMINO ACID DERIVATIVES

After having defined the criteria that any new selective ligation reaction would have to fulfil in order to be appropriate for bioconjugation, we decided to start with an establishment of a simple HPLC setting allowing determination of the potential of a testing reactive group. In order to achieve the objective, a pool of UV-detectable amino acid derivatives, namely their benzylamides (Fig. 90), was prepared.

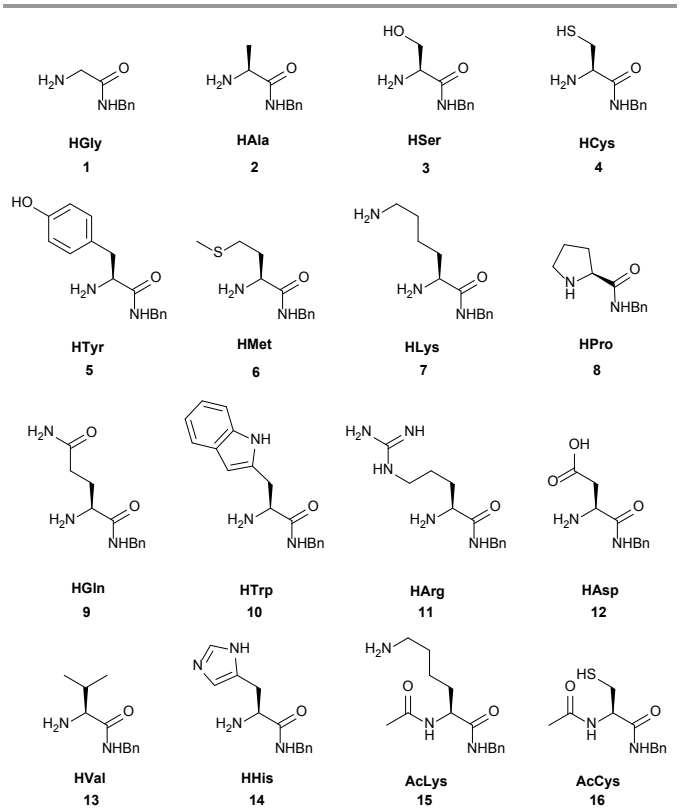


Fig. 90. Series of UV-traceable amino acid derivatives used in the HPLC-screening.

This series was composed from the reagents **1-16** representing all reactive functionalities which can be found in essential amino acid side chains. The presence of the UV-traceable benzyl moiety made it possible to easily evaluate the reactions' advancement by measuring the conversion either of starting materials or yielding products using simple quantitative HPLC technique. Because of anticipated solubility issues, the majority of the model compounds (**1-14**) corresponded to *N*-terminal amino acid derivatives, which generally possess better solubilities due to the presence of protonatable amino groups in their structures. For some tested reactive probes possessing ambiguous reactivity profiles, in-chain model derivatives (**15-16**) were used to test possible discrepancy in reactivity (Section B 3, page 50).

The reagents of interest are then subjected to the reaction with each among model amino acid derivative (**1-14**), as well as to the study of their hydrolytical stability in aqueous media. Obtained results being characteristic for the reactive probe represent its reactivity profile (see below). For the purpose of simplifying the representation and making it more illustrative, the results are represented in colour-coded form in this manuscript. Unless otherwise specified, the reaction advancement is monitored by HPLC using the conversion of UV-traceable reactive probe and the appearance of UV-traceable reaction product as a reference.

Several types of reactivity profiles were expected to be observed during the screening. The reactive probe was anticipated to fall into one of five possible cases (Fig. 91). It could thus be:

- Unreactive (the probe is hydrolytically stable, but no reaction is observed with any amino acid derivative).
- Unstable (the probe hydrolyses too fast and gives no coupling product with any of amino acid residues).
- Selective (the probe is stable enough and reacts preferentially with a specific amino acid derivative without producing any by-products).
- Promiscuous (the probe is reactive towards all amino acid derivatives).
- Intermediate (the probe represents some degree of selectivity; can be regarded as a combination of previously mentioned cases).

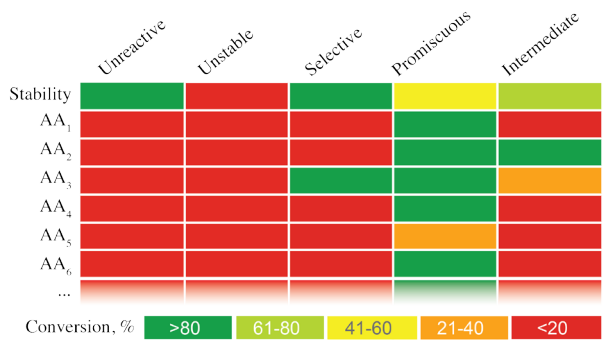
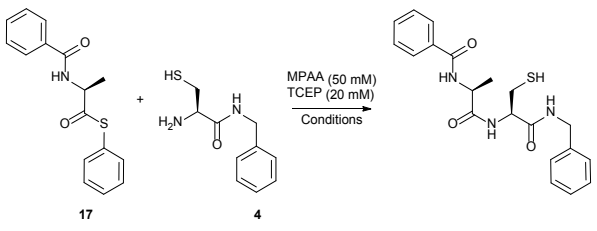


Fig. 91. Possible reactivity profiles expected to be obtained during the screening. Field stability corresponds to the percentage of the non-hydrolysed probe in aqueous media (100% - % of hydrolysis).

Before starting the design and synthesis of the reactive probes of interest we first decided to choose the reference technique to provide a threshold in terms of the reactivity appropriate for bioconjugation. Native chemical ligation (NCL, see Section A 2.3.1, page 33), one the most popular conjugation methodologies used for the preparation of native proteins and peptides, was chosen as a reference reaction on which we defined standard conditions such as concentrations of reacting partners, buffers that can be used as reaction media, the necessity of using the internal standard, and the timespan at which the screening should be conducted.

Thioester alanine derivative **17** and cysteine model probe **4** were used in the first trials to evaluate practically various reaction conditions, in order to better apprehend merits and possible drawbacks of such a well-established conjugation methodology as NCL. After having tested several buffer systems, temperatures and solubilisation additives, we were able to define optimal conditions for conducting the screening. PBS (1x, pH 7.5) was found to be the most appropriate buffer for the reaction to occur at 100 μM – 1 mM concentrations at reasonable timescale (monitoring of the reaction advancement with HPLC for 5 hours). A very low solubility of the resulting dipeptide allowed us to test various solubilisation additives in order to establish general reaction conditions appropriate even for such poorly soluble reacting partners. Among them, dimethylsulphoxide (DMSO, up to 50% depending on the solubility of reagents) has represented the technique of choice.



Conditions: Buffers: PB, PBS, HEPES, TRIS, NH_4HCO_3.
Temperature: 4-40 °C;
Concentrations: 100 μM - 10 mM;
Solubilisation agents: 6 M GuHCl, DMSO, ACN, TFE.

MPAA: 4-mercaptophenylacetic acid.

Scheme 1. Reference reaction used to define the optimal reaction conditions for the screening.

Furthermore, during these first trials we have realised that in order to reliably estimate the conversion level of the reagents by integrating the areas of their peaks in chromatogram traces, the use of internal standard was crucial. Benzamide was chosen as the most appropriate substrate for this role because of its high visibility in UV, water solubility, and low reactivity with both nucleophiles and electrophiles. The mitigation of the discrepancy in the injection volume during HPLC analyses as well as other technical issues were thus resolved by simply normalizing the areas of the peaks of reacting partners on the peak of benzamide.

Lastly, we decided to study possible selectivity profiles by comparing previously mentioned NCL model reaction with a similar transformation supposedly being much less selective. The tested reaction involved two benzoylalanine derivatives: the previously mentioned thioester **17** and the corresponding phenolate **18** (Fig. 92). Indeed, the difference in the reaction selectivities of the two model cases was salient even when conducted under rough reaction conditions with high concentrations of reagents and absence of exogenous thiol additive.

These first results on model reactions were especially important for assessing (and thus being prepared) for the majority of possible difficulties which one can encounter while conducting the screening. As such, occurring simultaneously with the studied reaction, the hydrolysis of both starting materials and resulting products, or their degradation (*e.g.* by oxidation) may strongly

affect, or even completely flaw the obtained results, and therefore should be addressed attentively in every case.

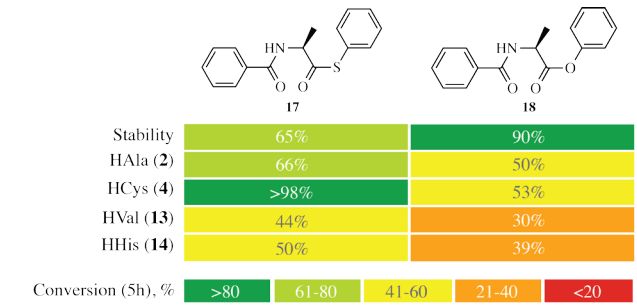


Fig. 92. Comparison of the reaction profiles of the two model reactions: native chemical ligation involving thioester-activated alanine derivative **17** and the reaction with a phenolate-activated analogue **18**. Reaction conditions: PBS (pH 7.5), TCEP (10 mM), 1 mM concentrations of reaction partners. Stability corresponds to the percentage of the non-hydrolysed probe (100% - % of hydrolysis).

1.2 Structure-reactivity study

While non-substituted reactive probes can be used for rough selection of potentially promising among them, further investigations are usually required for improvement of their characteristics and defining the appropriate anchor points for the introduction of other functional entities (so called tags). Very often, even a slight change of the scaffold substituents may result in large modification of its reactivity. Therefore, a structure-reactivity study should first be conducted for defining general trends in the influence of substituents on the reactivity profile of the reactive probe.

HPLC-based screening approach represents an appropriate system for assigning the reagent's reactivity profile towards various amino acid side chains. However, it gives no information about the selectivity of the probe in complex media where various functional groups are present together. Obviously, a simple study can be made by testing the reaction on mixtures of amino acid models. However, there would be a big gap between the results for selectivity, obtained on a mixture of simple amino acid derivatives, and any real biological system. In order to expand the applicability of the results obtained in the first step of the screening, the elaboration of a cheap and an easily producible complex model system to test the reactivity of the selected probes in close to biological conditions was important.

1.3 Tryptic digest of lysozyme

We decided to use tryptic digest (peptide mixtures obtained after the digestion of a protein by trypsin) of a hen egg white lysozyme – an average-length protein consisting of 124 amino acid residues – as such a model system. Lysozyme and trypsin were chosen because of their high accessibility, fineness and availability from various commercial sources at affordable prices. Trypsin's high substrate specificity and good reproducibility of digestion have already made it one the most commonly used enzyme in mass spectrometry. As a result, many automated algorithms are present today and are used for the identification of peptides present in complex mixtures. Tryptic digests are generated by following "Keil rule". [802] Generally speaking, it says, that during trypsin digestion of a protein, cuttings exclusively occur after arginine and lysine. The rule is almost absolute and is not applicable only if proline is situated next to arginine or lysine in the sequence.

In the case of a hen egg white lysozyme, such a digestion results in obtaining a mixture of 59 detectable peptides with an average size of 20 amino acid residues (Fig. 93, see Annex 1 for a full list of peptides generated by *in silico* digestion). The majority of these peptides are detectable by nanoHPLC coupled to a Q-TOF mass spectrometer (Section B 4.5, page 57). Further MS/MS analysis of tagged peptides can provide data allowing, for each peptide to determine the exact number and type of side chains which participated in the reaction, as well as the efficacy of the tagging (comparing the intensity of the target peptide before and after the labelling).

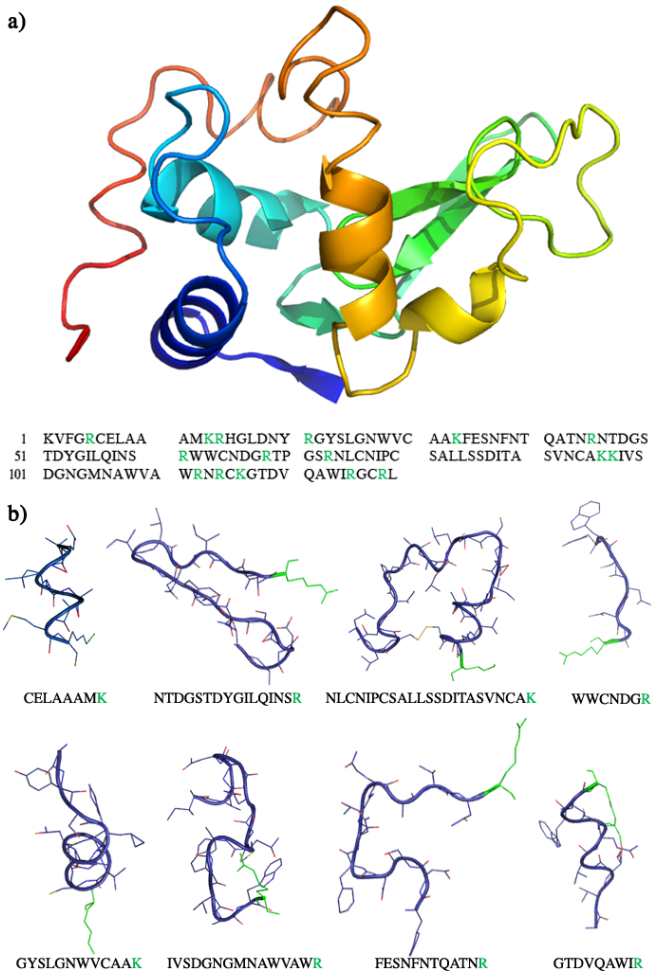


Fig. 93. Schematic representation of the procedure used for generating of the model tryptic digest of a hen egg white lysozyme (pdb: 1LYS). **(a)** Structure and amino acid sequence of the protein with loci of trypsin digestion shown in green. **(b)** Selected examples of the peptides generated during the digestion. The eight most abundant peptides in mass spectra out of 59 predicted by *in silico* digestion (see Annex 1) are depicted.

A rather complex peptide mixture could thus be easily and reproducibly prepared from a non-expensive protein source with no need for specific equipment or expertise. It is noteworthy that the complexity and the constitution of the generated peptide mixtures can be modified by changing the protein substrate used for digestion.

In order to increase the sensitivity of the approach the TMPP-tagging methodology – an approach previously described by Watson and collaborators[803,804] as a highly-sensitive picomole-scale method for charge derivatisation of peptides – was used. For this purpose, a tris(2,4,6-TriMethoxyPhenyl)Phosphonium (TMPP) probe containing a functional group of interest should be synthesised. The permanent charge of TMPP results in increasing of the ion abundance of the peptides to which it is attached during the reaction and, as a result, allows enhancing the sensitivity of the HPLC detection of low abundant side reactions.

Moreover, heighten hydrophobicity of TMPP results in increasing the retention time of modified peptides in reverse-phase HPLC and shifting the peaks thereof to a less encumbered zone, further increasing the sensitivity of the approach (Fig. 94).

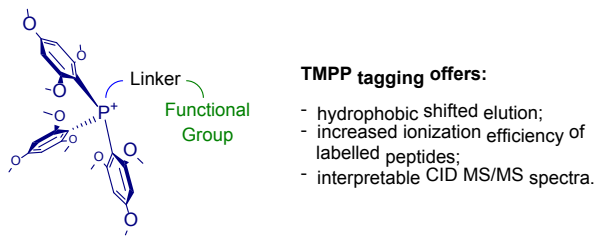


Fig. 94. Structure of TMPP-probe: tris(2,4,6-trimethoxyphenyl)phosphonium fragment (shown in blue) responsible for the improvement of the sensitivity of the detection by increasing the ion abundance and hydrophobic shifting of the modified peptides; functional group of interest (shown in green); and linker connecting the two previously mentioned fragments together.

2 DESIGN AND SYNTHESIS OF THE PROBES

At the time the project started, we decided to define appropriate candidates for bioconjugation by first conducting an extensive survey of literature. As a result, three classes of reagents of interest were identified: imidoyl halides and analogues, which we expected to generate active reactive species *in situ*; a series of activated Michael acceptors envisioned for amino-selective modifications; and a set of electrophiles presumably possessing selectivity towards nucleophilic amino acid side chains. Synthetic approaches used for the preparation of each reagent will be discussed below.

2.1 HYDROXYIMIDOYL HALIDES AND ANALOGUES

The class of hydroxyimidoyl halide and analogues with general structure -C(X)=*N*-OH (Fig. 95) was envisioned because these were known to possess unfamiliar reactivities, namely resulting from the tendency of the electrophilic centre to generating highly reactive nitriloxides *in situ*.[805] Because at the time the project was initiated, no reports on nitrile oxides' reactivity towards nucleophiles were published, we decided to synthesise a pool of different scaffolds and test them in our screening system (Fig. 95).

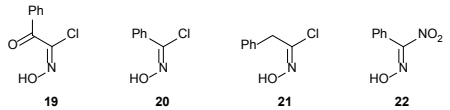
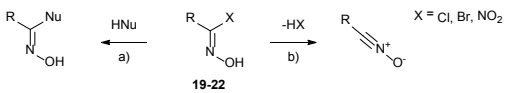


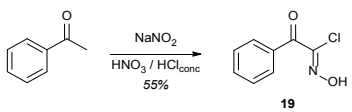
Fig. 95. Hydroxyimidoyl halides and analogous structures used in the screening.

We selected the four structures **19-22** that we decided to test for their reactivities towards amino acid derivatives **1-14**. These scaffolds were of especial interest, because another project devoted to bioorthogonal methodologies was ongoing in our research group. We expected nucleophilic amino acid side chains to react either by attacking the electrophilic hydroxyimidoyl moiety directly (Scheme 2, pathway a) or by reacting with nitriloxides generated *in situ* (Scheme 2, pathway b). Importantly, the electrophiles that would be found unreactive towards tested amino acid side chains could be further tested for their reactivity as possible bioorthogonal reagents.



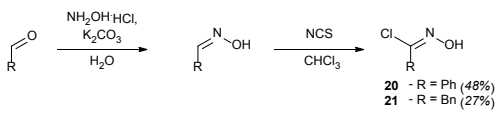
Scheme 2. Two possible pathways for hydroxyimidoyl halides' and analogues' reactivity. **a)** Direct reaction with nucleophiles (HNu). **b)** *In situ* generation of highly reactive nitriloxides.

Probably one of the most reactive among hydroxyimidoyl halides, the probe **19** was synthesised from acetophenone *via* direct oxidative chlorination with the aqueous HCl/HNO_3 mixture (*aqua regia*) using procedure developed by Hamersak *et al.* [806]



Scheme 3. Synthesis of *N*-hydroxybenzimidoyl chloride (**19**).

Both phenyl- and corresponding benzyl-substituted hydroxyimidoyl chlorides (**20-21**) were synthesised accordingly to a classic protocol having the chlorination of the corresponding oximes with *N*-chlorosuccinimide in the presence of catalytic amount of pyridine as the key step of the synthesis.



Scheme 4. Synthesis of *N*-hydroxybenzimidoyl (**20**) and *N*-hydroxy-2-phenylacetimidoyl (**21**) chlorides. The overall yield is presented in the parentheses.

Nitro(phenyl)methanone oximes (also referred to as nitrolic acids) attracted our attention because they were reported in early publication of Kornblum and Weaver[807] to possess peculi-

ar reactivity towards nucleophiles. Because arylnitrolic acids were reported to be easily hydrolysable, we decided to devote our efforts for synthesising the corresponding alkyl-substituted analogue which we expected to be more stable. The probe **22** was synthesised using a slightly modified classical procedure of Komblum and co-workers,[807] which consists in an unfamiliar one-step transformation from (2-bromoethyl)benzene with 18% yield. The reaction and the hypothesised mechanism of this transformation are depicted in Fig. 96.

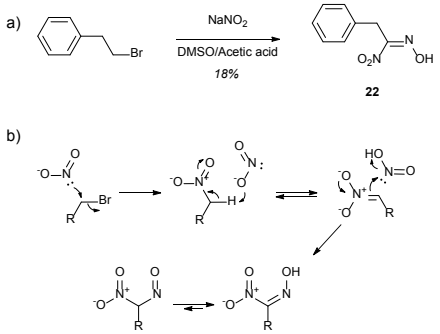


Fig. 96. **a)** Synthesis of nitrolic acid **22**. **b)** Mechanism of the reaction proposed by Kornblum and Weaver. [807]

2.2 ACTIVATED MICHAEL ACCEPTORS

Another class of reagents which was chosen for screening is a series of electrophiles which were expected to possess peculiar reactivities, especially towards specific *N*-terminal amino acid residues (Fig. 97). All of these probes contain electrophilic centres of different reactivity including stabilised enolate analogues (**23**-**25**) and activated carbonyls (**26**-**27**).

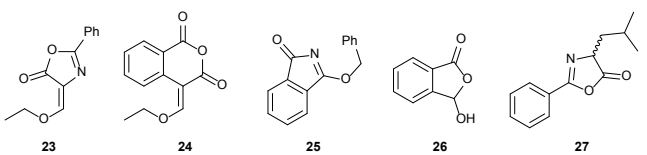
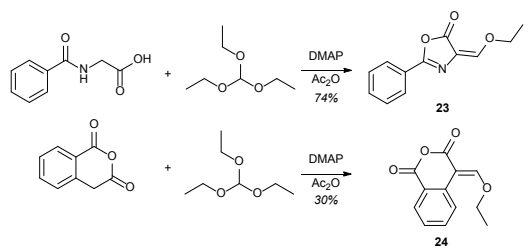


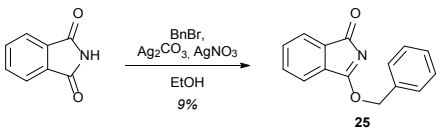
Fig. 97. Activated Michael acceptors selected for the screening.

The two structurally similar probes **23** and **24** were synthesised from hippuric acid (Scheme 5) and isochromane-1,3-dione respectively by their reaction with triethyl orthoformate in acetic anhydride in the presence of catalytic amount of 4-dimethylaminopyridine (DMAP).[808]



Scheme 5. Synthesis of 4-(ethoxymethylene)-2-phenyloxazol-5(*4H*)-one (**23**) and 4-(ethoxymethylene)isochromane-1,3-dione (**24**).

Alkyl-imidates – the class of compounds to which the probe **25** belongs – are closely related to imidoesters known for their exceptional reactivity in bioconjugation (see Section A 1.3.10, page 18). We therefore expected **25** to possess interesting reactivity profile, because of the presence of a moderately activated electrophilic centre and a very good leaving group (isoindoline-1,3-dione). The probe was synthesised using a procedure developed by Matoba *et al.*[809] for the preparation of *O*-alkylated isoindoline-1,3-diones. Preferential *O*-alkylation is achieved by using oxophilic base – silver carbonate.



Scheme 6. Synthesis of 3-(benzyloxy)-*1H*-isoindol-1-one (**25**).

2-formylbenzoic acid **26**, which under physiological conditions exists nearly exclusively in its tautomeric 3-hydroxyisobenzofuran-1(*3H*)-one form (Fig. 98), was selected for the screening because it was hypothesised to possess slightly reduced electrophilicity compared to arylaldehydes, which are known to react with the free amines present in proteins. We anticipated the formation of the cyclic form to be beneficial for the selectivity of otherwise promiscuous arylaldehydes. Moreover, the presence of the neighbouring carboxylic acid functionality would result in the theoretical possibility of modifying the reactivity of the probe in its reaction with nucleophilic amino acid side chains.

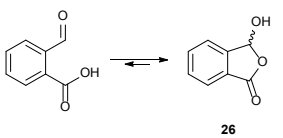
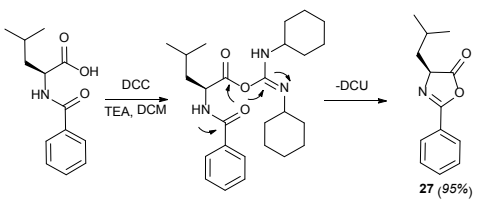


Fig. 98. Equilibrium between cyclic and open forms of 2-formylbenzoic acid (**26**).

Probe **27** represents an example of a well-documented class of compounds – oxazol-5-ones – which represent by-products often obtained during peptide synthesis.[810] These are obtained by intramolecular attack of the carbonyl group of the adjacent amide bond if no nucleophile is present in the reaction media. The reaction is accelerated when benzoic acid amide is used. We anticipated oxazol-5-one **27** to be subjected to a ring opening with nucleophilic amino acid side-chains.



Scheme 7. Synthesis of 4-isobutyl-2-phenyloxazol-5-one (**27**) from benzoyl-*L*-leucin.

2.3 VARIOUS CHEMICAL FUNCTIONS

Finally, the last group of reagents **28-32** designed for screening involved chemical functionalities which were chosen because of their reported peculiar reactivity towards specific chemical groups, although most often in organic solvents (Fig. 99).

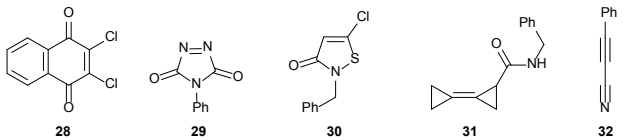
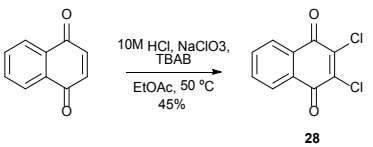


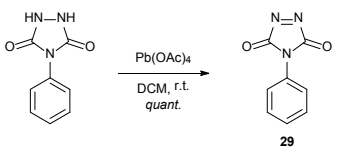
Fig. 99. Structures of chemical probes used in the screening.

The probe **28** represents an example of an activated analogue of a class of substituted 1,4-naphtaquinones. These compounds were recently reported as irreversible protease inhibitors, [811] which supposedly act by reacting with nucleophilic amino acid residues of the enzymes' active sites. The target reagent was synthesised by biphasic chlorination of non-substituted 1,4-naphtaquinone by sodium perchlorate in the presence of tetra-*N*-butylammonium bromide (TBAB) as a phase transfer catalyst using a procedure described by Cheng *et al.*[812]



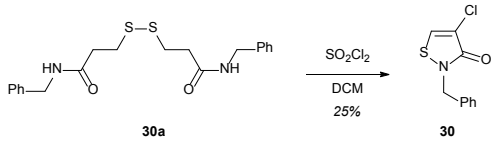
Scheme 8. Synthesis of 2,3-dichloronaphthalene-1,4-dione (**28**).

1,2,4-triazole-3,5-diones, a class of activated electrophiles to which probe **29** belongs, are known for decades, especially for their reactivity as dienophiles in the Diels-Alder reaction and with strong nucleophiles.[813] As the phenyl-substituted 1,2,4-triazole-3,5-diones were reported to possess higher stability than corresponding alkyl-substituted derivatives, **29** was selected as a probe most appropriate for conducting the screening. Moreover, its synthesis was easily achievable by one-step oxidation of the commercially available phenyl urazole. Unfortunately for us, right at the time we had finished the screening, Barbas and co-workers have reported[522] and further expanded[523] the application of the scaffold for selective tyrosine labeling.



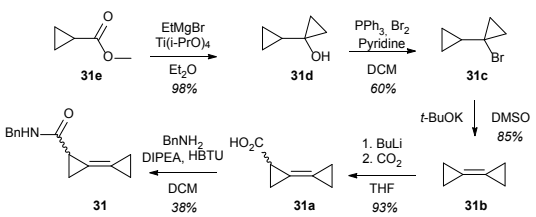
Scheme 9. Synthesis of 2,3-dichloronaphthalene-1,4-dione (**29**).

5-Chloroisothiozolones, represented in our screening by the reactive probe **30**, were reported to possess inhibitory activity towards thiol-dependent cysteine protease cathepsin B,[814] presumably by their reaction with cysteine residues in the active site of the enzyme. However, no studies of their selectivity towards cysteine or hydrolytic stability have been conducted to date. The probe is obtained during sulfuryl chloride-mediated oxidation of an easily obtainable precursor **30a** (Scheme 10).



Scheme 10. Synthesis of 2-benzyl-5-chloroisothiazol-3(2H)-one.

The interest in the preparation of **31** aroused from the effectiveness mentioned by Kozhushkov *et al.*[815]of thiol addition to the double bond of bicyclopropylidene in organic solvents. The synthesis of a key intermediate – non-substituted bicyclopropylidene – was accomplished following the protocol reported by the same research group.[816] It begins with a Kulinkovich reaction between methyl cyclopropanecarboxylate **31e** and ethyl magnesium bromide to form the second cyclopropyl moiety of the 1-cyclopropylcyclopropanol **31d** with excellent yields. The next step is the alcohol substitution with bromine by means of Appel reaction to yield **31c**. The following reaction with potassium *tert*-butoxide results in obtaining the strained bicyclopropylidene **31b** – the key intermediate of the synthesis. This could easily be transformed into the corresponding acid **31a** by quenching the carbanion generated with n-butyl lithium with carbon dioxide. Remarkably, the obtained acid was stable enough to be introduced into peptide coupling under classical conditions with no specific precautions to be met.



Scheme 11. Synthesis of 2,3-dichloronaphthalene-1,4-dione (**31**).

Lastly, conjugated alkynenitriles have been the subject of various studies devoted to the development of new synthetic approaches to new carbocyclic and heterocyclic compounds under nucleophilic addition and cycloaddition reaction conditions.[416,817-819] A recent publication of Trofimov and co-workers[416] has attracted our attention because it described the synthesis of new functionalised cysteine and methionine derivatives by reaction between these amino acids and 3-phenylpropiolonitrile (**32**). Interestingly, the reaction was found to occur at fast rate in water/acetone mixtures (50:50), which gave a hint onto possible sufficient enough stability of 3-arylpropiolonitriles.

3 SCREENING

After having synthesised the reacting probes **19-32** and optimised the screening conditions, we could turn to the screening of their reciprocal reactivity with amino acid derivatives **1-16**. The reactions were conducted directly in HPLC vials at 1 mM concentrations of the reacting partners (see Section D 2.2, for practical details). The reaction mixtures, which contained benzamide as internal control in order to mitigate the uncertainties related to variations of injection volume, were analysed by comparing the HPLC traces at T_0 of about 1 minute (first injection after the mixing) and T_{end} of 5 hours. The obtained results are depicted in Fig. 100.

For the reagents which possessed significant level of hydrolysis, the level of the probe's conversion was regarded as partially due to hydrolysis (calculated from the area of the peak of the hydrolysed product) and to the reaction with amino acid derivative (calculated from the area of the product's peak).

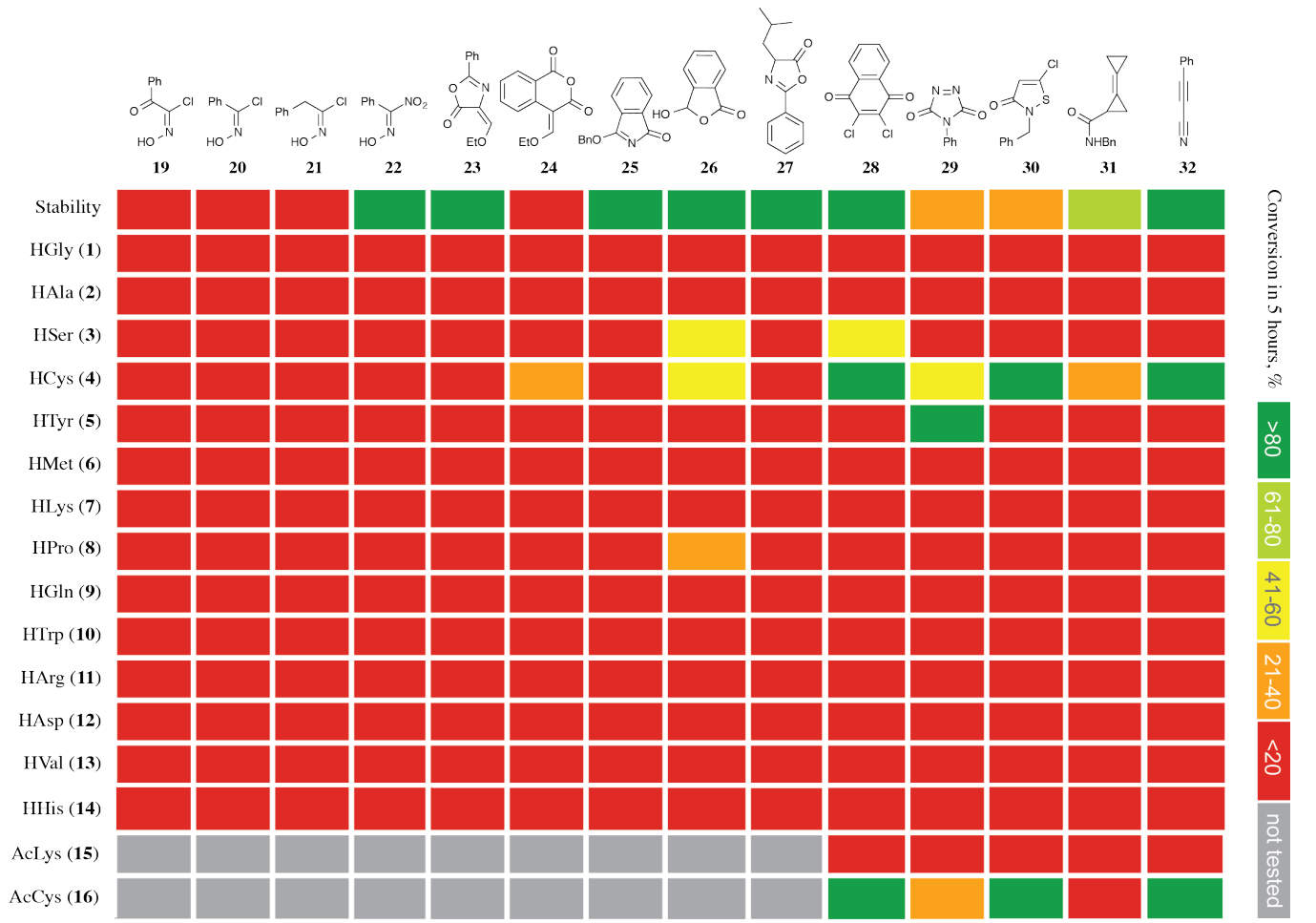


Fig. 100. Results of the screening. Reaction conditions: PBS (1x, pH 7.5), 25 °C, 1 mM concentrations of the reagent partners, 5 mM concentration of benzamide. Stability corresponds to the percentage of the non-hydrolysed probe.

The majority of the selected probes turned out to be either unstable in aqueous media (**19-21**, **24**), or too unreactive towards any of amino acid derivatives (**22-23**, **25**, **27**). Some represented several degrees of selectivity (**24**, **26**), but the reactivity was not high enough and only modest yields were observed after 5 hours of reaction at 1 mM concentrations of the reagents, which automatically makes these reactions unsuitable for bioconjugation. Finally, probes **28-32** demonstrated high selectivity towards specific nucleophilic amino acid side chains. As it was previously noted, by the time the screening was over, the dehydrourazole **29**, which was found to possess interesting selectivity profile towards tyrosine, the very same scaffold was reported by Barbas group.[522] Although being selective towards cysteine, the probe **31** represented only moderate reactivity and thus was not further investigated.

Our focus was therefore set on the three most promising cysteine-selective scaffolds: 2,3-Dichloronaphthalene-1,4-dione

(**28**), 2-benzyl-5-chloroisothiazol-3(2H)-one (**30**), and 3-phenylpropiolonitrile (**32**).

2,3-Dichloronaphthalene-1,4-dione (**28**) possessed an important drawback we became aware of while conducting first trials. Namely, an important level of cysteine oxidation was observed along with the conjugation, which would eventually result in low over-all efficiency of the conjugation.

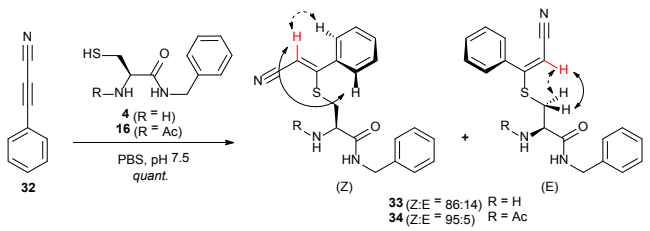
Despite its high level of conversion and exceptional reaction rates (complete conversion in 20 minutes) towards both *N*-terminal and in-chain model cysteine derivatives (**4** and **16** respectively), 2-benzyl-5-chloroisothiazol-3(2H)-one (**30**) was rather unstable in aqueous solutions and upon storage, thus requiring the preparation of its fresh solutions for each screening essay. Moreover, both the starting products and the obtained reaction products were highly reactive towards TCEP - a classical reducing agent used in bioconjugation. Because the use of thiol-containing reducing agent such as DTT or β-mercaptoethanol is not compatible with any among thiol-selective methodologies, this reactivity towards TCEP would largely limit the applicability of the scaffold **30** for the majority of the applications in bioconjugation on cysteine, for they generally imply prior reduction of disulphide bonds to uncover the reactive thiolate side chains thereof.

3-Phenylpropiolonitrile (**32**) – an example of a larger family of 3-arylpropiolonitriles (APN) – undoubtedly came forth in terms of results obtained during the first step of the screening and thus was decided to be elaborated further in more details.

4 3-ARYLPROPIOLONITIRILES (APN)

4.1 Analysis of addition products

First, the reaction of cysteine derivatives **4** and **16** with **32** was analysed by scaling it up and isolating the obtained addition products. In accordance with the results obtained by Trofimov and co-workers for non-substituted cysteine, [416] these were found to be a mixture of *(E)*- and *(Z)*-isomers of the corresponding vinylthioethers **33** and **34** (Scheme 12).



Scheme 12. Model reactions of cysteine derivatives **4** and **16** with 4-phenylpropiolonitrile **32**.

4.1.1 Stereoselectivity study

The ratio between isomers could be defined either by HPLC or by NMR. The exact constitution of two geometric isomers can be easily attributed using 2D-NOESY spectrometry. While both isomers possess a very characteristic signal in ^{1}H NMR at 5.0-5.5 ppm, which corresponds to the vinylic proton of the adduct (Scheme 12, shown in red), the correlation of this signal varies depending on the isomer. Accordingly, the *(Z)*-isomers of **33** and **34** represent a well distinguished correlation signal between the vinylic proton and two aromatic protons of the neighbouring phenyl ring, while for the *(E)*-isomers a correlation between this proton and the neighbouring CH_2 group was observed (see Annex 2 for an example of NOESY spectra of the addition products).

As one could already assume from the above-depicted results, the ratio between two geometric isomers appeared to be dependent on the structure of the thiol participating in the reaction. In order to test this assumption, a study involving differently substituted thiols was conducted. We found that, in all tested cases, the main product was the *(Z)*-isomer of the vinylthioethers. The ratio is slightly shifted towards the *(E)*-isomer (with the *(Z)*-isomer still being major) when highly encumbered thiols or those possessing positively charged functional groups situated in a close proximity to their sulfhydryl groups are used (Table 3). While the first observation could be logically attended (basing on the relative size of the hydrogen atom and the cyanogroup), the influence of the neighbouring positive charge on *(Z)*:*(E)* ratio is unclear and requires deeper study of the reaction mechanism.

Table 3. Influence of the thiol structure on *(Z)*:*(E)* ratio of the addition products obtained in the reaction with **32**[a].

Thiol	**4**	**16**	BnSH	PhSH	EtSH	*t*-BuSH
(Z):*(E)*	86:14	95:5	93:7	97:3	96:4	90:10

[a] Reaction conditions: PBS:DMSO (80:20, pH 7.5), 25 ˚C, 1 mM concentrations of both reaction partners.

Lastly, the reaction of in-chain cysteine model **16** with **32** was conducted at different temperatures, keeping all other reaction conditions unchanged. Interestingly, we have found that there was no or a very slight influence of the temperature on the stereoselectivity of the addition (Table 4).

Table 4. Influence of the temperature on *(Z)*:*(E)* ratio of the addition products obtained in the reaction of **16** with **32**[a].

Thiol	**16** (4 ˚C)	**16** (25 ˚C)	**16** (37 ˚C)
(Z):*(E)*	95:5	95:5	96:4

[a] Reaction conditions: PBS:DMSO (80:20, pH 7.5), 1 mM concentrations of both reaction partners.

4.2 Benchmarking with other methods

After having explored the reaction of **32** with various thiol-containing probes, we decided to compare the reactivity profile of **32** with other well-established methods used for cysteine modification. For this, a series of chemical scaffolds including maleimide (**35**), iodoacetamide (**36**), 4-vinylpyridine (**37**), alkynoic amide (**38**), and alkynone (**39**) derivatives was prepared (Fig. 101). Some of these probes (**35**-**37**) belong to already established methodologies for cysteine modification, [28] while other scaffolds (**38**-**39**) represent two classes of activated al-

kynes recently described by Che and collaborators[409] as promising reagents for the modification of cysteine-containing peptides.

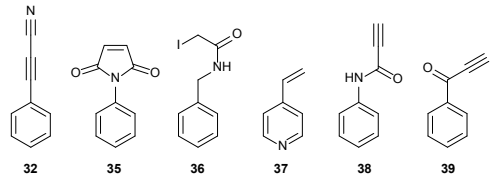


Fig. 101. Set of cysteine reactive probes investigated during the benchmarking.

In spite of being widely used for modification of cysteine in peptides and proteins, all the approaches available to date possess drawbacks limiting their general applicability, which are mainly evoked by the presence of undesired side reactions during conjugation (in case of iodoacetamides and maleimides),[28,820,821] and the relative instability of resulting conjugates in aqueous media.[28,822]

4.2.1 Hydrolytic stability and reactivity

We thus decided to proceed by conducting a side-by-side comparative study of the aqueous stability and reactivity of probe **32**, **35-39**. All tests were done in standard conditions (PBS buffer, pH 7.6, 25 ˚C). The reactivity of the probes was studied on their reaction with Cys derivatives **4** and **16** at 1 mM concentration of both reacting partners. The obtained results are summed up in Table 5.

Table 5. Hydrolytic stability and reactivity of reactive probes **32, 35-39** with a UV-traceable cysteine derivative.

	32	**35**	**36**	**37**	**38**	**39**
Degradation[a], %	0	25	0	0	0	0.9
k_{obs}, s^{-1}	-	$7x10^{-5}$	-	-	-	$2x10^{-6}$
Yield[a], %	94	99	81	60	74	99
k_2, $M^{-1}s^{-1}$	3.1	>50	1.2	0.4	0.8	>50

[a] Conversion of electrophile in 1 hour; measured by calibrated UV-HPLC traces, using benzamide as an internal standard at 1 mM concentrations of reagents in PBS:DMSO (4:1) at pH 7.6 at 25 ˚C (see Annex 3 and Annex 4 for corresponding HPLC traces).

Concerning aqueous stability of reactive probes, we found that maleimide **35** underwent 25% of hydrolysis after 1 hour in PBS at 1 mM concentration (k_{obs} = $7x10^{-5}s^{-1}$), while very little (for alkynone **39**) or no degradation (probes **32**, **36-38**) could be detected for the other tested electrophiles (see Annex 3). Interestingly, even after a month in PBS at 25 ˚C no traces of degradation could be detected for **32**. Noteworthy, maleimides' sensitivity to hydrolysis may explain in part the need to add an excess of maleimide-containing probe when performing labelling of complex biological mixtures. [823]

Concerning the reactivity, it was found that all reagents **32**, **35-39** gave from moderate (**36-38**) to very good yields (**32**, **35** and **39**). APN **32** reacted with cysteine derivative with a second-order rate constant of $k_2 = 3.1\ M^{-1}s^{-1}$, which is 2.6 times higher than that exhibited by iodoacetamide (**36**, $k_2 = 1.2\ M^{-1}s^{-1}$), about 4 times higher, compared to alkynoic amide (**38**, $k_2 = 0.8\ M^{-1}s^{-1}$) and exceeds that of 4-vinylpyridine (**37**, $k_2 = 0.4\ M^{-1}s^{-1}$) by the factor of 8.

The presence of an activated ketone group in the structure of alkynone **39** resulted in obtaining complex mixtures of the addition products when *N*-terminal cysteine derivative **4** was used, which is why we had to in-chain cysteine probe **16**, which do not possess a free amino group, to study the reactivity of alkynone **39** (see Annex 5 for HPLC traces).

Even though, the kinetics of the reaction of **32** was slower than that of maleimide **35** and alkynone **39**, where complete conversion was achieved in less than 30 minutes of reaction (k_2 was estimated to be higher than $50\ M^{-1}s^{-1}$), the chromatographic traces corresponding to APN **32** were mostly much neater and generally better reproducible. This observation gave us a hint on possible higher selectivity of APN and prompted us to further investigate it in details.

4.2.2 Selectivity

Having these first promising results in hand, we thus decided to broaden the study and search for possible side reactivities with other nucleophilic amino acids. For this purpose the reactivity of the probes **32** and **35** was tested on a full series of amino acid derivatives **1-14** used in the first step of the screening (see Section B 1.1, page 45). As previously, all reactions were analysed by HPLC after 5 hours of reaction. We found that both studied electrophiles appeared quite chemoselective for cysteine. However, in our rather stringent conditions (1 mM concentration of reagents) maleimide **35** showed a significant conversion level in the presence of such amino acids as glycine, serine and histidine derivatives (Table 6).

Table 6. Conversion of UV-traceable amino acid derivatives **1-14** in the presence of **32** and **35.**

	32	**35**		**32**	**35**
HGly (**1**)	0.2%	**3.3%**	HGln (**9**)	0.1%	1.7%
HAla (**2**)	0.1%	0.2%	HTrp (**10**)	1.3%	1.0%
HSer (**3**)	0.3%	**2.8%**	HArg (**11**)	0.4%	2.9%
HCys (**4**)	98%	100%	HAsp (**12**)	0.6%	1.3%
HTyr (**5**)	1.6%	2.4%	HVal (**13**)	0.7%	0.9%
HMet (**6**)	0.1%	1.9%	HHis (**14**)	0.6%	**8.5%**

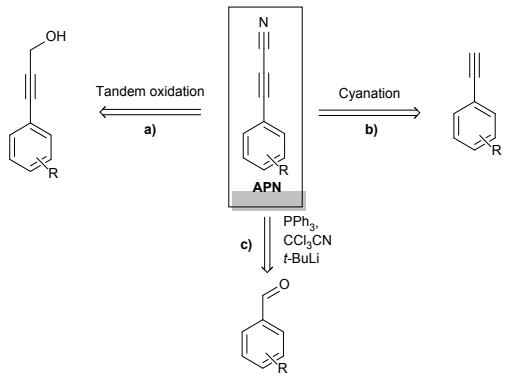
Conversion of amino acid derivatives in 1 hour; measured by calibrated UV-HPLC traces, using benzamide as an internal standard at 1 mM concentrations of reagents in PBS:DMSO (80:20) at pH 7.6 at 25 ˚C. In cases shown in bold masses of corresponding adducts were detected by ESI-LCMS.

After having compared **32** with other known methodologies and defined its highly promising potential for cysteine conjugation, we decided to test APN in the second step of the screening: on the tryptic digest of lysozyme. However, before being able to conduct the study, we had to investigate the influence of the substituent introduction on the reactivity and stability of APN. The section to follow is thus devoted to the synthesis of substi-

tuted APNs and the study of their reactivity towards Cys moiety.

4.3 Synthesis of substituted APN

The synthesis of 4-arylpropiolonitriles has been the subject of several studies. The three most elaborated approaches allowing for synthesising APN are: MnO_2-mediated free radical oxidation of the corresponding propargylic alcohols in the presence of ammonia, cyanation of the corresponding arylalkynes, and transformation of arylaldehydes to chlorovinyl nitriles followed by the elimination reaction in the presence of very strong bases (Scheme 13).

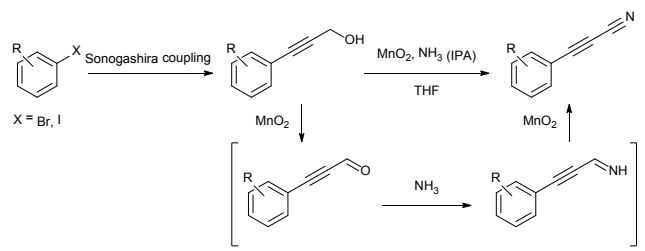


Scheme 13. Known synthetic approaches to APN. **a)** MnO_2-mediated tandem oxidation of 4-arylpropargylic alcohols. [824] **b)** Cyanation of arylalkynes. [825,826] **c)** Elimination of HCl from the α-chlorovinyl nitrile obtained *in situ via* Wittig reaction. [827]

After having conducted first trials on simple non-substituted substrates, we chose only two methodologies (Scheme 13, **a** and **b**) for further applications, for the procedure of the last approach (Scheme 13, **c**) was considerably more laborious.

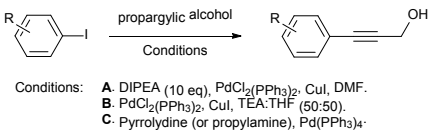
4.3.1 Tandem oxidation of 4-arylpropargylic alcohols

The approach described by McAllister and co-workers[824] has attracted our attention because of its ease and general applicability for substrates containing various substituents. The key step of the transformation consists in an *in situ* oxidation-imination-oxidation sequence of the propargylic alcohols, which can be prepared from corresponding haloarenes *via* well-elaborated Sonogashira coupling reaction.



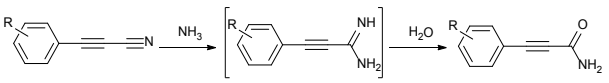
Scheme 14. Synthesis of APN from haloarene *via* Sonogashira coupling with propargyl alcohol followed by tandem oxidation with MnO_2 in the presence of ammonia solution.

Overall three different synthetic approaches were used for Sonogashira coupling depending on solubility and reactivity of the starting haloarene and were mainly introduced in response to emerging issues encountered during the synthesis (Scheme 15). During the first trials, the oxidation of arylpropargylic alcohols to corresponding APN presented several difficulties. These were found to be mainly due to the low activity of MnO_2 available in the laboratory at that time. To overcome the issue, highly active form of MnO_2 was prepared using classical approach consisting in the reaction of equimolar quantities of $MnCl_2$ and $KMnO_4$ in water with a careful control of the reaction temperature in order to obtain highly dispersed form of the oxidant. The reaction with this newly obtained oxidising agent was smooth for the vast majority of APN. However, the necessity of conducting the oxidation in the presence of ammonia (see mechanism, Scheme 14) results in several limitations of the approach.



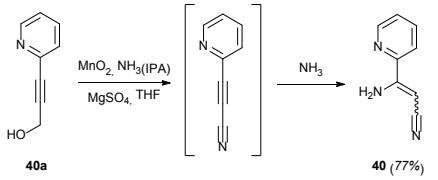
Scheme 15. Sonogashira coupling used for the synthesis of arylpropargylic alcohols (see Section D 2.9, page 77)

Firstly, the obtained 3-arylpropiolonitriles can hydrolyse in presence of ammonia to the corresponding 3-arylpropiolamides (Scheme 16), which were found to be almost completely unreactive towards cysteine. Consequently, the passing of every reaction should be carefully controlled by TLC in order to quench it at the proper time, which may vary largely depending on the temperature, the activity of MnO_2, and the substitution of the APN core. Generally, the reaction time lay in the range of 1-24 hours.



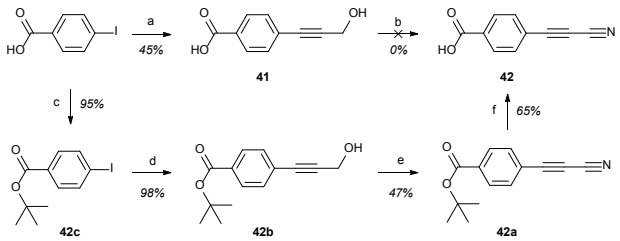
Scheme 16. Hydrolysis of APN in the presence of ammonia.

Present ammonia can also act as a base when activated APN are generated during oxidation. This was the case, for instance, during our tries to synthesise (2-pyridyl)-substituted APN derivative, where the reaction quantitatively yielded to the corresponding APN-NH_2 adduct **40** (Scheme 17).



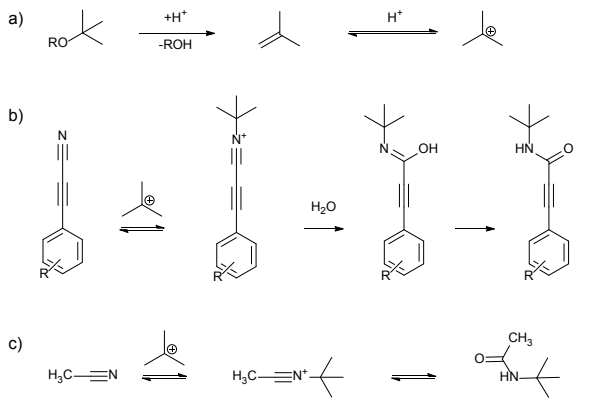
Scheme 17. Addition of ammonia to activated APN (see Annex 6 for NMR spectra of the starting material and the product).

Secondly, because the reaction is conducted under highly basic conditions, all acidic functional groups will be deprotonated. This, in its turn, may result in lowering the solubility of the starting material. For instance, because of the solubility issues, *p*-COOH substituted APN **42** could not be synthesised in a straightforward way from the corresponding propargylic alcohol, but instead a 4-step synthetic pathway including protection-deprotection steps was used (Scheme 18).



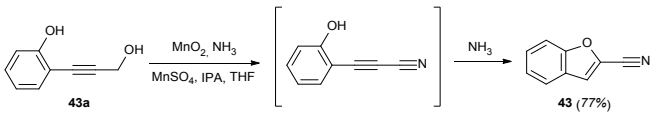
Scheme 18. Synthesis of APN-*p*-COOH using a 4-step pathway including protection and deprotection of the carboxylate. Reaction conditions: a) and d) $PdCl_2(PPh_3)_4$, CuI, propargylic alcohol, DIPEA, DMF; b) and e) THF-IPA(NH_3), $MgSO_4$, MnO_2; c) Boc_2O, Et_2O; f) TFA, ACN.

It is noteworthy that the last step of *tert*-butyl deprotection (Scheme 18f) must not be conducted in dichloromethane (under classical conditions), lest Ritter reaction takes place and the corresponding *N*-*tert*-butyl amides are obtained as by-products (Scheme 19).



Scheme 19. Ritter reaction observed during *tert*-butyl deprotonation. **(a)** Generation of the carbocation upon deprotection. **(b)** Ritter reaction on APN resulting in obtaining the corresponding *tert*-butyl amide. **(c)** Scavenging of the generated carbocation by acetonitrile.

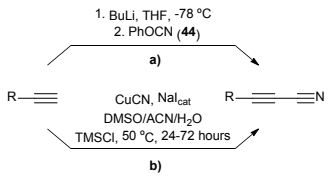
Instead, the reactions should be done in solvents (or with additives) which can trap generated *tert*-butyl cation and prevent the formation of the nitrilium ion. As such, acetonitrile, dioxane, and THF were used. While no reaction occurred in dioxane and THF (even after 24 hours of stirring at 25 ˚C), the deprotection of **42a** in acetonitrile passed smoothly and gave no by-products. Finally, because of the increase basicity of the media, some functional groups may increase their reactivity because of deprotonation. This was the issue we encountered during the synthesis of the *o*-OH substituted APN analogue, where instead of the target compound its cyclised isomer **43** was obtained.



Scheme 20. Cyclisation APN-*o*-OH under basic conditions.

4.3.2 Cyanation of alkynes

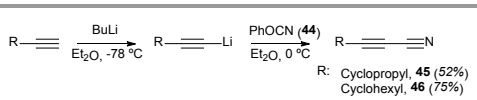
Mainly because of the limitations of McAllister's approach we were continuously developing alternative methodologies for APN synthesis. Out of many transformations described in literature, cyanation of alkynes has drawn our attention, because it allowed to broaden the list of propiolonitriles towards several scaffolds not accessible by McAllister's approach, namely alkyl-substituted propiolonitriles. We started with testing two main protocols of cyanation described in literature: by means of electrophilic cyanates and utilising oxidative cyanation with copper (I) cyanide (Scheme 21).



Scheme 21. Two main approaches used for arylalkynes cyanation: **a)** with phenylcyanate (**44**); **b)** by means of oxidative cyanation in the presence of CuCN.

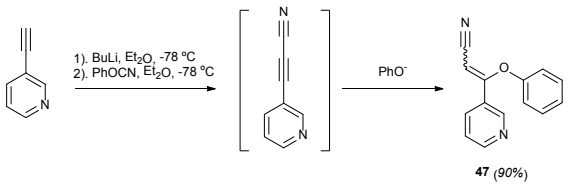
As a result of this model study, the cyanation of the carbanions generated from arylalkynes with phenyl cyanate described by Murray and Zweifel (Scheme 21a)[826] became our procedure of choice, because the oxidative cyanation with copper (I) cyanide (Scheme 21b) was less smooth, required longer reaction times, and implied the utilisation of more complex isolation procedure.

The chosen approach was especially helpful for the preparation of two 3-alkylpropiolonitriles: 3-cyclopropylpropiolonitrile (**45**) and 3-cyclohexylpropiolonitrile (**46**) (Scheme 22) which were useful during the study of APN reactivity (see Section 4.4)



Scheme 22. Direct alkynes cyanation with phenyl cyanate.

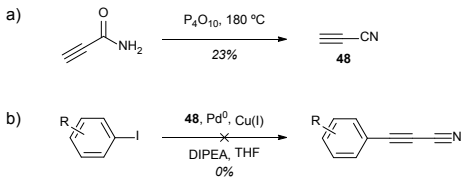
Cyanation using phenylcyanate was found to be very effective for the preparation of non-activated APN; however, all our efforts to synthesise the activated (pyridyl)-substituted APN probes failed. Even at low temperature the reaction of the less reactive 3-pyridyl-substituted isomer resulted exclusively in obtaining the addition product **47** between APN and phenolate leaving group (Scheme 23).



Scheme 23. Addition of phenolate to activated APN.

4.3.3 Direct arylation of propynenitrile

Finally, we decided to try to develop a methodology which would allow us to introduce propiolonitrile functionality in one step on late stages of synthesis. We thus tried to elaborate an approach for the preparation of APN from the corresponding arylhalides by Sonogashira coupling with propynenitrile (**48**). The synthesis of **48** is well described and consists in dehydratation of the commercially available propynamide with phosphorus pentoxide (Scheme 24a).



Scheme 24. Direct approach towards APN. **a)** Preparation of propynenitrile (**48**) from propynamide. [828] **b)** Sonogashira coupling of **48** with arylhalides. No product was obtained due to a fast degradation of **48**.

Unfortunately, propynenitrile was found to be especially unstable in basic media and in presence of Palladium complexes required for conducting the Sonogashira coupling. Complete degradation of propynenitrile resulted in yielding no target product at any of the tested reaction conditions including using of various bases (TEA, DIPEA, and propylamine), different catalysts ($Pd(PPh_3)_4$ $Pd(PPh_3)_2Cl_2$), and temperatures (0 - 60 ˚C).

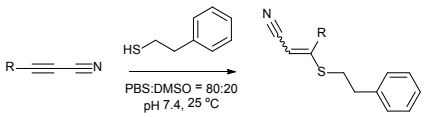
4.4 Structure-reactivity study

The developed synthetic approaches opened for us the possibility to prepare differently substituted APN derivatives, as well as corresponding alkyl analogues, which we have chosen to begin our structure-reactivity study with.

The influence of the aryl moiety on the reactivity of propiolonitriles was thus analysed by conducting a comparative study involving a non-substituted APN **32**, its cyclopropyl- (**45**), and cyclohexyl-substituted (**46**) analogues. Because alkyl-substituted propiolonitriles are not UV-visible, we decided to use phenylethylenemercaptane as a reacting partner in order to make the reaction followable by HPLC.

As it clearly follows from the obtained results (Table 7), the presence of an aryl moiety is of crucial importance for the reactivity of substituted propiolonitriles towards thiols. A simple substitution of phenyl residue by cyclopropyl results in 4.4-fold decrease of the reaction constant, while the introduction of cyclohexyl entails even more drastic 10-fold slump.

Table 7. Reaction 4-phenyl-, 4-cyclopropyl-, and 4-cyclohexyl-substituted propiolonitriles with phenylethylenemercaptane.



R:			
	32	**45**	**46**
Yield[a], %	77	43	25
k_2, $M^{-1}s^{-1}$	0.944	0.216	0.096

[a] Measured by calibrated UV-HPLC traces basing on the signal of phenylethylenemercaptane, using benzamide as an internal standard at 1 mM concentrations of reagents in PBS:DMSO (4:1) at pH 7.6 at 25 ˚C (see Annex 8 for corresponding HPLC traces).

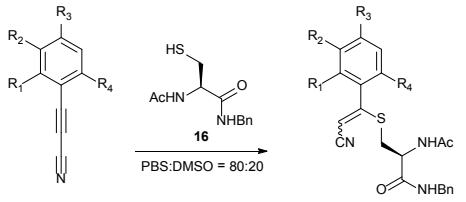
We thus decided to continue our investigation on 4-aryl-substituted propiolonitriles exclusively and synthesised a series of differently substituted APN. All of these substrates were tested for reciprocal reactivity with the model in-chain cysteine derivative **16**. In order to be able to easily compare the reactivity of differently substituted APN without need to conduct exhaustive kinetic measurements, all reactions were decided to be tested in the same reaction conditions. Following comparison of the APN conversions would in this case correlate with their relative reactivity towards thiols.

The concentrations were selected at such a level as to provide about 50% conversion of a non-substituted APN **32** after 30 minutes of the reaction (time of one HPLC run). This concentration was experimentally defined to be 50 μM in PBS at pH 7.5. The measured conversions of APN **32, 49-60** are summed up in Table 8.

Interestingly, the reaction was found to be extremely sensitive to steric hindrances induced by substituents in *ortho*-position to propiolonitrile group (compounds **52**, **57**) as well as to the electronic effect of the substituent. Concerning the effect of substituent in the *para*-position, we noted that electron-donating groups decreased the yield, while the electron-withdrawing had the opposite effect. For instance, APN-*p*-NH_2 (**55**) gave an 7% yield of cysteine adduct, while APN-*p*-NHAc (**59**) afforded 52% yield that is comparable to an nonsubstituted APN **32**, and APN-*p*-CONHMe (**60**) yielded 86%. The fact that the reactivity of APN can be easily modulated either by changing the nature of the substituent or its position on the aromatic ring, represented an important advantage compared to other methodologies.

According to obtained results, we defined the *para*-position of the APN phenyl ring as the most appropriate locus for the attachment of functional tags required for further steps of the screening, namely, the subject of the next section – screening of the APN selectivity on the tryptic digest of lysozyme. XXX

Table 8. Reaction of substituted APN with **16**.



	APN	R_1	R_2	R_3	R_4	Conversion[a]
32	-	H	H	H	H	56.1%
49	*o*-OMe	OMe	H	H	H	10.1%
50	*m*-OMe	H	OMe	H	H	46.8%
51	*p*-OMe	H	H	OMe	H	14.6%
52	*o,o'*-di-OMe	OMe	H	H	OMe	4.8%[b]
53	*o*-NH_2	NH_2	H	H	H	10.8%
54	*m*-NH_2	H	NH_2	H	H	42.0%
55	*p*-NH_2	H	H	NH_2	H	7.4%
56	*o*-Me	Me	H	H	H	10.3%
57	*o,o'*-di-Me	Me	H	H	Me	2.3%[b]
58	*o*-NO_2	NO_2	H	H	H	70.4%[c]
59	*p*-NHAc	H	H	NHAc	H	52.3%
60	*p*-CONHMe	H	H	CONHMe	H	85.5%

[a] Reaction conditions: PBS:DMSO (80:20, pH 7.5), 25 °C, 50 μM concentrations of both reaction partners. [b] Conversion in 60 minutes. [c] By-products were observed.

4.5 Selectivity of APN

In order to test the selectivity of APN in the labelling of complex peptide mixtures generated using trypsin digestion of lysozyme, a set of probes containing the APN electrophilic tag attached to TMPP, a highly responsive MS tag (see Section B 1.3, page 47) should have been synthesised. As previously discussed, once conjugated to a native peptide, TMPP modifies its chromatographic behaviour by delaying retention time in reverse phase LC and strongly enhances its signal detection. As a consequence, we expected to detect any non-specific reaction of APN with a very high sensitivity.

The synthesis of APN-TMPP probes was first envisioned starting from the commercially available TMPP precursor: TMPP-Ac-OSu (Fig. 102).

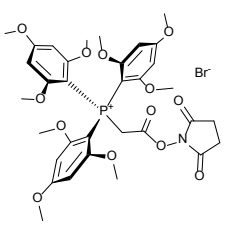
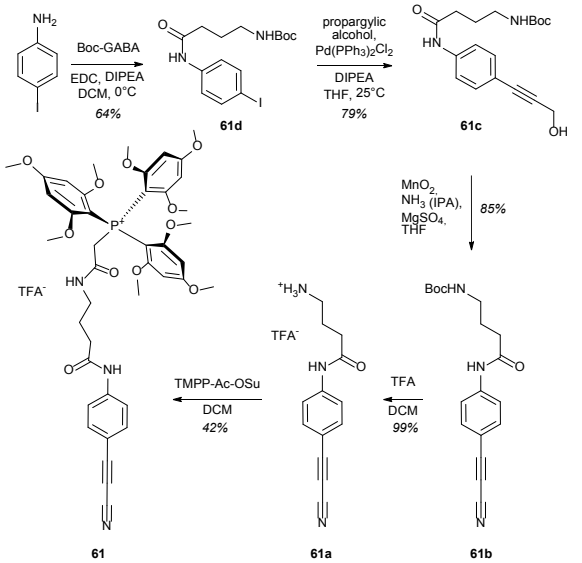


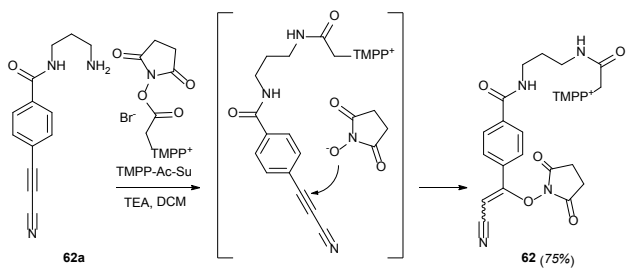
Fig. 102. Structure of the commercially available TMPP-Ac-OSu reagent.

First, a probe based on APN-*p*-NHAc scaffold was synthesised with sufficiently high overall yield. The last step being the coupling reaction of an activated ester (TMPP-Ac-OSu) with free amine (Scheme 25), however, represented several unexpected issues. Namely, long reaction times were found to be detrimental for the preparation of the target product, mainly due to reaction of the generated APN with NHS leaving group present in the media. Fortunately, the problem could simply be avoided by shortening the reaction time to 15 minutes.



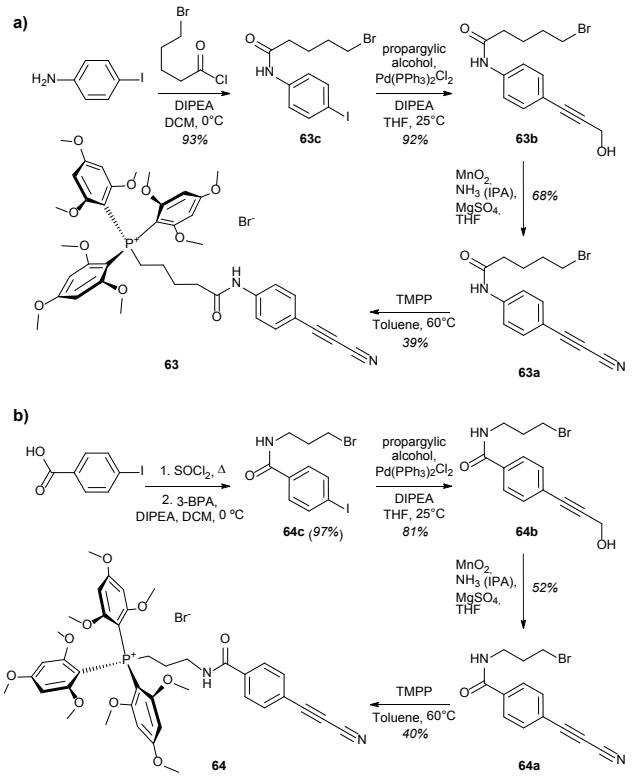
Scheme 25. Synthesis of APN-TMPP probe **61**.

In contrary, the synthesis of TMPP probe containing more reactive *p*-CONH scaffold in its structure could not be accomplished, because the corresponding APN-NHS addition product **62** was the only compound we isolated after the coupling (Scheme 26).



Scheme 26. Addition of NHS on the APN during the synthesis of APN-TMPP probe.

Because of these synthetic issues we decided to design a strategy for the synthesis of APN-TMPP probes which would not include NHS-mediated coupling. Instead, a new alternative approach, which included the direct alkylation of TMPP by the corresponding alkyl bromide, was envisioned. Following it, two APN-containing scaffolds possessing different reactivity – *p*-NHAc and *p*-CONH – could be introduced into the structures of probes APN-TMPP probes **63** and **64** without any difficulties (Scheme 27).



Scheme 27. Synthesis of APN-TMPP probe **63** and **64**.

All synthesised TMPP-containing probes were first tested for their stability in aqueous media. For this purpose, aliquots of their solutions in PBS (pH 7.4) were analysed using HPLC for 24 hours. No detectable degradation product was observed in any of the studied cases; however, the scaffold **63** gave somewhat less neat stability profile (the peak became wider over time) in HPLC and was eliminated from further tests on APN selectivity.

The labelling effectiveness of the probes **61** and **64** was tested on a model peptide mixture generated from lysozyme, which was reduced by TCEP and submitted to trypsin digestion (see Section B 1.3, page 47). The obtained peptide mixtures were analysed using NanoLC-MS/MS. The digestion was conducted with a lower than classical amount of TCEP (1 mM instead of usually used 10-100 mM concentration), for as we have found that APN underwent the reversible addition of TCEP when high concentration (100 mM) of the reductant was used (Fig. 103). While the corresponding APN-TCEP adducts could be detected by MS (see Annex 9), none of our efforts to isolate them were successful.

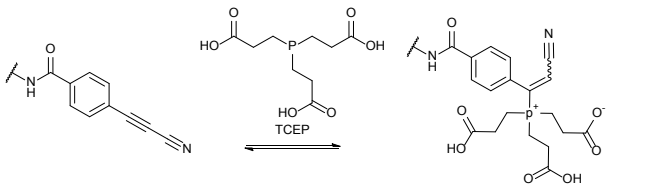


Fig. 103. Hypothesised structure of APN-TCEP adducts. See Annex 9 for the corresponding analytical data.

Evaluation of the chemoselectivity was carried out by studying the reaction of probes **61** and **64** with the model digest at a 200:1 molar ratio of APN-TMPP (1 µM) to protein (around 10:1 to cysteine moieties) at 25°C for 1 hour. Peptide mixtures obtained with and without chemical derivatisation were analysed by LC-MS/MS.

As we expected, all detectable cysteine-containing peptides reacted with both probes and were delayed while cysteine-free peptides were unaffected (Table 3). The labelling efficiency was estimated based on the ratio between intensities of labelled and non-labelled peptides by LC-MS.

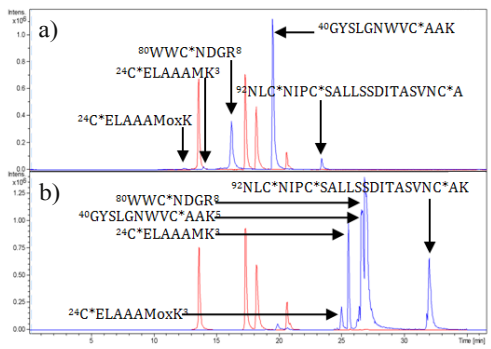


Fig. 104. Comparison of extracted ion chromatograms (EIC) from nanoLC-MS/MS of lysozyme digest. **a)** Without cysteine derivatisation. **b)** Derivatisation with APN-TMPP probe **61**.

More than 98% of the detected peptides were completely tagged. Labelled cysteine-containing peptides had an increased retention time due to the addition of the hydrophobic TMPP group, whereas the retention time of all other peptides remained unchanged (Fig. 104). Due to the introduction of a permanent positive charge given by the tag, ionisation efficiency of modified peptides in ESI-MS was increased. Enhancement of the peptide signal ranged from 2-fold to roughly 8-fold for the both tested APN-TMPP probes **61** and **64**. The chromatograms of peptide mixtures before and after the labelling were carefully analysed (Fig. 104) and the obtained results for the probe **61** are shown in Table 9. We estimated the labeling efficiency based on the ratio between intensities of labelled and nonlabelled peptides by LC-MS. More than 98% of the detected peptides were found completely tagged.

Table 9. LC-MS analyses of tryptic digest of lysozyme before and after reaction with **61**.

	Before tagging		After tagging			
Peptide sequence[a]	m/z (charge state)	RT (min)	m/z (charge state)	RT (min)	ΔRT (min)	Number of tags
^{24}**C**ELAAAMK^{31}	418.70 (+2)	13.9	545.89 (+3)	24.2	10.3	1
^{24}**C**ELAAAM$_{ox}$K^{31}	426.70 (+2)	12.6	551.23 (+3)	23.7	11.1	1
^{40}GYSLGNWV**C**AAK^{51}	634.81 (+2)	19.8	689.96 (+3)	25.3	5.5	1
^{80}WW**C**NDGR^{86}	468.69 (+2)	16.5	579.22 (+3)	25.7	9.2	1
^{92}NL**C**NIP**C**SALLSSDITASVN**C**AK^{114}	779.71 (+2)	23.7	949.20 (+5)	30.3	6.6	3
^{52}FESNFNTQATNR^{63}	714.83 (+2)	13.5	714.83 (+2)	13.5	0	0
^{135}GTDVQAWIR^{143}	523.27 (+2)	17.3	523.27 (+2)	17.3	0	0
^{64}NTDGSTDYGILQINSR^{79}	585.28 (+2)	18.2	585.28 (+2)	18.2	0	0
^{116}IVSDGNGMNAWVAWR^{130}	559.27 (+2)	20.9	559.27 (+2)	20.9	0	0

[a] Cysteine residues are in bold red.

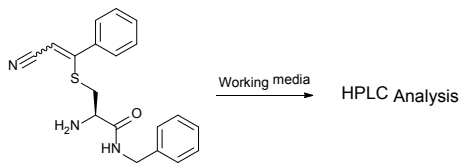
4.6 Stability of APN-thiol conjugates

As maleimide-mediated conjugation on cysteine was to date the most often used methodology in bioconjugation (see Section A 1.3.3, page 13),[28] we decided to conduct a comparative side-by-side study of two types of conjugates: containing thiol-maleimide linkage and their APN analogues.

Because of the gain in popularity of the preparation of the biological agents (biologics),[829] some previously neglected features of the conjugation methodologies were found to be responsible for the majority of observable detrimental effects on new prominent therapeutic agents.[830] These are, namely, the biostability of the linkage used for the preparation of the conjugate, the distribution of the products obtained during bioconjugation, and their exact constitution.

The study of APN-thiol conjugates' stability was first conducted on the model conjugate **33**. Unexpectedly, this was found to be remarkably stable in a wide range of pH (from 0 to 14), in the presence of an excess of such nucleophiles as thiophenol, cysteine, thioethanol, glutathione (100-1000 equivalents, 10-100 mM), as well as reducing agents such as TCEP or DTT (0.1-1 M) at pH 7.4 (see Annex 10 for HPLC traces). Thiol exchange, the process hypothesised to be the main reason responsible for the *in vivo* instability of maleimide-thiol conjugates, was then studied by incubation **33** with 100 equivalents of thiophenol at slightly basic conditions (TRIS buffer, pH 9.5). HPLC monitoring showed no significant exchange even after 2 hours of incubation. In comparison, the two other activated alkynes **38** and **39** described by Che and collaborators,[409] underwent complete cleavage after only 30 minutes under the same conditions.

Table 10. Tests on **33** stability in various media.[a]



33

#	Working media	Degradation[b]
1	100 mM PhSH in PBS:DMSO (80:20), pH 7.4	0.7%
2	1M H_2O_2 (pH 7.0)	1.8%
3	1M HCl (pH 0)	0.1%
4	1M NaOH (pH 14)	3.4%
5	1M Reduced glutathione (GSH) in PBS, pH 7.4	0.4%
6[c]	1M Imidazole in PBS, pH 7.4	0.5%
7	50 mM PhSH in TRIS:DMSO (80:20), pH 9.5	0.3%
8	1M cysteine in PBS, pH 7.4	0.3%
9	100 mM BME[d] in TRIS:DMSO (80:20), pH 9.5	0.0%
10	100 mM TCEP in PBS, pH 7.4	0.0%
11	1M DTT in PBS, pH 7.4	0.5%

[a] See Annex 10 for HPLC traces. [b] Conversion of **33** in 1 hour; measured by calibrated UV-HPLC traces, using benzamide as an internal standard in working media at 25 °C. [c] Reported conditions for the cleavage of maleimide-thiol conjugates.[312] [d] 2-Mercaptoethanol.

In order to evaluate the biostability of the APN-thiol linkage, a side-by-side stability comparison of APN- and maleimide-conjugates was conducted. To monitor bond breakage at low concentrations in biological environments we synthesised two FRET-based probes **65** and **66**, containing APN- and maleimide-thioether scaffold respectively (Fig. 105a). In these probes, the cleavage of any bond between fluorescent dye (5'-TAMRA) and quencher (BHQ-2) will result in the appearance of the fluorescence signal. Since these two probes differ only by their core APN- or maleimide-thiol linkage, we assume that appearance of fluorescence will account for instability of the

scaffold. Moreover, because of the very similar structure of the two probes, we expected their cellular uptake to be comparable.

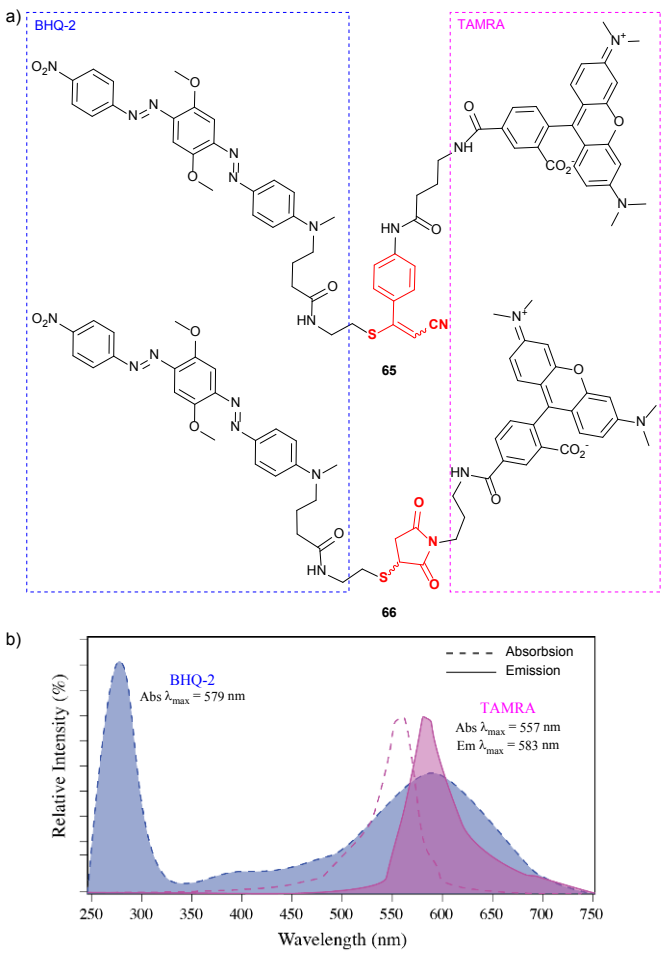


Fig. 105. **a)** Structures of FRET-probes **65** and **66** used in the study. APN- and maleimide-thiol linkages are shown in red. **b)** Absorption (dashed) and emission (solid) spectra of BHQ-2 (blue) and TAMRA (magenta).

To test the stability in human plasma, both probes (1 μM) were incubated at 37°C, while monitoring the appearance of fluorescence at 580 nm over 15 hours (Fig. 106). We detected that incubation of **65** showed only 1.5 times increase of fluorescence, while **66** gave a much stronger increase (4-fold) due to degradation in biological medium. This was in a good accordance with previously described plasma-instability of maleimide conjugates.[830]

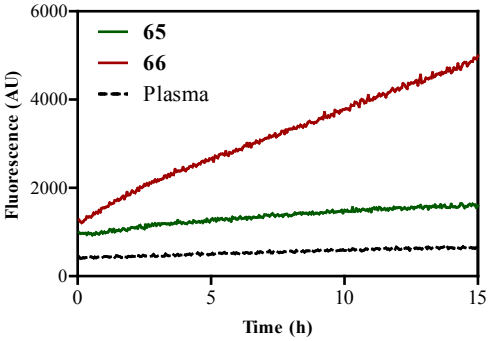


Fig. 106. Side-by-side comparison of the instability of probes **65** and **66** (1 μM) in human plasma at 37 °C. Higher fluorescence level corresponds to higher level of degradation.

The biolabilities of the probes were then studied in living cells (BNL CL.2). To provide quantitative and spatial information about probe hydrolysis, fluorescence activated-cells were imaged by flow cytometry and confocal microscopy. First, cells were loaded with the FRET-based probes (1 μM, 2 hours) and the fluorescence activation was recorded over time by flow cytometry (Fig. 106). As in human plasma, the obtained results indicated higher stability of APN probe **65** *vs*. maleimide probe **66** over the given time. For instance, after one day of incubation, maleimide probe exhibited a 4-fold higher level of fluorescence than the APN probe.

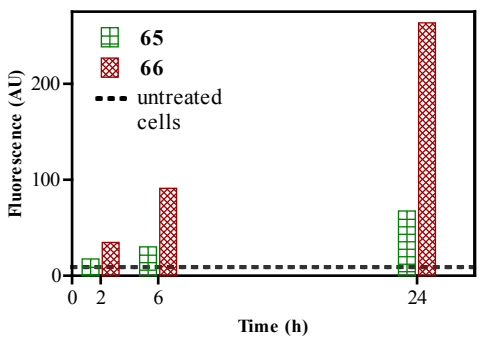


Fig. 107. Side-by-side comparison of the instability of probes **65** and **66** (1 μM) in in BNL CL.2 cells. Blue dashed line corresponds to the fluorescence level on untreated cells. Higher fluorescence level corresponds to higher level of degradation.

For confocal imaging, cells were loaded with both probes (5 μM, 90 minutes) and nuclei were stained with Hoechst 33258 (5 μg/ml, 30 minutes). Resulting images are shown in Fig. 108 (probe **65** – image A; probe **66** - image B). Compared to **65**, maleimide-thiol conjugate **66** exhibited higher fluorescent signal with a vesicular distribution, probably corresponding to endosomal uptake of the probe. The results are congruent with flow cytometry measurements and confirmed higher stability of APN- than maleimide-thiol conjugates.

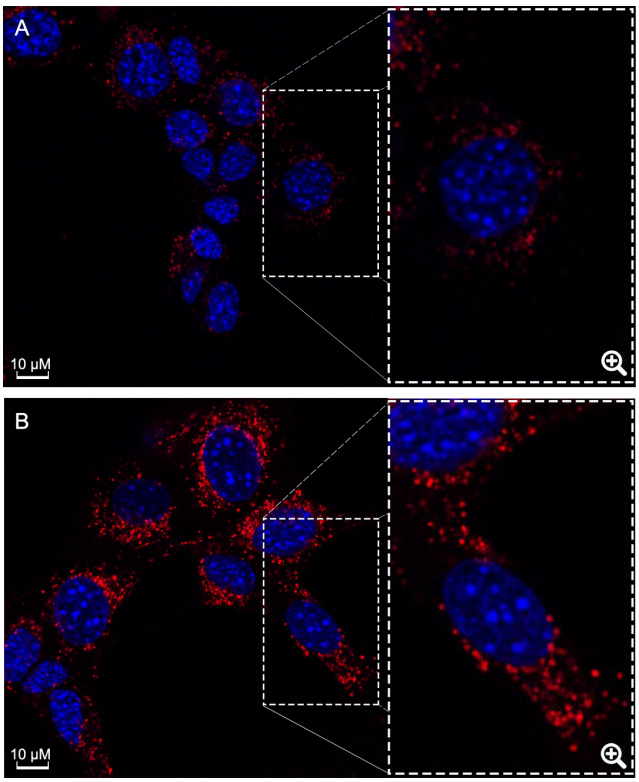


Fig. 108. *In vitro* imaging of BNL CL.2 cells with probe **65** (A) and probe **66** (B). Viable cell imaging was carried out by first staining with the different probes (5 μM, 90 min) followed by staining with Hoechst 33258 (5 μg/ml, 30 min). Confocal fluorescence microscopy imaging of *via*ble cells was performed at a magnification of x630.

In regard with above-described studies, we have shown that APN scaffold can serve as a powerful chemical tool for coupling purposes because of its good reactivity and excellent selectivity with thiol moieties, and remarkable stability in physiological media. Therefore, we were fully convinced that APN-mediated conjugation may open an interesting prospect for the preparation of new biologics.

4.7 Functionalised APN

In order to further expend this new class of coupling reagents, we have started to explore the possibility of synthesising relevant functionalised APN derivatives representing the major classes of probes and reagents used in bioconjugation.

4.7.1 Thiol-thiol reactive linkers

Three isomeric homobifunctional APN linkers **67-69** were synthesised from the corresponding aryl-propargylic alcohols using MnO_2-mediated oxidation in the presence of ammonia (see Section B 4.3.1, page 54). This last oxidation step posed several difficulties mainly due to a big amount of partially hydrolysed products containing propiolonitrile and propiolamide reactive functions; as a result rather low yields were obtained (Fig. 109). The distance between reacting moieties varies depending on the isomer, therefore allowing the use of the probes as thiol-thiol "tweezers" for either connecting two reacting partners or rebridging of reduced disulfide bonds present in the molecule.

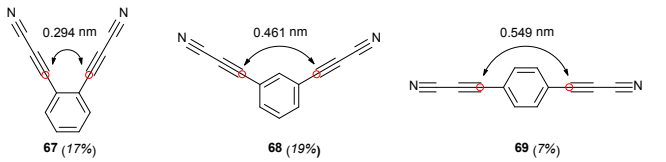
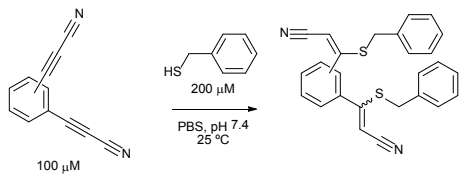


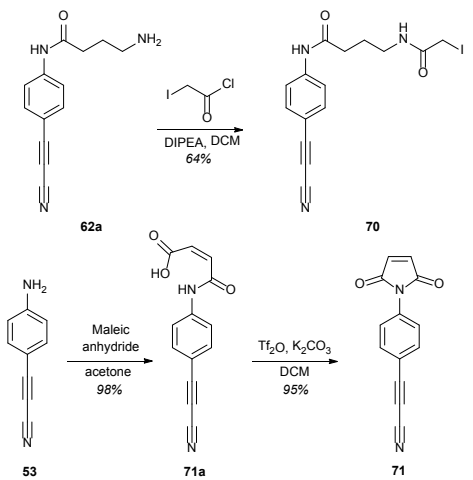
Fig. 109. Structure of homobifunctional APN. The distance between reacting centres is indicated.

The reactivity of probes **67-69** was first tested on the reaction with benzylmercaptane. All reactions were smooth and total conversion of probes was achieved in 30 minutes at 100 μM and 250 μM concentration of the probes and the thiol respectively. However, in the case of the *ortho*-substituted probe **67**, various products were obtained, supposedly due to polymerisation and cyclisation. While probes **68** and **69** were completely stable in aqueous media, the probe **67** turned black even upon storage in dry DMSO at -20 ˚C. Because the two reacting functionalities possess similar reactivity towards sulfhydryls, homobifunctional APN reagents should be regarded as effective dimerisation and rebridging thiol selective reagents.



Scheme 28. Model study of APN-tweezers reactivity.

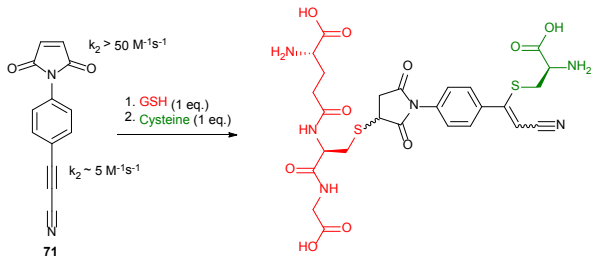
Another type of probes we have developed allow for successive linking of two different thiol molecules. These APN-containing chemical linkers possess chemical functions which reactivity towards thiols differs largely enough, which makes the kinetic resolution of two addition steps possible.



Scheme 29. Synthesis of heterobifunctional thiol reactive APN probes.

As such, probes combining APN and iodoacetamide, or APN and maleimide moieties were synthesised (Scheme 29). Both probes contained the same core APN-*p*-NH_2 scaffold in their structures, and were obtained *via* simple synthetic pathways by acetylation (probe **70**), or maleination followed by cyclisation in the presence of triflate anhydride (probe **71**).

Looking more interesting for bioconjugation APN-maleimide probe **71** was decided to be studied in more details. First, its stability to hydrolysis was found sufficiently high to conduct a model study on its applications for the coupling of two thiol-containing fragments - glutathione and free cysteine. The efficiency of the linker was found substantially high when the conjugation was carried in PBS at pH 7.4 (Scheme 29, see Annex 11 for HPLC traces).



Scheme 30. Kinetically resolved thiol-thiol coupling using the probe **71**.

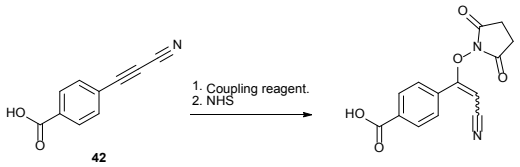
4.7.2 Thiol-amine reactive linkers

A class of reagents allowing for linking thiols and amines is to date one of the most often used family of linkers for bioconjugation. Indeed, the majority of approaches for the preparation of conjugates consist first in the introduction of thiol-reactive moieties onto the surface of a biomolecule by means of the modification of its free amino groups. Following reaction of the newly introduced thiol-reactive group allow introducing of virtually any functional entity.

The previously synthesised APN-*p*-COOH probe **42** containing free acid group could, to some extent, already be regarded as such an amine-thiol conjugative linker, but its application would require the utilisation of coupling reagents. This is often not acceptable in bioconjugation, which is why our main efforts were dedicated to the preparation of APN-containing probes possessing activated ester moieties in their structure in order to be utilisable as ready-to-use linkers for bioconjugation.

We first started with the straightforward preparation of the corresponding NHS-ester of the acid **42**, being aware of the possible reaction between the APN moiety in its structure and *N*-hydroxysuccinimide. Indeed, none of reaction conditions were successful for the preparation of the target compound, but exclusively resulted in the isolation of the corresponding APN-NHS addition product (Scheme 30).

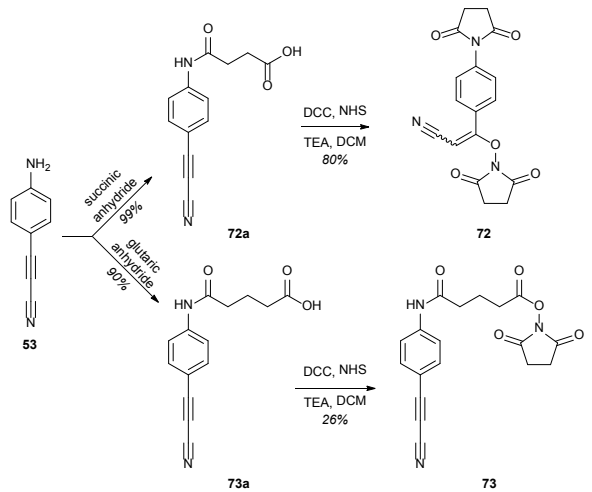
The reason for such an elevated kinetics of NHS addition to **42** and impossibility of isolation of the targeted NHS-activated ester of APN was clear. Possessing a good electron-withdrawing carboxylate group in *para*-position, the APN-*p*-COOH scaffold clearly belonged to a class of activated APN.



Scheme 31. Addition of NHS to the APN moiety during the synthesis of APN-NHS probe (observed by LCMS).

The first alternative strategy we started to explore implied thus the preparation of a probe containing less reactive core scaffold. Exploiting the results obtained during the synthesis of APN-TMPP probe **61**, we assumed that the presence of *p*-NHAc fragment may be helpful for the preparation of the target NHS-esters.

The synthesis of the corresponding acids were started from APN-*p*-NH_2 key-intermediate **53** and was similar to the already elaborated for the preparation of the APN-maleimide reactive probe **71** reaction with cyclic anhydrides. As a result, two APN-containing acids **72a** and **73a** could be obtained smoothly with very high yields. While no difficulties were encountered during the last step of the preparation of the probe **73**, the main product obtained from the acid **72a** was not the target NHS-ester, but a product of a two-step transformation involving cyclisation to APN-succinimide derivative followed by the addition of NHS to its propiolonitrile moiety (**72**).



Scheme 32. Alternative strategy used for the synthesis of APN-NHS probes.

In parallel, realising a very important disadvantage of the presence of *N*-hydroxysuccinimide as leaving group in APN-NHS probes, we started to develop alternative activated ester scaffolds that would be deprived from any possible side-reactions. We have conducted a comparative study on APN reactivity towards leaving groups most often used for the preparation of activated esters: *N*-hydroxysuccinimide, *N*-hydroxybenzotriazole, *N*-hydroxytriazolopyridine, perfluorophenol, and perfluoro-4-hydroxybenzenesulfonic acid (Fig. 110).

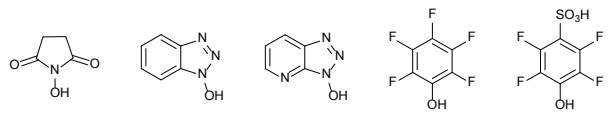


Fig. 110. Leaving groups tested for their reactivity towards APN (from left to right): *N*-hydroxysuccinimide, *N*-hydroxybenzotriazole, *N*-hydroxytriazolopyridine, perfluorophenol, perfluoro-4-hydroxybenzenesulfonic acid. Reaction conditions: 10 mM concentration of the reaction partners (APN **32** and the leaving group), controlled by HPLC after 1 hour of incubation.

As a result of the screening, we could identify perfluorated phenols to be absolutely appropriate for our purposes, because no addition product between them and a non-substituted APN **32** could be detected neither in aqueous (PBS, pH 7.4), nor in organic solvents (DCM, THF).

Two probes containing perfluorophenol leaving groups (**74-75**) were synthesised using a DCC-mediated coupling in DCM with moderate yields and no corresponding addition products between APN moiety and leaving groups.

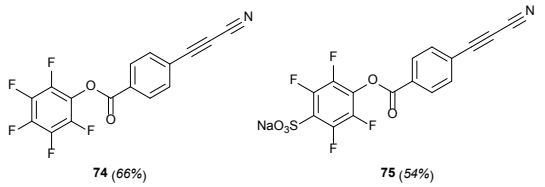
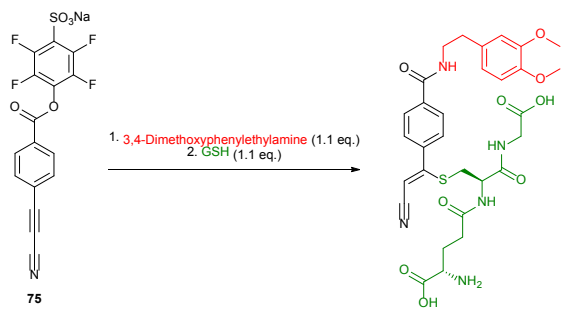


Fig. 111. Structures of APN-containing activated perfluorophenolates probes.

Both probes were then tested for their efficiency for the conjugation of a model amine to glutathione (GSH, Scheme 33). This was found to be remarkably high (no traces of by-products in HPLC), especially when conducted in PBS (pH 7.4) and borate buffer (pH 8.6). To further test the applicability of the linker **75** for conjugation, we conducted the same reaction, but using an excess of the amine (5 eq.) on the first step. After 5 minutes of the reaction, GSH (1.1 eq.) was thus added to the reaction mixture containing the excess of the amine. The efficiency of the conjugation remained unaffected.

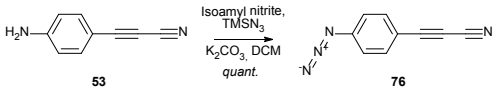


Scheme 33. Conjugation of 3,4-dimethoxyphenylethylamine with glutathione (GSH) using probe **75**.

4.7.3 Bioorthogonal APN linkers

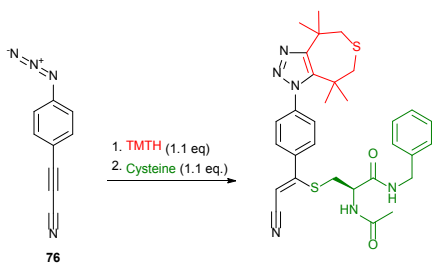
Bioorthogonal linkers are currently gaining popularity in many domains of modern chemical biology. This fact has prompted us to expand our series of APN-containing linkers towards this class of reagents.

First, the synthesis of APN-azide linker **76** was carried starting from a usual key-intermediate – APN-*p*-NH_2 (**53**). The transformation was unexpectedly smooth and did not require any optimisations of reaction conditions, and gave quantitative yield of the target compound.



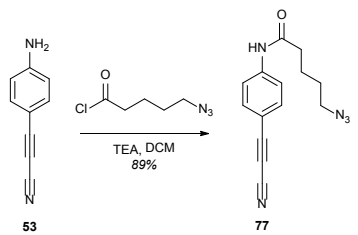
Scheme 34. Synthesis of APN-azide probe **76**.

Because of known instability, or even explosibility, of non-stabilised arylazides, we first conducted a stability study of the probe **76** in aqueous media as well as upon heating. Interestingly, this was found to be absolutely stable in both cases. While heating to 100 °C undoubtedly provoked some degradation (the compound turned black and clammy upon melting), no traces of any degradation products could be detected by HPLC even after 48 hours of incubation in PBS at pH ranging from 5.5 to 10.5 at 25 °C. The model study of the applicability of **76** as a conjugation linker was carried out by conducting the linking of a strain-promoted alkyne (TMTH) with free thiolate group of a cysteine model in-chain model **16** (Scheme 35, see Annex 12 for HPLC traces).



Scheme 35. Conjugation of phenylethylamine with cysteine using APN-azide linker **76**.

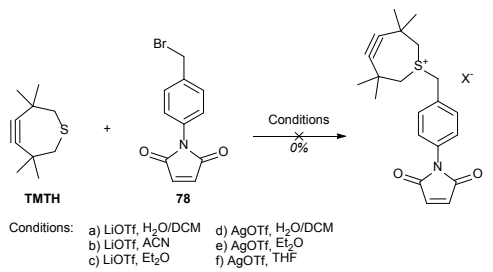
A more classical alkylazide-containing APN probe was also synthesised starting from the same key-intermediate in one step by acylation with chloroanhydride of azidopentanoic acid in high yield and was used in copper-catalysed click reaction with terminal alkyne fragment for the preparation of APN-HAZA-biotin probe **82** (see Section B 4.7.4, page 65).



Scheme 36. Preparation of APN-alkylazide probe **77**.

Because the development of new linkers for strain-promoted alkyne-azide click reaction was one of the central subjects in our research group, we decided to focus on the development of a TMTH probe containing APN functionality in order to be functionalisable by a large variety of already developed thiol-containing probes.

The results obtained in the laboratory suggested that the only feasible way for functionalisation of TMTH was its alkylation by benzylhalides in the presence of lithium or argentum triflate to generate the corresponding TMTI derivative.[831] The reaction was rather slow and thus required reaction times of 24-96 hours of stirring at 25 ˚C in two-phase water-DCM system. Before the APN stability was discovered, the major efforts in the laboratory were consecrated to the synthesis maleimide-based TMTH probe which was supposed to become a key intermediate for the synthesise a wide variety of TMTI-containing functionalised reagents. These efforts however were unsuccessful because of complete hydrolysis of maleimide at required reaction conditions already after less than 30 minutes of the reaction (Scheme 37).

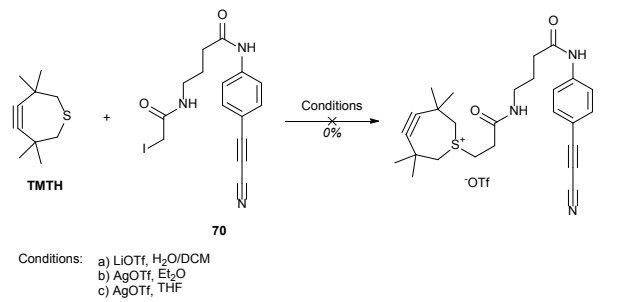


Scheme 37. Tested synthetic approaches towards maleimide-TMTI probe. None of the tested conditions gave the target compound.

Another way involving more stable iodoacetamide fragment was not feasible because iodoacetamide itself was reactive (even though much less than benzylbromide) towards TMTH which resulted in obtaining non-separable mixtures.

By the time APN was engaged into the reaction pathway, many reaction conditions were tested, and none of them was found working for the preparation of the targeted TMTI derivative. Because our studies on the stability of APN in various reactive media have demonstrated an unexpectedly elevated stability of the propiolonitrile moiety, we started considering synthetic approaches towards APN-TMTI probe.

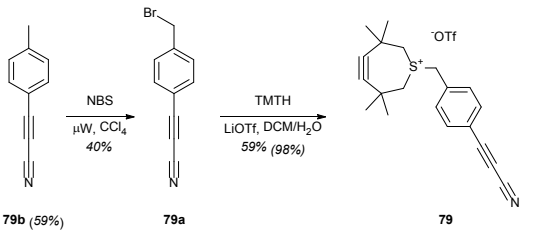
First trials were started on a previously synthesised probe **70** containing iodoacetamide and APN moieties in its structure. Unfortunately, **70** was found to be completely unreactive with TMTH and only the starting materials could be isolated even after several days of reaction.



Scheme 38. First unsuccessful trials on the preparation of APN-TMTI probe starting from iodoacetamide **70**. Only starting materials could be isolated.

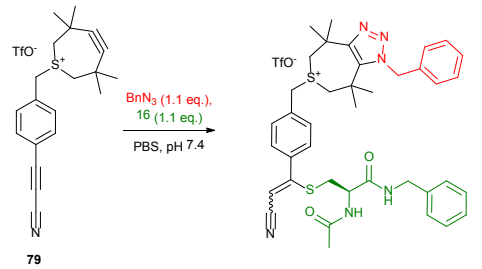
Even though we were unable to get the targeted compound, obtained results were extremely valuable, for they have confirmed the stability of APN in the rather stringent conditions required for the reaction to occur.

Another APN building block containing benzylbromide moiety in its structure was envisaged for the synthesis. For this purpose, APN-*p*-Me probe **79b** was first monobrominated, which turned out to be rather non-trivial task. Only radical NBS-mediated brominating under μW-irradiation (heating was found completely ineffective and resulted in inseparable complex mixtures of starting materials, monobrominated, dibrominated, and many other by-products). Pure monobrominated APN **79a** could, however, be obtained in modest yields when short-time intensive μW irradiation was used instead of plain heating. The obtained APN-benzyl bromide **79a** was then engaged into the reaction with TMTH under above-described conditions. The products could be obtained with high yields after 120 hours of reaction (the starting material could completely be recovered giving the virtual quantitative recovered yield).



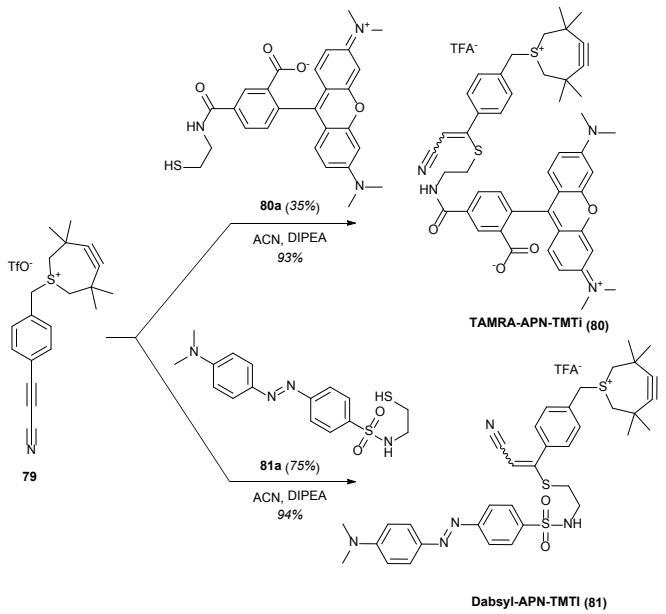
Scheme 39. Synthesis of APN-TMTI probe **79**.

Linker **79** represents an extremely valuable intermediate for the synthesis of various TMTI-containing probes not accessible otherwise. As before, we first started with the tests on stability of **79** and its reactivity towards thiol-containing reagents. It was found to be completely stable in aqueous buffers at neutral pH and reacted neatly with thiol-containing compounds to give targeted addition products. The applicability of the probe **79** for conjugation was then tested on its reaction with benzyl azide and in-chain cysteine model **16**. The conjugate was obtained quantitatively in less than 30 minutes of reaction at 100 μM concentrations of the reacting partners (controlled by LCMS).



Scheme 40. Conjugation of benzylazide (BnN_3) with cysteine derivative **16** using APN-TMTI linker **79**.

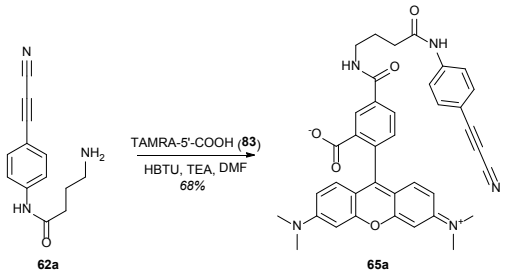
Two TMTI-containing probes, one containing a fluorescent dye (**80**), and another possessing a quencher (**81**), were synthesised using APN as an anchor point for the reaction with corresponding thiol-containing intermediates (Scheme 41). In both cases, the reactions were found to give the targeted products quantitatively after 5 minutes of reaction if conducted in acetonitrile containing an excess of triethylamine or DIPEA. The synthesised probes are currently further elaborated in the laboratory.



Scheme 41. Synthesis of TAMRA- and Dabsyl-containing TMTI probes using APN-TMTI linker **79**.

4.7.4 Various APN probes

Several APN compounds were synthesised for other specific applications. First, APN-TAMRA fluorescent probe **65a** was obtained for the preparation of FRET-based probe **65** used to study the stability of APN-thiol linkage (see Section 4.6, page 59). The probe was also used for the labeling of BSA (possesses one exposed free Cys residue on its surface) as a proof-of-concept of APN applicability for the labelling of full-size proteins. The obtained results showed high efficiency of APN-mediated labeling at low concentration of the probe (1 μM).



Scheme 42. Synthetic pathway for the preparation of APN-TAMRA **65a**.

Another compound that was synthesised, in this case for proteomics applications, is the probe **82** containing an APN moiety connected to a biotin residue by means of a cleavable HAZA linker, previously developed in our laboratory.[832] The probe was obtained by assembling two fragments by means of copper-catalysed azide-alkyne click reaction: APN-alkylazide **77** and alkyne-HAZA-biotin probe **82a**. The application of this probe is envisaged in the near close future in the laboratory for substracting cysteine-containing subproteome *via* streptavidin-biotin-based extraction approach involving first immobilisation of all cysteine-containing peptides on streptavidin-coated magnetic beads followed by specific cleavage of HAZA-linker with dithionite.

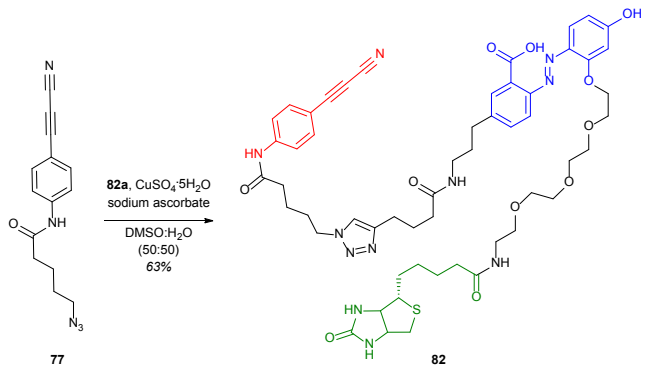


Fig. 112. Structure of the probe **82** consisting of an APN fragment (shown in red), a HAZA cleavable linker (shown in blue), and a biotin moiety (shown in green).

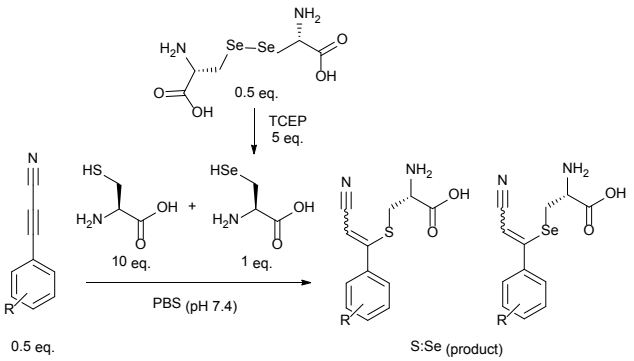
4.8 Cysteine vs selenocysteine labeling

One of the directions we started to elaborate after having obtained substituted APN derivatives and defined their reactivity profiles was the development and optimisation of an approach allowing selective labelling of selenocysteine in the present cysteine. These two amino acids possess similar, but differing reactivity, namely evoked by higher acidity and lower oxidation potential of selenols compared to thiols (see Section A 2.6, page 41). Moreover, because it is well-established that selenols are better nucleophiles than thiols,[833,834] we suggested that fine tuning of APN reactivity might be helpful for achieving some level of selectivity towards selenocysteine.

Such selectivity would represent a huge advantage for numerous applications, namely in the field of proteomics, where sele-

nocysteine is often difficult to detect because of its low abundance and propensity to oxidative degradation.

First, the reaction of non-substituted APN **32** with a mixture of cysteine and selenocysteine (generated *in situ* from selenocystine) was studied. Interestingly enough, already under non-optimised conditions (PBS, pH 7.4) the reaction mainly occurred on selenocysteine (ratio between addition products of cysteine *versus* selenocysteine was 1:99, Scheme 43).



Scheme 43. Scheme of the comparative study for S *vs* Se selectivity.

The influence of the APN structure on its selectivity towards selenocysteine was then studied. The ratio between the initial concentrations of cysteine and selenocysteine was increased (to 100:1) in order to obtain S:Se ratio of the addition products in an easily measurable range.

Table 11. S:Se ratio in the addition products of the reaction between a substituted APN and cysteine:selenocysteine (100:1) mixture.

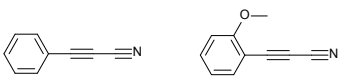
#	APN		S:Se (product)
1	**32**	-	10:90
2	**49**	*o*-MeO	8:92
3	**50**	*m*-MeO	11:89
4	**51**	*p*-MeO	12:88
5	**53**	*o*-NH_2	10:90
6	**55**	*p*-NH_2	12:88
7	**57**	*o,o'*-Me	9:91
8	**58**	*o*-NO_2	23:77

[a] Reaction conditions: PBS (pH 7.4), 25 ˚C, concentrations of reagents: APN (0.5 mM, 0.5 eq.), selenocysteine (1 mM, 1 eq.), cysteine (100 mM, 100 eq.), TCEP (5 mM, 5 eq.).

Apparently, the decreasing of APN reactivity by introduction an electron-donating group only slightly influenced its selectivity towards selenocysteine *vs.* cysteine (Table 12, entries 2-7), while the introducing of an electron-withdrawing nitro group decreased the selectivity more noticeably. Conversely, the introduction of neutral substituents into the *ortho*-position of APN aryl moiety (Table 12, entries 2, 7) appeared as slightly favourable for the selectivity.

Therefore, we inferred that the selectivity towards selenocysteine may be induced by two main factors – lower pK_a of the selenol group compared to a free thiol resulting in its preferential occurrence in more reactive deprotonated form at neutral pH, or longer C-Se bond length compared to C-S bond, which would explain slightly better selectivities obtained for the *ortho*-substituted APN. In order to test our assumptions, two APN – nonsubstituted **32** and APN-*o*-MeO **49** that demonstrated the highest level of selectivity in the above-described study – were side-by-side compared at varying pH. The reaction was decided to conduct at even higher 1000:1 initial ratio of cysteine *vs.* selenocysteine. The obtained results are shown in Table 12.

Table 12. Influence of the pH on S:Se ratio in the reaction of probes **32** and **49** with cysteine:selenocysteine (1000:1) mixture.[a]



32 **49**

#	APN	pH	S:Se (product)
1	**32**	5.5	31:69
2	**32**	7.4	34:66
3	**32**	9.6	94:6
4	**49**	5.5	24:76
5	**49**	7.4	25:75
6	**49**	9.6	78:22

[a] Reaction were carried in PBS at following concentrations of reagents: APN (0.5 mM, 0.5 eq.), selenocysteine (1 mM, 1 eq.), cysteine (1 M, 1000 eq.), TCEP (5 mM, 10 eq.).

Indeed, the pH was found to have the most significant influence on the S:Se ratio. Evidently, the deprotonation of cysteine's thiol group (pK_a = 8.33) at pH 9.6 largely increases the reactivity thereof and therefore the S:Se ratio. As a result of these first trials, high level of selectivity towards selenocysteine (1:3300) was achieved for the probe **49** when the reaction was conducted at neutral and slightly acidic pH levels. Further optimisations are, however, still required in order to make APN applicable for real application in proteomics.

After having discovered the reactivity of APN towards selenocysteine, we decided to identify a biological target where such reactivity might represent an important advantage.

4.9 THIOREDOXIN REDUCTASE INHIBITION

Thioredoxin reductases (TrxR) belong to a class of ubiquitous selenocysteine containing oxidoreductases, the only known enzymes to reduce Thioredoxin (Trx). Two families of TrxR have been identified to date: one in animals and another in prokaryotes and some eukaryotes.

Selenocysteine plays an important role in the reaction mechanism of TrxR by participating in two key thiol-disulfide exchange reactions. In one of them, Sec participates as the donor of electrons to the Trx, while in the other it acts by accepting electrons from the N-terminal redox centre of TrxR to selenosulfide bond placed in the C-terminus of the enzyme (Fig. 113). Curiously, one of the first reports on selenocysteine itself were

published on TrxR, and further experimental evidence of its importance was mainly obtained as the result of an observation that its replacement by cysteine markedly reduced catalytic activity of the enzyme.

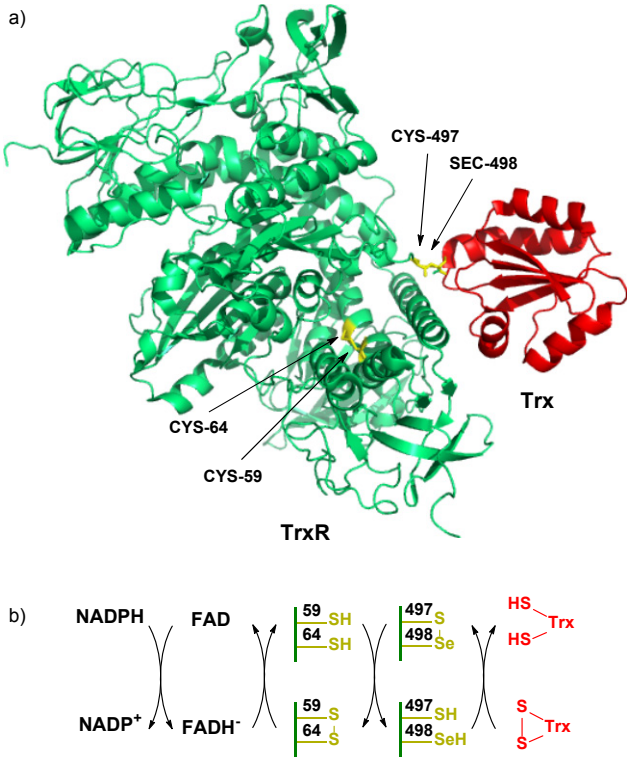


Fig. 113. a) Structure of the human thioredoxin reductase (TrxR, shown in green) complex with thioredoxin (Trx, shown in red); pdb: 3QFB b) Proposed mechanism of TrxR1-mediated reduction of Trx. [833]

An increasing number of compounds acting on TrxR have been reported over the last years, namely due to their promising potential for application in cancer therapy.[835] To date, these are represented by nitrosureas (IC50 ~ 430 μM), Au- and Pt-phospoles (IC50: 2-50 μM), quinoid compounds (IC50: 0.2-4 μM) and *para*-quinones (IC50: 1-5 μM), generally acting as irreversible inhibitors of TrxR or Trx.[836]

After having discovered the exceptional reactivity and selectivity of APN in the reaction with Cys and Sec residues, we decided to measure the inhibitory activity of differently substituted APN (Fig. 114) on human TrxR1. This study was accomplished using Thioredoxin Reductase Assay Kit (Sigma Aldrich), which exploits TrxR-mediated reduction reaction of 5,5'-dithiobis(2-nitrobenzoic) acid (DTNB) to 5-thio-2-nitrobenzoic acid (TNB) with NADPH (produces a strong yellow colour that is measured at 412 nm). This colorimetric assay was used for a direct measurement of the inhibitory activity of some of the previously obtained substituted APN.

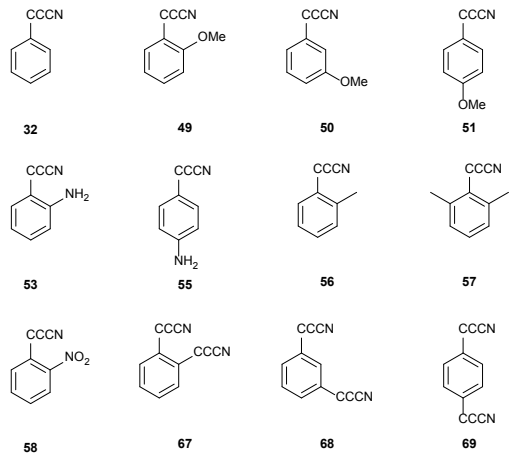


Fig. 114. APN probes tested for their TrxR inhibitory activity.

APN-tweezers (**67-69**) were chosen for this study because we expected them to be reactive either with two-thiolate or thiol-selenolate active centre of TrxR or Trx. The other probes were selected basing on the results obtained in the study on APN reactivity (Section B 4.4, page 56), and selenocysteine *vs.* cysteine selectivity (Section B 4.8, page 65) in such a way to cover APN presumably possessing substantially different reactivities.

First, the inhibitory activity of the non-substituted APN **32** was evaluated by conducting the measurement at concentrations varying in a wide range of 1 μM – 1 mM. The probe was found to possess rather low inhibitory activity for TrxR with IC50 of 35.7 ± 0.1 μM. All measurements to follow were thus decided to conduct at 1 μM – 100 μM concentrations. The most active among tested APN were then remeasured at lower concentrations in order determine their exact IC50 value. The obtained results are summed in Table 13

Table 13. TrxR inhibitory activity of APN probes.

#	APN		IC50, μM[a]
1	**32**	-	35.7 ± 0.1
2	**49**	*o*-OMe	> 100
3	**50**	*m*-OMe	22.0 ± 0.4
4	**51**	*p*-OMe	> 100
5	**53**	*o*-NH_2	> 100
6	**55**	*p*-NH_2	95.1 ± 0.8
7	**56**	*o*-Me	> 100
8	**57**	*o,o'*-diMe	> 100
9	**58**	*o*-NO_2	15.0 ± 0.1
10	**67**	*o*-di-APN	2.9 ± 0.2
11	**68**	*m*-di-APN	4.1 ± 0.1
12	**69**	*p*-di-APN	8.9 ± 0.1

[a] Working buffer: 100 mM potassium phosphate with 10 mM EDTA and 0.24 mM NADPH. See Annex 13 for the corresponding plots.

According to the obtained results, *o*-substituted APNs demonstrated low inhibitory activity, while all APN tweezers **67-69** possessed IC50 in micromolar range. Generally speaking, the profile of APN inhibitory activity corresponded to the general reactivity of APN: activated APN, containing electron withdrawing group were generally much more active as TrxR inhibitors, while APN containing electron-donating groups or *ortho*-substituents in their structure demonstrated rather low activities.

Surprisingly, the probe **49**, which was previously found to possess increased selectivity towards selenocysteine, showed very low inhibitory activity. We hypothesise it to be due either to its low overall reactivity towards cysteine/selenocysteine, or to the inaccessibility of the active centre. Further study is still required in order to determine both the target (ThrR or Trx) and the exact site of APN reactivity on TrxR and therefore to optimise the structure of the scaffold in order to further increase the inhibitory activity and ThrR selectivity of substituted APN.

5 CONCLUSIONS

To summarise this chapter, we have presented the development and application of a general approach aimed at the discovery of new chemical functionalities possessing selectivity towards specific amino acid side chains found in native proteins.

The first step of the screening consisted in testing the reciprocal reactivity of selected reactive functional groups and a series of UV-detectable amino acid derivatives allowing their fast and easy benchmarking for the applicability and appropriateness for bioconjugation.

A lead structure came out of the screening – 3-arylpropiolonitrile (APN) – was then further investigated in details for its reactivity towards Cys moiety. Furthermore, we have conducted a structure-reactivity study which allowed us to optimise the reactivity of APN and to define the most appropriate position for an anchor point introduction. This could only be done by further exploring and enhancing the available synthetic approaches. As a result, a series of ready-to-use APN-containing probes was synthesised, as well as a set of APN probes possessing a highly responsible TMPP MS-tag in their structures.

The latter were used in the second step of the screening representing the utilisation of the model peptide mixtures, which opened the possibility to test APN selectivity and reactivity in close-to-real reaction conditions involving the presence of different peptides at high dilution. This screening methology is easy to accomplish, for it consist in a simple and reproducible generation of a rather complex peptide mixtures *via* tryptic digestion of a non-expensive peptide source (*e.g.* lysozyme).

All in all, APN-based methodology allowed us to surpass the main drawbacks of existing strategies, notably side reactions with other nucleophilic amino acid residues and instability of educt and addition products. These results open an interesting prospect in the field of antibody-drug conjugates and biologics, where high selectivity and biostability are key features.

Following the intriguing results on the discrepancy between APNs' reactivity towards cysteine *vs.* selenocysteine and utilising previously obtained data on APN reactivity, we were able to achieve prominent preliminary results (3300-to-1 selectivity for selenocysteine) which are currently further investigating in the laboratory.

Finally, we conducted a study of APN utilisation as Thioredoxin reductase inhibitors and achieved interesting results in terms of activities (although still rather modest). Nonetheless of the found structure-reactivity pattern, the mechanism of APN action and real potential of APN as Thioredoxin reductase inhibitors are still to be determined.

C. DISCOVERY OF SEQUENCE-SELECTIVE REACTIONS

This chapter is devoted to the development and applications of a novel on-beads screening system allowing fast discovery of sequences possessing elevated reactivity towards residue-selective chemical functionalities. The proof-of-concept study was conducted on a library of hexapeptides with two cysteine-selective probes: one – containing the newly discovered APN group, and another – possessing a maleimide residue. Difference in the reactivity of the probes towards cysteine depending on the neighbouring amino acid side chains was determined and is currently under further investigation in the laboratory.

The previously described screening methodology (Section B, page 45) was designed to address three main properties of a bioselective reaction: the reactivity profile, the selectivity towards specific amino acid residue, and the efficacy at high dilution in complex media.

It is, however, a well-established fact that in the majority of cases, the reactivity and the selectivity of a functional group can be largely influenced by neighbouring amino acid residues. For example, Francis and collaborators[602] have demonstrated that otherwise promiscuous to any *N*-terminal residue PLP-mediated transamination reaction (see Section A 2.1.3) was very sensitive to the presence of Lys residues which, when situated in close proximity to the reaction centre, largely increased the reaction rate of transamination.

1 DESIGN OF THE SCREENING SYSTEM

Inspired by Francis' approach, we decided to envision a simple assay in which the reagent of choice would be tested on the combinatorial library of peptides synthesised using split-and-mix methodology. However, in order to make the sequencing of the on-beads peptide less laborious, an elegant algorithm developed by Reymond and collaborators[837,838] for designing combinatorial peptide libraries has been utilised.

This so-called "unique pair" approach consists in assigning to each Amino Acid (AA) two variable positions in the sequence. This arrangement limits sequence design and maximum number of different AA per variable position and is therefore perfectly suitable for focused libraries. A great advantage of this approach is the an easy to accomplish decoding by simple Amino Acid Analysis (AAA) of the beads.

AAA is a standard service analysis available in most biochemical laboratories, and it allows one to control the quality of the peptide on each bead as well as to decode the peptide sequence with no need for specific equipment. Indeed, in contrast to most library decoding protocols, a simple HPLC setting without mass-detector is the only piece of equipment required for these operation to be conducted. This is an important advantage of AAA compared to other known techniques such as, for instance, utilisation of a truncation ladder in addition to the full-length peptide on the beads, which only allows carrying out the sequencing using elaborated MALDI-TOF MS.[602]

Several limitations however have to be met when utilising AAA for peptide sequencing. Because the detection of amino acids constituting the analysed peptide is achieved after its complete hydrolysis (refluxing of beads for 1 hour in concentrated HCl), the combinatorial library cannot contain interconvertible pairs of amino acids (notably Asn-Asp and Gln-Glu pairs, which both amino acid residues in a pair would yield the same aspartic and glutamic acid respectively upon hydrolysis), as well as Trp residues, which simply do not survive harsh hydrolysis conditions.

Schematic representation of the on-beads screening approach is depicted on the Fig. 115. It starts with a classical Split&Mix solid-phase peptide synthesis, in which, however, the positions of each amino acid in the sequence are predefined. TAGSFREE software developed by Reymond's group is used to design the library. Each step of the library preparation consists of 5 splits followed by coupling with a corresponding amino acid followed by a mix step. For our first trials (see Section C 3, page 71) two focused libraries for cysteine selective reagents have been synthesises. As a result of such 5 sequential splits and mixings, $5^6 = 15625$ different peptides per library can be synthesised in less than 15 hours of work.

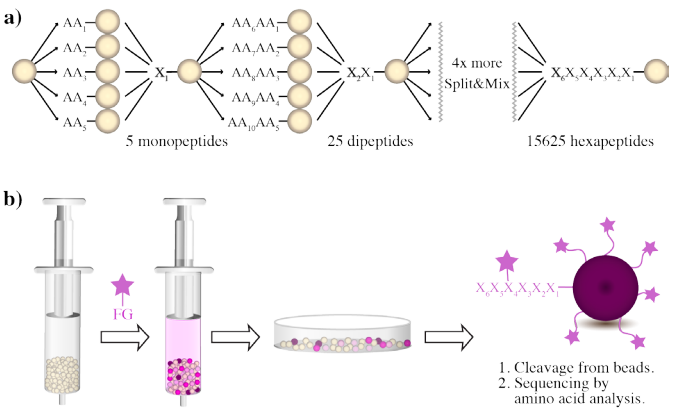


Fig. 115.Schematic representation of on-beads screening. **a)** Preparation of on-beads combinatorial libraries by Split&Mix synthesis of pentapeptides on TentaGel™ beads. **b)** Reaction of the on-beads combinatorial library with a probe containing a functional group (FG) of interest. Peptides sequences disposed to the reaction with the probe can be colorimetrically distinguished from the unreactive ones.

Incubation of the reactive probe of interest (ideally possessing a moiety simplifying the detection) with these combinatorial libraries of peptides can serve for fast detection of the sequences especially reactive towards studied functional group. If a col-

oured probe is used in this on-beads screening, simple colorimetrical distinction of different beads under microscope is enough for selecting the sequences disposing the highest reactivities towards the probe.

In order to overcome the main difficulty known for on-beads screening systems, which consists in non-covalent interactions between the probe and polymeric bead, we first decided to conduct a model study with two probes: **83**, containing a non-reactive acid functionality (negative control), and **84**, disposing a reactive activated ester moiety (positive control; Fig. 116). In both cases, 5-TAMRA (5-carboxytetramethyl-rhodamine) scaffold has been used as the moiety responsible for fluorescence. Different reaction conditions as well as rinsing and swelling systems were tested in order to develop a general protocol appropriate for further tests.

These first trials were conducted on two simple monopeptide-covered TentaGel® beads: **85**, containing a free amino group of alanine, and **86** where the amino group was acetylated by acetic anhydride in order to make beads completely unreactive towards NHS-activated ester **84**.

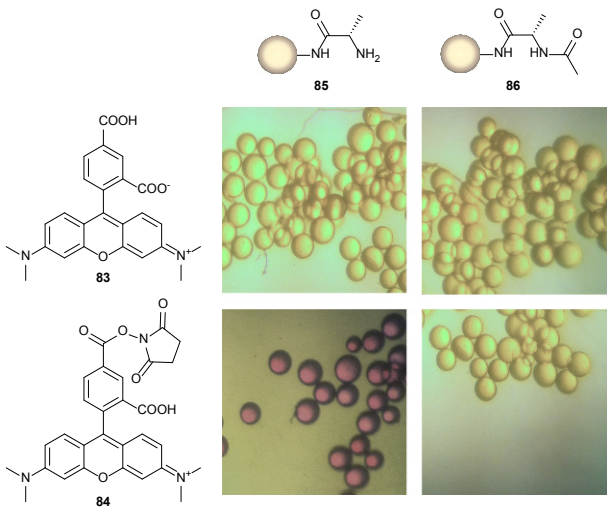


Fig. 116. Two TAMRA-containing probes used for establishing the reaction conditions of on-beads screening. Reaction conditions: 5 mg of TentaGel® HL NH_2 beads per 1 ml of TRIS:DMSO = 80:20 (pH 7.5), 100 μM concentration of fluorescent probes; 30 minutes of incubation, rinsing with methanol (2 times) and DMSO (2 times).

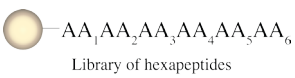
We have found that the concentration of the fluorescent probe must be rather low (10 μM – 1 mM), for when higher concentrations of the probes were used, the unspecific staining was observed: even 48 hours of rinsing with pure DMSO, DMF, or methanol were not helpful for the elimination of the non-covalent staining of beads. The residual beads' staining could however be largely decreased by using elevated amounts of DMSO during the labelling (See Section D 2.8, page 76 for practical details). Appropriate reaction conditions were thus identified to consist in the carry of the reaction in PBS-DMSO working solutions, preceded by careful beads' swelling in DCM, and followed by thorough rinsing with DMSO, DMF, and methanol until the staining completely disappears in the negative control

2 PREPARATION OF ON-BEADS LIBRARIES OF PEPTIDED AND REACTING PROBES

As it was discussed before, the peptide libraries were decided to be synthesised using the approach of "unique pairs". The positions for 15 different amino acids (2 positions per amino acid) were selected in such a way that would cover peptides possessing cysteine residues neighbouring with the reactive amino acid side-chains.

Only minor difficulties were encountered during the synthesis and analysis of the beads. For instance, the first among synthesised on-beads libraries (#1) turned out to possess an unexpected drawback making it difficult to be analysed by AAA. That being said, the presence of methionine in the sequences was found to somewhat complicate the analysis of the obtained results, for the integral intensity of the peak corresponding to methionine was systematically lower and added some degree of uncertainty to the generated data. This was mainly due to its oxidation and degradation, which made the attribution of peptide structures ambiguous in several cases (see Annex 14). This prompted us to design and synthesise an analogous library (#2) deprived from this shortcoming. The compositions of both libraries are shown in Table 14.

Table 14. Composition of the combinatorial peptide libraries.[a]



Library #1

AA_1	AA_2	AA_3	AA_4	AA_5	AA_6
Ala	Met	Arg	**Cys**	**Cys**	Lys
Tyr	Leu	Ser	Lys	Tyr	Ala
Pro	Phe	Pro	Arg	Ser	His
Leu	Gln	Asp	Phe	Met	Asp
Thr	Gly	Gly	Thr	His	Gln

Library #2

AA_1	AA_2	AA_3	AA_4	AA_5	AA_6
Lys	Asp	Ile	Ala	Ala	Ser
Arg	Gly	Gln	Ser	Arg	Lys
Phe	Thr	Phe	Ile	Glu	**Cys**
Gly	Tyr	His	Thr	Asp	His
Pro	Leu	Leu	Pro	**Cys**	Tyr

[a] Positions of Cys residues are shown in bold red.

Two probes containing the same fluorescent TAMRA moiety in their structures were synthesised: **66a** possessing a maleimide moiety, and its APN analogue **65a**. Both probe being of similar structure and only differing by their reactive hook.

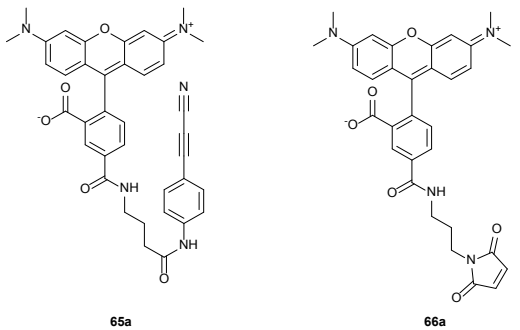


Fig. 117. Structures of fluorescent probes used in on-beads screening for sequence-dependent selectivity.

Statistically, both libraries were designed in such a way to contain at least one cysteine residue in 43.75% of all hexapeptides, 6.25% of which being vicinal dicysteine hexapeptides of general structure: $AA_1AA_2AA_3$**CysCys**AA_6 for the Library #1 and $AA_1AA_2AA_3AA_4$**CysCys** for the Library #1 (AA_n – any of the amino acid residues depicted in Table 14).

3 SCREENING

The screening was conducted under the previously optimised reaction, swelling, and rinsing conditions. Probes were used at such concentrations as to provide 10-20 eq. of the probe to cysteine (5-10 μM). Higher concentrations of the probes (100 eq. and more) gave too high level of non-specific staining, while lower concentrations (1 eq.) resulted in only slight staining, which in both cases complicated the selection of beads.

For the proof of concept, the on-bead peptide Library #2 was completely deprotected, washed, swelled, and incubated first in the working buffer (PBS:DMSO, 80:20) and then with the probe. Visually, similar distribution of beads depending on their colour was obtained in the case of both probes **65a** and **66a** (Fig. 102). Beads were divided into 4 different categories, depending on the appearing intensity of their colouration, taking into consideration following thresholds: 0% - the intensity of not-labelled swelled beads; 100% - the intensity of the beads incubated with 100 eq. of the probes (not shown). The apparent difference of colouration was hypothesised to be due to either different level of the peptide reactivity towards the tested probe, or to a different level of non-covalent interaction with beads. In order to minimise the contribution of the latter, beads were carefully washed by three different solvents (DCM, methanol, DMF; 3 times each, see Section D 2.8, page 76), and left agitated overnight in DMSO before selecting the beads.

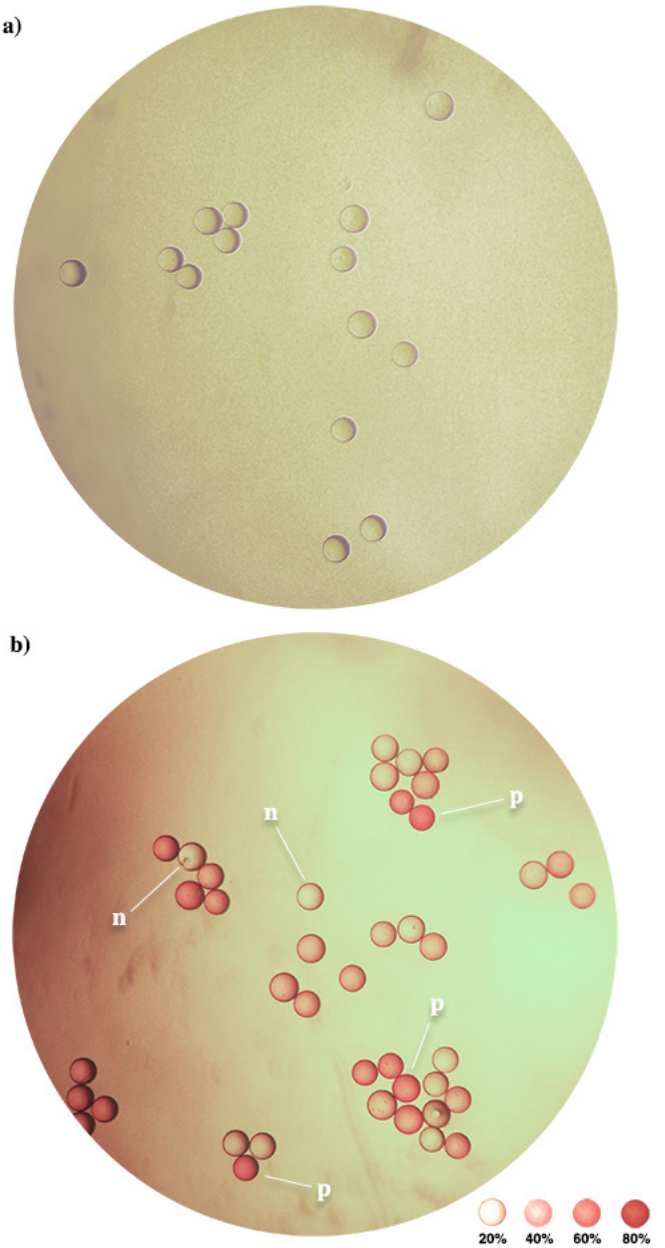


Fig. 118. Peptide library before (a) and after (b) the incubation with the reactive probe **66a**.

The most and the les intensively coloured beads (80% and 20 %, **p** and **n** respectively, Fig. 118) were picked and subjected to AAA for sequencing. The obtained data is summed up in Table 15 (*NB*: for some compositions there are more than one corresponding peptide sequence (constitution), which is an inevitable shortcoming of the approach of "fixed pairs").

Table 15. Results of on-beads screening of probes **65a** and **66a.**[a]

Probe 65a (APN)

#	Amino acid present	Sequence	
1	Asp, Gln, Gln, Ser, Phe, Lys	FDQSQK	Colourless beads (**n**)
2	Gln, Gly, Thr, Pro, Tyr, Asp	GTQPDY / PGQTDY	
3	Gln, Ser, Gly, Ala, Pro, Gly	GGQPAS	
4	Gln, Gly, Ala, Tyr, Ile, Leu	GLIAQY / GLQIAY	
5	Gln, Gly, His, Ile, Leu, Phe	FGLIQH / GLFIQH	
6	Gly, His, Ala, Ile, Leu, Leu	GLLIAH	
7	Gln, His, Thr, Pro, Pro, Ile	PTIPQH	
8	Asp, Gly, His, His, Thr, Ala	GDHTAH / GTHADH	
9	Asp, His, Thr, Ala, Pro, Leu	PDLTAH / PTLADH	
10	Gly, Arg, Thr, Tyr, Leu, **Cys**	RGLTCY / GYLTRC	Coloured beads (**p**)
11	Pro, **Cys**, Leu, Leu, Lys, Lys	KLLPCK	
12	Gln, Ser, Arg, Thr, **Cys**, Lys	RTQSCK / KTQSRC	
13	Phe, **Cys**, Tyr, Arg, Ser, Tyr	RYFSCY	
14	Ala, **Cys**, Ile, Leu, Lys, Lys	KLIACK	
15	Asp, Ser, Arg, **Cys**, Leu, Lys	RDLSCK / KDLSRC	
16	Asp, Arg, Pro, **Cys**, Phe, Arg	RDFPRC	
17	Asp, Gly, His, Tyr, Ile, **Cys**	GDHICY / GYHIDC	

Probe 66a (maleimide)

#	Amino acid present	Sequence	
1	Gly, His, Ala, Pro, Ile, Leu	PGLIAH / GLIPAH	Colourless beads (**n**)
2	Glu, Glu, Ile, Leu, Phe, Ser	FLQIQS	
3	Asp, Glu, Ser, Gly, Ala, Pro	GDQPAS / PGQADS	
4	Asp, Glu, Gly, Ala, Ile, lys	GDIAQK / GDQIAK	
5	Glu, His, Ala, Pro, Tyr Ile	PYIAQH / PYQIAH	
6	Glu, Gly, Tyr, Ile, Ile, Gly	GGIIQY	
7	Gly, Ala, Ile, Leu, Lys, Phe	KAIFLG	
8	Gly, His, Ala, Ala, Pro, Leu	PGLAAH	
9	Asp, Ala, Pro, Ile, Leu, Lys	PDLIAK / PLIADK	
10	Ser, Ser, Ala, Pro, Ile, Leu	PLISAS	
11	His, Thr, Ala, **Cys**, Lys, Lys	KTHACK	Coloured beads (**p**)
12	Ser, His, Arg, Ala, Tyr, **Cys**	RYHACS / RYHSAC	
13	Pro, Pro, Tyr, Tyr, Leu, **Cys**	PYLPCY	
14	Glu, Gly, Gly, Ile, Lys, **Cys**	GGQICK	
15	Gly, His, Thr, **Cys**, Ile, Thr	GTITCH	
16	Gly, His, Thr, Tyr, **Cys**, Ile	GTHICY / GYITCH	
17	Glu, Gly, His, Tyr, **Cys**, Ile	GYQICH / GYHIQC	
18	Asp, Ser, **Cys**, Leu, Lys, Lys	KDLSCK	
19	Gly, Gly, His, Thr, **Cys**, Ile	GGITCH	
20	Gly, His, Ala, Tyr, Leu, **Cys**	GLHACY / GYLACH	
21	Gly, Pro, Tyr, **Cys**, Ile, Lys	KGIPCY / GYIPCK	
22	Asp, Gly, His, Ala, Tyr, **Cys**	GYHADC / GDHACY	
23	Glu, Gly, Arg, Ala, Phe, **Cys**	RGFAQC / FGQARC	

[a] Incubation for 20 minutes in 5 μM solution of the probes in PBS:DMSO mixtures (80:20, pH 7.4). Cys residues are in bold red; neighbouring Lys, Tyr, and Arg – in blue.

The first important fact we have found is that both of the tested probes demonstrated high level of selectivity towards cysteine. Indeed, all the intensively coloured beads were found to contain cysteine in their sequences, while none of those remaining colourless possessed this amino acid residue. The odds for such a result to be a coincidence can be calculated using Eq.1 and turn out to be almost negligible: 7.57^{-4} % and 6.82^{-6} % for probes **65a** and **66a** respectively.

$$\boldsymbol{Eq.1}: P = (Cys_+)^{\boldsymbol{p}}(1 - Cys_+)^{\boldsymbol{n}} = (0.4375)^{\boldsymbol{p}}(0.5625)^{\boldsymbol{n}}$$

This first observation clearly demonstrates the applicability of the on-beads methodology for the screening of residue-selective functional groups and, obviously, makes it an appropriate alternative to the previously described HPLC-based screening (Section B 1.1, page 45). While representing a much faster approach for testing of reactivity (50 beads can be sequenced in 1 day), on-beads screening requires the preparation of detectable (*e.g.* fluorescent, or simply coloured) probes containing the tested functional group of interest, which makes it much more laborious in the beginning of the screening.

Curiously, even on such a restricted number of sequenced beads we could notice several interesting trends. The majority of the reactive beads contained three amino acids residues next to cysteine – lysine, tyrosine, and arginine (shown in blue, Table 15). This tendency is less remarkable for the probe **66a** containing more reactive and promiscuous maleimidyl residue.

Interestingly, none of dicysteinyl peptides (represent 6.25% of all peptides) were detected among reactive peptide sequences. This may either be due to the oxidation of dicysteinyl peptides during swelling and washing, or simply be a statistical coincidence due to a small number of analysed beads. This brings us to a very important thing to be taken into consideration while conducting the screening of combinatorial libraries. Even though the obtained results clearly demonstrated the applicability of the approach of on-beads screening for sequence selectivity of chemical reagents, the percentage of the targeted sequences must be rather low in order to make the obtained data credible statistically. Therefore, the design of the library represents a very important step to be accomplished before starting the study.

4 CONCLUSIONS

All in all, in this chapter we have presented our investigations on the development and first proof-of-concept studies of an on-beads model screening system which allows determination of selectivities for a specific sequence of amino acids. The libraries were synthesised following the approach of fixed pairs, which opens the possibility of using a simple amino acid analysis (AAA) to conduct sequencing. AAA does not imply utilisation of MS or any other expensive and sophisticated techniques, which represents an important advantage compared to ordinary used approaches. We thus expect our screening approach to take its place among other available screening.

As a result of the screening, we have obtained first insights into the discrepancy between APN reactivity depending of the amino acid side chain neighbouring the reactive cysteine moiety. Interestingly, such a difference was somewhat less remarkable for a more reactive maleimide-containing probe. However, in order to make safe assumptions, much more data should be collected by sequencing more beads. This work is currently ongoing and full reactivity patterns of new and already described functional groups are to be determined.

D. EXPERIMENTAL PART

1 GENERAL METHODS

1.1 Experimental procedures

Unless otherwise indicated, reactions were carried out under an atmosphere of argon in flame-dried glassware with magnetic stirring. Air and/or moisture-sensitive liquids were transferred *via* syringe. When required, solutions were degassed by bubbling of argon through a needle. Organic solutions were concentrated by rotary evaporation at 25-60 °C at 15-30 torr. Analytical thin layer chromatography (TLC) was performed using plates cut from glass sheets (silica gel 60F-254 from Merck). Visualisation was achieved under a 254 or 365 nm UV light and by immersion in an appropriate revelation solution.

1.2 Materials

All reagents were obtained from commercial sources and used without any further purifications. The anhydrous solvents utilised in experiments were obtained from Sigma-Aldrich or Alfa Aesar. All reagents used in the experiments were purchased from Aldrich, Alfa Aesar, Acros or TCI and were used without any further purification. Silica gel for column chromatography was purchased from Merck (Geduran® Si 60, 40-63 µm). Column flash chromatography was carried out using silica gel G-25 (40-63 µm) from Macherey-Nagel.

1.3 Instrumentation

1.3.1 NMR spectrometry

^{1}H and ^{13}C NMR spectra were recorded at 23°C on Bruker 300, 400 and 500 spectrometers. Recorded shifts are reported in parts per million (δ) and calibrated using residual non-deuterated solvent. Data are represented as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, quint = quintet, m = multiplet, br = broad), coupling constant (*J*, Hz) and integration.

1.3.2 Mass spectrometry

High resolution mass spectra (HRMS) were obtained using a Agilent Q-TOF (time of flight) 6520 and low resolution mass spectra using a Agilent MSD 1200 SL (ESI/APCI) with a Agilent HPLC1200 SL. GC-MS analyses were performed by means of Agilent 7890A Gas Chromatograph equipped with DB-5MS 30 m × 0.25 mm column and JEOL AccuTOF-GCv. The semi-preparative HPLC system consisted of a Waters 600 pump, a 2487 detector (Waters), a 5 mL sample loop and a SunFire C18 column (150 mm × 19 mm i.d., 5 µm, Waters) with a 40 min gradient from 5% to 95% acetonitrile.

1.3.3 Microscopy

Fluorescence images were recorded on an inverted widefield microscope DM IRB (Leica Microsystems) equipped with a mercury metal halide EL 6000 (Leica Microsystems) for fluorescence excitation and a variety of objectives (×10/×20/×40/×63) have been applied. Filter cubes utilised were A4 for Hoechst imaging (BP 360/40 – 400 – BP 470/40) and Cy3 for TAMRA imaging (BP 545/40 – 565 – BP 610/75). Confocal images were recorded on a TCS SPE-II (Leica Microsystems) using HXC PL APO 20×/0.7 CS (×20) and HXC PL APO 63×/1.40 OIL CS (×63) objectives. Excitation was done with 405 nm (Hoechst experiment), 488 nm (Oregon Green experiment) or 561 nm (TAMRA experiment) and fluorescence was detected from 430 – 480 nm (Hoechst), 500 – 560 nm (Oregon Green) and 570 – 625 nm (TAMRA). Analysis was carried out with the software Coolnap and ImageJ.

1.3.4 UV spectroscopy

UV-Vis spectra were recorded on Shimadzu UV-1800 spectrophotometer. IR spectra were recorded on a Nicolet 380 FT-IR spectrometer from Thermo Electron Corporation as a DCM solution or solid on a diamond plate.

1.3.5 HPLC

Analytic HPLC experiments were carried out on a Shimadzu system (pump: LC 20-AD, detector: SPD 20-A, autosampler: SIL 20-A) using a SunFire™ C18 5 µM 4.6 × 150 mm column (Waters). HPLC parameters were as follows: flow rate 1 mL/min, gradient from 5 to 95% of mobile phase B from 0 to 20 min, followed by 5 min at 95% of mobile phase and post time of 5 min. Mobile phase A was 0.05% TFA in water (mQ) (v/v), and mobile phase B was acetonitrile (HPLC grade). Unless otherwise specified, the detection was done at 254 nm.

Preparative HPLC procedures were performed on semi-preparative HPLC system (pump: Waters 600, detector: Waters 2489, collector: Waters fraction collector III, 5 mL sample loop) using a Sunfire C18 (150 mm × 19 mm i.d., 5 µm, Waters) at a flow of 17 mL/min. Per sample 1 mL were injected and as eluent system water/ACN containing 0.05% TFA was used. The gradient applied was 5% to 95% ACN in 40 min and 10 min of re-equilibration. Detection was done at 254 nm.

2 GENERAL PROCEDURES

2.1 Stock solutions

100 mM stock solutions of each reagent used in the screening were prepared in DMSO and stored at -20 °C. A 100 mM stock solution of benzamide (used as an internal standard) was prepared in distilled water and stored at 25 °C.

2.2 HPLC screening of reactivities

10 µL of the stock solution of amino acid derivative (**1-16**) and 2.5 µL of the stock solution of benzamide were added to a vial containing 977.5 µL of PBS (1x, pH 7.6). The solution was vortexed and 10 µL of the stock solution of electrophile are added to give 1 mM final concentrations of reagents and 0.25 mM concentration of benzamide. Aliquots of the reaction

mixture (50 μL) were analysed at T_0 of about 1 minute and at T_{end} of 5 hours by HPLC. Areas under peaks of starting materials and hydrolysis products are normalised according to the area of peak of the internal standard.

2.3 Study of hydrolytic stability of reacting probes 32, 35-39

To a vial, containing 980 μL of PBS (1x, pH 7.6), were subsequently added 10 μL of the stock solution of benzamide and 10 μL of the stock solution of the probe to give final concentration of 1 mM (both internal standard and the probe). Aliquots of the reaction mixture (50 μL) were analysed by HPLC for 5 hours (injection every 30 min). Areas under the peaks of the starting materials and the products of hydrolysis were normalised according to the area of the peak of the internal standard.

2.4 Reactivity of APN 32, 49-60 towards cysteine derivative 16.

To a vial, containing 985 μL of PBS (1x, pH 7.6), were subsequently added 5 μL of the stock solution of benzamide (10 mM in water), 5 μL of the stock solution of APN (10 mM in DMSO) and 5 μL of stock solution of AcCys (**16**, 10 mM in DMSO) to give final concentration of 50 μM (each reagent and internal standard). Aliquots of the reaction mixture (50 μL) were analysed by HPLC (injection at 0 and 30 minutes of reaction). Areas under the peaks of the starting materials and hydrolysis products were normalised according to the area of the peak of the internal standard.

2.5 Stability of 33 in different working media

A 100 mM stock solution of **33** was prepared in DMSO and stored at -20 ˚C. 1 μL of the stock solution was added to 999 μL of working media to give 100 μM final concentration of substrate. Aliquots were analysed at different time points. Areas under peak of starting material (**33**) were normalised according to the area of the peak of the internal standard (benzamide). All measurements were carried on at 25 ˚C.

2.6 Fluorescence stability measurements

Solutions of the FRET-probe in the buffer are prepared in a 96-well plate at specified concentration. The 96-well plate is placed in the Victor plate reader and the measurements are started. Fluorescence is detected every 120 s with the interval including 10 s of shaking and 5 s of delay before recording every measurement. Excitation was done with λ_{max} of 560 nm, the fluorescence was detected at 580 nm.

2.7 Solid-phase peptide synthesis

Unless otherwise specified, SPPS is done in 10 mL syringes on 100 mg of TentaGel® HL (Fmoc protected).

2.7.1 Coupling

Following protocol allowed attaching one amino acid residue on the surface of Fmoc protected beads.

1. ***Swelling.*** The resin was agitated in DCM (5-6 mL) for 10 min. Solvent was sucked out.
2. ***Fmoc-deprotection.*** (2 times): amino groups of the resin were deprotected by adding 7 mL of piperidine:DMF solution (20:80) and were agitated for 10 minutes. The resin was washed with DCM (to get beads back to syringe), NMP (3 times), MeOH (3 times) and DCM (3 times).
3. ***Control.*** The presence of free amino groups was checked by the reaction with 1% DMF solution of DNTB with 10% DMF solution of DIPEA (yellow beads = no amino groups; red beads = free amino groups are present). *NB*: the presence (or absence) of proline was verified using the colour reaction with of 2% chloranil in DMF and 2% acetaldehyde in DMF.
4. ***Peptide coupling.*** Previously prepared solution of Fmoc-protected amino acid and Pybop in NMP (1 mL) (5-10 min before addition) was added to carefully washed resin (last wash with DCM for swelling), a syringe is filled to 70% of its full volume, DIPEA is added and stirring continued for 1 hour at 25 ˚C. Solvents were sucked out, the resin was washed 3 times with NMP, 3 times with MeOH, 3 times with DCM.
5. ***Control.*** The absence of free amino groups was controlled by the same test as described in the step 4 (beads should not become red-coloured).
6. ***Acetylation.*** Washed and swelled resin was stirred for 15 minutes with a mixture Ac_2O:DCM solution (1:1).
7. ***Washing.*** Resin was washed with NMP (3 times), MeOH (3 times) and DCM (3 times).

2.7.2 Complete deprotection

1. Resin was washed, swelled and deprotected (**Coupling procedure**, step 7, 1-2).
2. Cleavage mixture (94TFA : 1TIS : 2.5EDT : 2.5H2O, v:v) was added and agitation was continued for 4.5 hours. Solvents were sucked out, resin was washed with NMP (5 times), MeOH (5 times) and DCM (5 times).

2.8 On-beads screening of reactive probes

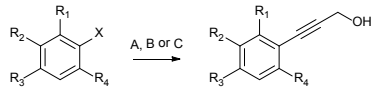
Completely deprotected on-beads peptide library (5-20 mg) was swelled for 5 minutes in DCM, washed with MeOH (5 minutes) and working solution (5 minutes). Required amount of the stock solution of the probe was added to the working solution. After being vortexed, the obtained solution was added to syringes with beads, and the incubation continues for 10-60 minutes. The beads were washed with working solution, NMP, MeOH and DCM (3 times each).

Standard working solution: PBS:DMSO (80:20), pH 7.4.

Standard working concentration of the testing probe: 10 μM.

2.9 SONOGASHIRA COUPLING

Reaction scheme:



Standard reaction protocols:

A. To a degased solution of the proper aryl halide (1 eq., 1 mmol) in DMF (5 mL) were added DIPEA (10 eq., 10 mmol), premixed $PdCl_2(PPh_3)_2$ (0.03 eq., 30 µmol), and CuI (0.06 eq., 60 µmol). The resulting mixture was degased, stirred for another 5 minutes, followed by the addition of propargyl alcohol (1.2 eq., 1.2 mmol). The resulting mixture was stirred for 1-24 hours (monitored by TLC). 1M HCl (50 mL) was added (if contains free amino groups, 50 mL of water were added instead) and the reaction mixture was extracted with ethyl acetate (3x20 mL). The combined ethyl acetate fractions were washed with water (1x10 mL) dried over $MgSO_4$ and evaporated under reduced pressure. Crude products were purified by flash chromatography (gradient of 20 minutes from 100% of cyclohexane to 100% of ethyl acetate).

B. To a degased solution of the proper aryl halide (1 eq., 1 mmol) in THF (5 mL) and TEA (5 mL), were added premixed $PdCl_2(PPh_3)_2$ (0.03 eq., 30 µmol) and CuI (0.06 eq., 60 µmol), followed by the addition of propargyl alcohol (1.2 eq., 1.2 mmol). The reaction mixture was stirred for 1-24 hours (monitored by TLC). THF and TEA were evaporated under reduced pressure and the crude product was purified by flash chromatography (gradient of 20 minutes from 100% of cyclohexane to 100% of ethyl acetate).

C. To a degased solution of the proper aryl halide (1 eq., 1 mmol) in propylamine or pyrrolidine (3 mL), $Pd(PPh_3)_4$ (0.05 eq., 50 µmol) was added. The resulting mixture was heated overnight (30-50 ˚C), evaporated and the crude product was purified by flash chromatography (gradient of 20 minutes from 100% of cyclohexane to 100% of ethyl acetate).

2.10 PREPARATION OF HIGHLY ACTIVE MnO_2

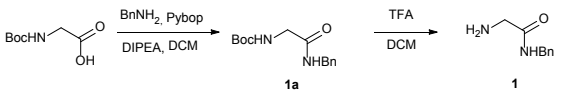
A solution of $MnCl_2 \cdot 4H_2O$ (1 eq., 1 mole, 200 g) in water (2 L) at 70 °C was gradually added during 10 minutes, with stirring, to a solution of $KMnO_4$ (1 eq., 1 mole, 160 g) in water (2 L) at 60 °C. A vigorous reaction ensued with evolution of chlorine; the suspension was stirred for 2 hours and kept overnight at 25 ˚C. The precipitate was filtered off, washed thoroughly with water (4 L) until pH 6.5-7 and the washing gave a negligible chloride test. The filter cake was then dried at 120-130 ˚C for 18 h to give a chocolate-brown, highly disperse amorphous powder.

2.11 MnO_2 OXIDATION

Slightly modified procedure described by McAllister *et al.* [824] To a solution of the proper propargylic alcohol (1 eq., 1 mmol) in THF (4.5 mL) were added $MgSO_4$ (15 eq., 15 mmol), highly active MnO_2 (25 eq., 25 mmol) and 2M NH_3 solution in IPA (4 eq., 4 mmol, 2 mL). The resulting mixture was vigorously stirred at 25 ˚C for 0.5-12 hours (monitored by TLC). DCM (20 mL) was added and the crude mixture was filtered through Celite, evaporated under reduced pressure to give crude product and purified by flash chromatography if required.

3 SYNTHESIS

3.1 (S)-2-AMINO-*N*-BENZYLACETAMIDE (1):



1a: *tert*-Butyl (2-(benzylamino)-2-oxoethyl)carbamate

To a cooled to 0 ˚C solution of Boc-Gly-OH (1 eq., 500 mg, 2.85 mmol) in DCM (10.8 mL) were added DIPEA (3 eq., 1.42 mL, 8.56 mmol) and PYBOP (1.1 eq., 1630 mg, 3.14 mmol). The resulting mixture was stirred for another 5 min at 0 ˚C. Benzylamine (1.5 eq., 468 µL, 4.28 mmol) was added and the solution was allowed to warm up to room temperature and stirred for 3 hours. The reaction mixture was washed with 1M HCl (20 mL), $NaHCO_3$ (sat.) (10 mL) and brine (10 mL), dried over $MgSO_4$ and evaporated. The crude material was purified by flash chromatography (20 minutes gradient EtOAc/Cyclohexane) and evaporated under reduced pressure to yield **1a** (680 mg, 2.57 mmol, 90 %) as a white solid.

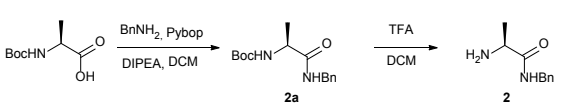
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.23-7.53 (m, 5H), 6.53 (br. s., 1H), 4.47 (d, J = 5.77 Hz, 2H), 3.84 (d, J = 5.02 Hz, 2H), 1.44 (s, 9H); ESI-MS: $C_{14}H_{21}N_2O_3^+$ $[M+H]^+$, 264.1; found 264.2.

1: 2-Amino-*N*-benzylacetamide (TFA salt).

To a solution of **1a** (1 eq., 300 mg, 1.13 mmol) in DCM (5 mL) was added TFA (20 eq., 1.7 mL, 22.7 mmol). The resulting mixture was stirred for 2 hours at 25 ˚C. The solvents were evaporated under reduced pressure to yield **1** (309 mg, 1.11 mmol, 98 %) as a white solid which was used without any further purification.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.24-7.41 (m, 5H), 4.45 (s, 2H), 3.72 (s, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 169.6, 138.9, 129.7, 128.7, 128.5, 44.3, 41.5; ESI-MS: $C_9H_{13}N_2O^+$ $[M+H]^+$, 165.1; found 165.1.

3.2 (S)-2-AMINO-*N*-BENZYLPROPANAMIDE (2):



2a: (S)-*tert*-Butyl (1-(benzylamino)-1-oxopropan-2-yl)carbamate.

This compound was synthesised following the same procedure as for the synthesis of **1a**. White solid; yield: 92%.

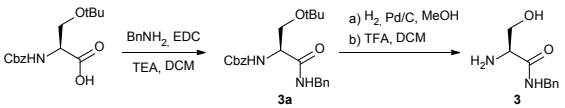
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.21-7.40 (m, 5H), 5.04 (br. s., 1H), 4.35-4.60 (m, 2H), 4.21 (br. s., 1H), 1.43 (s, 9H), 1.40 (d, *J* = 7.03 Hz, 3H); ESI-MS: $C_{15}H_{23}N_2O_2^+$ $[M+H]^+$, 278.2; found 278.2.

2: (S)-2-Amino-*N*-benzylpropanamide (TFA salt).

This compound was synthesised following the same procedure as for the synthesis of **1**. Yellowish oil; yield: 98%.

Yield: 98%, yellowish oil. ^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.20-7.44 (m, 5H), 4.44 (s, 2H), 3.96 (q, *J* = 7.03 Hz, 1H), 1.53 (d, *J* = 7.03 Hz, 3H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 170.8, 139.5, 129.7, 128.7, 128.5, 50.3, 44.3, 17.7; ESI-MS: $C_{10}H_{15}N_2O^+$ $[M+H]^+$, 179.1; found 179.2.

3.3 (S)-2-Amino-*N*-benzyl-3-hydroxypropanamide (3):



3a: (S)-*tert*-Butyl (1-(benzylamino)-3-(*tert*-butoxy)-1-oxopropan-2-yl)carbamate.

To a solution of N-Cbz-O(*t*Bu)-L-serine (1 eq., 370 mg, 1.25 mmol) and TEA (3 eq., 960 mg, 1.23 mL, 7.43 mmol) in DCM (20 mL) cooled to 0 ˚C was added EDC (1.10 eq., 264 mg, 1.38 mmol). The obtained solution was stirred for 10 min at 0 ˚C, benzylamine (1.1 eq., 151 μL, 1.38 mmol) was added and the reaction mass was stirred overnight at 25 ˚C. The resulting mixture was washed with 1M HCl (2x20 mL). The organic layer was dried over $MgSO_4$ and evaporated under reduced pressure. The crude product was purified by flash chromatography (DCM:MeOH, 20 min gradient from 0% to 5% of MeOH) to give **3a** (303 mg, 63%) as a white solid.

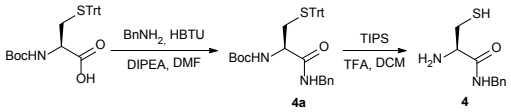
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.14-7.38 (m, 10H), 6.78 (br. s., 1H), 5.66 (br. s., 1H), 4.93-5.20 (m, 2H), 4.31-4.52 (m, 2H), 4.20 (br. s., 1H), 3.70-3.88 (m, 1H), 3.21-3.43 (m, 1H), 1.06 (s, 9H); ESI-MS: $C_{22}H_{29}N_2O_4^+$ $[M+H]^+$, 384.2; found 384.2.

3: (S)-2-Amino-*N*-benzyl-3-hydroxypropanamide.

To a degased solution of **3a** (150 mg, 0.39 mmol) in MeOH (10 mL) was added Pd/C (15 mg). The resulting mixture was stirred overnight under hydrogen atmosphere at 25 ˚C. The solution was filtered through Celite 545, evaporated, the obtained oil was dissolved in 1:1 mixture of DCM and TFA (5 mL), stirred for 2 hours at 25 ˚C and evaporated under reduced pressure to give the crude product, which was purified by flash chromatography (reverse phase, H_2O (0.5% TFA):MeOH, gradient from 0% to 100% of MeOH in 20 min) to give **3** (TFA salt, 40 mg, 40%) as a colorless oil.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.22-7.43 (m, 5H), 4.45 (d, *J* = 5.77 Hz, 2H), 3.91-4.01 (m, 2H), 3.80-3.91 (m, 1H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 168.1, 139.4, 129.7, 128.7, 128.5, 61.7, 56.4, 44.4; ESI-MS: $C_{10}H_{15}N_2O_2^+$ $[M+H]^+$, 195.1; found 195.2.

3.4 (R)-2-Amino-*N*-benzyl-3-mercaptopropanamide (4):



4a: (R)-*tert*-Butyl (1-(benzylamino)-1-oxo-3-(tritylthio)propan-2-yl)carbamate.

To a cooled to 0 ˚C solution of *N*-Boc-*S*-Trityl-L-cysteine (1 eq., 1.15 g, 2.48 mmol) and DIPEA (3 eq., 1.23 mL, 7.43 mmol) in DMF (8 mL), HBTU (1.01 eq., 0.95 g, 2.5 mmol) was added. The obtained solution was stirred for 10 min at 0 ˚C, benzylamine (1.1 eq., 0.3 mL, 2.72 mmol) was added and the reaction mass was stirred overnight at 25 ˚C. 1M HCl (80 mL) was added and the obtained solution was extracted with ethyl acetate (3x60 mL). The combined organic fractions were dried over $MgSO_4$ and evaporated. Product was purified by flash chromatography (DCM:MeOH, 20 min gradient from 0% to 10% of MeOH). **4a** was obtained as a white solid (1.1 g, 82%).

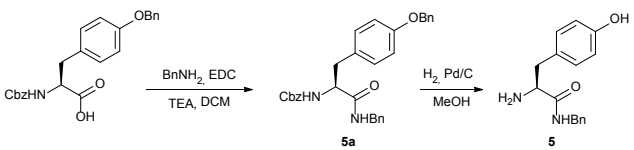
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.39 (d, *J* = 7.53 Hz, 6H), 7.15-7.35 (m, 14H), 4.62 (br. s., 1H), 4.42 (d, *J* = 15.06 Hz, 1H), 4.27 (d, *J* = 15.06 Hz, 1H), 4.02 (t, *J* = 6.15 Hz, 1H), 2.40-2.63 (m, 2H), 1.44 (s, 9H); ESI-MS: $C_{34}H_{37}N_2O_3S^+$ $[M+H]^+$, 552.2; found 552.2.

4: (R)-2-Amino-*N*-benzyl-3-mercaptopropanamide.

To a degased solution of **4a** (1 eq., 500 mg, 0.905 mmol) in DCM (15 mL), TFA (20 eq., 1.34 mL, 18.1 mmol) was added dropwise to give brightly yellow solution. Triisopropylsilane (TIPS, 5 eq., 0.927 mL, 4.52 mmol) was added, and the resulting mixture was stirred for another 2 hours at 25 ˚C (solution turned colorless). The crude mixture was evaporated under reduced pressure under reduced pressure, washed with cyclohexane, and the residue was purified by reverse phase flash chromatography (H_2O (0.5% TFA):MeOH, gradient from 0% to 100% of MeOH in 20 min) to yield **4** (TFA salt, 275 mg, 0.85 mmol, 94 %) as a colourless oil.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.28-7.34 (m, 5H), 4.46 (d, *J* = 4.27 Hz, 2H), 4.05 (t, *J* = 5.90 Hz, 1H), 3.03 (dd, *J* = 5.90, 19.70 Hz, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 168.3, 139.3, 129.7, 128.8, 128.6, 56.2, 44.5, 26.4; ESI-MS: $C_{10}H_{15}N_2OS^+$ $[M+H]^+$, 211.0; found 211.0.

3.5 (S)-2-Amino-*N*-benzyl-3-(4-hydroxyphenyl)propan-amide (5):



5a: (S)-Benzyl (1-(benzylamino)-3-(4-(benzyloxy)phenyl)-1-oxopropan-2-yl)carbamate.

To a solution of N-Cbz-OBn-L-Tyrosine (1 eq., 478 mg, 1.18 mmol) and TEA (3 eq., 0.490 mL, 3.54 mmol) in DCM (20 mL) at 0 ˚C was added EDC (1.10 eq., 249 mg, 1.30 mmol). The resulting solution was stirred for 10 min at 0 ˚C before benzylamine (1.1 eq., 0.142 mL, 1.30 mmol) was added and the reaction mixture was stirred overnight at 25 ˚C. The clude mixture was washed with 1M HCl (2x20 mL), dried over $MgSO_4$ and evaporated under reduced pressure. The crude product was purified by flash chromatography (DCM:MeOH, gradient from 100% of DCM to 90% of DCM in 20 min) to give **5a** (525 mg, 90%) as a yellowish oil.

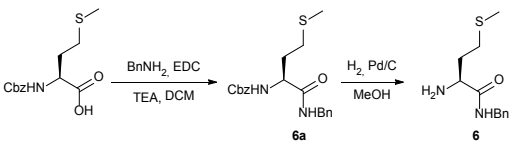
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.19-7.42 (m, 13H), 6.92-7.08 (m, 4H), 6.78 (m, 2H), 5.86 (br. s., 1H), 5.25 (br. s., 1H), 5.00 (s, 2H), 4.94 (s, 2H), 4.18-4.30 (m, 3H), 3.03 (dd, *J* = 5.52, 13.80 Hz, 1H), 2.89 (dd, *J* = 7.78, 13.80 Hz, 1H); ESI-MS: $C_{31}H_{31}N_2O_4^+$ $[M+H]^+$, 494.2; found 494.2.

5: (S)-2-Amino-*N*-benzyl-3-(4-hydroxyphenyl)propanamide.

To a degased solution of **5a** (500 mg, 1.01 mmol) in MeOH (10 mL) was added Pd/C (50 mg). The resuting mixture was stirred overnight under hydrogen atmosphere at 25 ˚C. The obtained solution was filtered through Celite 545 and evaporated under reduced pressure. The crude product was purified by flash chromatography (reverse phase, H_2O (0.5% TFA):MeOH, gradient from 0% to 100% of MeOH in 20 min) to give **5** (TFA salt, 311 mg, 80%) as a white solid.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.06-7.28 (m, 6H), 6.80-6.92 (m, *J* = 8.3 Hz, 2H), 6.47-6.69 (m, *J* = 8.3 Hz, 2H), 4.27 (d, *J* = 15.1 Hz, 1H), 4.13 (d, *J* = 15.1 Hz, 1H), 3.40 (t, *J* = 7.2 Hz, 1H), 2.76 (td, *J* = 7.2, 13.5 Hz, 1H), 2.66 (td, *J* = 7.2, 13.5 Hz, 1H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 177.0, 158.0, 139.9, 131.8, 129.9, 129.7, 129.0, 128.6, 116.8, 58.5, 44.4, 42.2; ESI-MS: $C_{16}H_{19}N_2O_2^+$ $[M+H]^+$, 271.1; found 271.2.

3.6 (S)-2-AMINO-*N*-BENZYL-4-(METHYLTHIO)BUTANAMIDE (6):



6a: (S)-Benzyl (1-(benzylamino)-4-(methylthio)-1-oxobutan-2-yl)carbamate.

To a solution of N-Cbz-L-Methionine (1 eq., 209 mg, 0.74 mmol) and TEA (3 eq., 0.308 mL, 2.21 mmol) in DCM (10 mL) was added EDC (1.10 eq., 156 mg, 0.81 mmol). The resulting mixture was stirred for 10 min at 0 ˚C before benzylamine (1.1 eq., 0.089 mL, 1.38 mmol) was added and the reaction mass was stirred overnight at 25 ˚C. The resulting mixture was washed with 1M HCl (2x10 mL), dried over $MgSO_4$ and evaporated. The crude product was purified by flash chromatography (DCM:MeOH, gradient from 100% of DCM to 90% of DCM in 20 min) to give **6a** (260 mg, 95%) as a white solid.

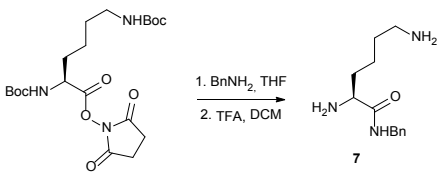
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.23-7.46 (m, 10H), 6.48 (br. s., 1H), 5.53 (br. s., 1H), 5.12 (s, 2H), 4.34-4.56 (m, 3H), 2.47-2.68 (m, 2H), 1.96-2.21 (m, 5H); ESI-MS: $C_{20}H_{25}N_2O_3S^+$ $[M+H]^+$, 373.2; found 373.1.

6: (S)-2-Amino-*N*-benzyl-4-(methylthio)butanamide.

To a degased solution of **6a** (200 mg, 0.54 mmol) in MeOH (10 mL) was added Pd/C (20 mg). The resulting mixture was stirred overnight under hydrogen atmosphere at 25 ˚C. The solution was filtered through Celite 545 and evaporated under reduced pressure. The crude product was purified by flash chromatography (reverse phase, H_2O (0.5% TFA):MeOH, 20 min gradient from 0 to 100% of MeOH) to give **6** (TFA salt, 120 mg, 63%) as a colourless oil.

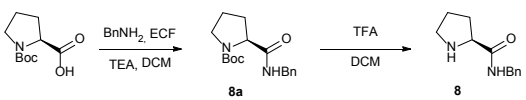
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.08-7.36 (m, 5H), 4.99 (s, 1H), 4.11-4.41 (m, 2H), 3.34 (t, *J* = 6.53 Hz, 1H), 2.26-2.53 (m, 2H), 1.90-1.98 (m, 3H), 1.61-1.90 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 177.2, 140.0, 129.6, 128.7, 128.3, 55.4, 44.1, 35.9, 31.2, 15.3; ESI-MS: $C_{12}H_{19}N_2OS^+$ $[M+H]^+$, 239.1; found 239.2.

3.7 (S)-2,6-DIAMINO-*N*-BENZYLHEXANAMIDE (7):



This compound was synthesised one-pot using a protocol reported by Huang *et al*.[839] starting from a commercially available Boc-Lys(Boc)-OSu.

3.8 (S)-*N*-BENZYLPYRROLIDINE-2-CARBOXAMIDE (8):



8b: *tert*-Butyl (S)-2-(benzylcarbamoyl)pyrrolidine-1-carboxylate.

N-Boc-L-proline (1 eq., 1000 mg, 4.7 mmol) and TEA (1 eq., 300 µL, 4.7 mmol) were dissolved in DCM (20 mL). The reaction mixture was cooled to 0 ˚C and ethyl chloroformate (1 eq., 265 µL, 4.7 mmol) was added dropwise over 10 minutes. Benzylamine (1.1 eq., 557 µL, 5.11 mmol) was added and the resulting mixture was stirred for 1 hour at 0 ˚C, at 25 ˚C for another 15 hours, and then heated to reflux for 1 hour. The reaction mixture was filtered through a Celite after having been cooled to room temperature. The crude product obtained after the evaporation of solvents was purified by flash chromatography (20 minutes gradient EtOAc/Cyclohexane) to give **8b** (545 mg, 35%) as white solid.

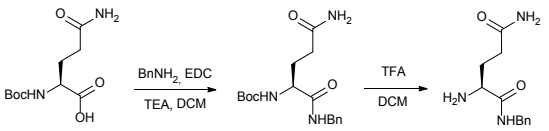
^{1}H NMR (400 MHz, DMSO-d_6) δ 8.10 (s, 1H), 7.15 – 7.25 (m, 5H), 4.36 (dd, *J* = 6.20, 15.13 Hz, 1H), 4.26 (dd, *J* = 5.95, 15.13 Hz, 1H), 4.10-4.15 (m, 1H), 3.35-3.46 (m, 1H), 3.28-3.35 (m, 1H), 2.15 (m, 1H), 1.70-1.90 (m, 3H), 1.35 (s, 9H); ESI-MS: $C_{17}H_{25}N_2O_3^+$ $[M+H]^+$, 304.2; found 304.2.

8: (S)-*N*-Benzylpyrrolidine-2-carboxamide.

8a (200 mg, 6.6 mmol) was dissolved in DCM (2 mL) and TFA (200 μL) was added. The resulting mixture was stirred for another 2 hours until complete completion of the reaction (TLC monitoring), the solvent was evaporated under reduced pressure and the obtained crude product was purified by reverse-phase flash chromatography (20 minutes gradient H_2O (0.5% TFA):MeOH).

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 8.10 (s, 1H), 7.21-7.29 (m, 5H), 4.35-4.39 (s, 2H), 3.81-3.85 (m, 1H), 3.00-3.12 (m, 1H), 2.85-2.89 (m, 1H), 2.76 (s, 1H), 2.11-2.17 (m, 1H), 1.90-1.95 (m, 1H), 1.65-1.71 (m, 2H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 174.7, 139.0, 127.9, 60.5, 47.1, 42.5, 30.8, 26.1; ESI-MS: $C_{12}H_{17}N_2O^+$ $[M+H]^+$, 205.1; found 205.3.

3.9 (S)-2-Amino-N^1-benzylpentanediamide (9):



9a: (S)-*tert*-Butyl (5-amino-1-(benzylamino)-1,5-dioxopentan-2-yl)carbamate.

To a cooled to 0 °C solution of Boc-Gln-OH (1 eq., 1.587 g, 6.44 mmol) in DCM (25 mL), TEA (3.4 eq., 3.05 mL, 21.91 mmol) and benzylamine (1.2 eq., 0.845 mL, 7.73 mmol) were subsequently added. EDC (1.2 eq., 1.482 g, 7.73 mmol) was added to the stirring solution and the resulting mixture was stirred for 3 hours at 25 °C, washed with 1M HCl (2x25 mL), brine (1x25 mL), dried over $MgSO_4$, and evaporated under reduced pressure to give **9a** (1.42 g, 66%) as a white solid.

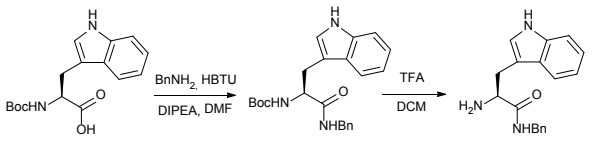
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.08-7.27 (m, 5H), 4.19-4.37 (m, 2H), 3.84-4.13 (m, 1H), 2.20 (t, *J* = 7.53 Hz, 2H), 1.94 (dd, *J* = 5.65, 13.43 Hz, 1H), 1.77 (dd, *J* = 8.03, 14.05 Hz, 1H), 1.34 (s, 9H); ESI-MS: $C_{17}H_{26}N_3O_4^+$ $[M+H]^+$, 335.2; found 335.2.

9: (S)-2-Amino-N^1-benzylpentanediamide.

To the solution of **9a** (1 eq., 490 mg, 1.46 mmol) in DCM (20 mL), TFA (20 eq., 2.2 mL) was added. The resulting mixture was stirred for 1 hour at 25 °C, then MeOH was added (10 mL), solvents were evaporated under reduced pressure to give **9** (TFA salt, 420 mg, 98%) as a yellowish oil.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.22-7.51 (m, 5H), 4.30-4.54 (m, 2H), 4.22 (dd, *J* = 4.77, 8.78 Hz, 1H), 2.23-2.52 (m, 3H), 1.87-2.15 (m, 1H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 181.6, 174.9, 139.8, 129.7, 128.7, 128.4, 58.3, 44.2, 30.6, 26.9; ESI-MS: $C_{12}H_{17}N_3NaO_2^+$ $[M+Na]^+$, 258.1; found 258.2.

3.10 (S)-2-Amino-*N*-benzyl-3-(1H-indol-3-yl)propanamide (10):



10a: (S)-*tert*-Butyl (1-(benzylamino)-3-(1H-indol-3-yl)-1-oxopropan-2-yl)carbamate.

To the solution of Boc-Trp-OH (1 eq., 1020 mg, 3.35 mmol) in DMF (40 mL), DIPEA (2.5 eq., 1.38 mL, 8.38 mmol) and HBTU (1 eq., 1271 mg, 3.35 mmol) were subsequently added. The resulting mixture was stirred for another 5 min, then benzylamine (1.2 eq., 0.44 mL, 4.02 mmol) was added. After the mixture was stirred overnight at 25 °C, 1M HCl (200 mL) was added and the resulting mixture was extracted with EtOAc (2x40 mL). Combined organic phases were washed with saturated $NaHCO_3$ (1x40 mL), dried over $MgSO_4$ and evaporated under reduced pressure to yield to the crude product, which was then purified by flash chromatography (DCM:MeOH, 20 min gradient from 0 to 10% of MeOH) to give **10a** (475 mg, 35%) as a white solid.

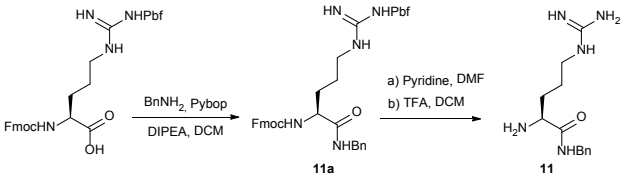
^{1}H NMR (400 MHz, DMSO-d_6,) δ 10.80 (s, 1H), 8.40 (br. s., 1H), 7.64 (d, *J* = 7.87 Hz, 1H), 7.33 (d, *J* = 7.87, 1H), 7.15 – 7.30 (m, 5H), 7.10 (s, 1H), 7.05 (t, *J* = 7.10 Hz, 1H), 6.96 (t, *J* = 7.10 Hz, 1H), 6.79 (d, *J* = 8.01 Hz, 1H), 3.11 (dd, *J* = 8.89, 14.10 Hz, 1H), 2.95 (dd, *J* = 8.89, 14.10 Hz, 1H), 1.31 (s, 9H); ESI-MS: $C_{23}H_{28}N_3O_3^+$ $[M+H]^+$, 393.2; found 393.1.

10: (S)-2-Amino-*N*-benzyl-3-(1H-indol-3-yl)propanamide.

To the solution of **10a** (1 eq., 85 mg, 2.16 mmol) in DCM (20 mL), TFA (10 eq., 160 μL, 2.16 mmol) was added. After 2 hours of stirring at 25 °C, the solution was quenched with MeOH (2 mL) and evaporated under reduced pressure to yield to the crude product, which was purified by preparative HPLC to give **10** (TFA salt, 70 mg, 79%) as a white solid.

^{1}H NMR (400 MHz, METHANOL-d_4,) δ 8.41 (br. s., 1H), 7.67 (d, *J* = 7.83 Hz, 1H), 7.44 (d, *J* = 7.82 Hz, 1H), 7.00 – 7.32 (m, 8H), 4.38 – 4.24 (m, 2H), 4.18 (t, *J* = 7.33 Hz, 1H), 3.41 (dd, *J* = 7.82, 14.34 Hz, 1H), 3.31 (dd, *J* = 7.82, 14.34 Hz, 1H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 170.1, 138.9, 138.3, 129.6, 128.8, 128.5, 125.7, 123.0, 120.4, 119.3, 112.8, 112.7, 108.2, 55.4, 44.6, 28.9; ; ESI-MS: $C_{18}H_{19}N_3NaO^+$ $[M+Na]^+$, 316.1; found 316.2

3.11 (S)-2-Amino-*N*-benzyl-5-guanidinopentanamide (11):



11a: (S)-(*9H*-Fluoren-9-yl)methyl (1-(benzylamino)-1-oxo-5-(3-((2,2,4,6,7-pentamethyl-2,3-dihydrobenzofuran-5-yl)sulfonyl)guanidino)pentan-2-yl)carbamate.

To a cooled to 0 ˚C solution of Fmoc-Arg(Pbf)-OH (1 eq., 500 mg, 0.771 mmol) in DCM (30 mL), Pybop (1 eq., 401 mg, 0.771 mmol) was added and the resulting mixture was stirred for 5 min at 0 ˚C. DIPEA (2 eq., 255 µL, 1.54 mmol) and benzylamine (5 eq., 412 mg, 0.421 mL, 3.85 mmol) were subsequently added. The resulting mixture was stirred for 1 hour at 25 ˚C before it was washed with 1M HCl (2x15 mL), with saturated $NaHCO_3$ (2x15 mL) and with water (1x30 mL). The combined organic fractions were dried over $MgSO_4$ and evaporated under reduced pressure to give **11a** that was used without further purification.

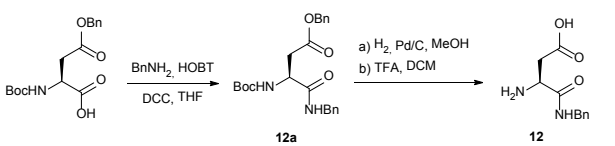
^{1}H NMR (400 MHz, DMSO-d6) δ 7.00 – 7.50 (m, 13H), 6.44 (m, 3H), 4.40 (s, 2H), 3.85 (dd, *J* = 6.45, 13.10 Hz, 1H), 3.05 – 3.60 (m, 2H), 2.95 (s, 2H), 2.10 (s, 3H), 2.49 (s, 3H), 2.58 (s, 3H), 1.50 – 1.88 (m, 2H), 1.07 – 1.30 (m, 2H), 1.15 (s, 3H), 1.11 (s, 3H); ESI-MS: $C_{41}H_{48}N_5O_6S^+$ $[M+H]^+$, 738.3; found 738.4.

11: (S)-2-Amino-*N*-benzyl-5-guanidinopentanamide.

To the solution of **11a** (200 mg, 0.426 mmol) in DMF (2 mL), piperidine (20 eq., 840 µL, 8.53 mmol) was added, the resulting mixture was stirred for 2 hours, solvents were evaporated and reaction crude was resolubilised in DCM. To the obtained solution, TFA (20 eq., 640 µL, 8.53 mmol) was added and mixture was stirred for 1 hour at 25 ˚C. Methanol (5 mL) was added and solvents were evaporated under reduced pressure. The crude product was purified by HPLC to yield **11** (85 mg, 40.6%) as a white solid.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.13-7.42 (m, 5H), 4.30-4.48 (m, 2H), 3.89 (t, *J* = 6.53 Hz, 1H), 3.19 (t, *J* = 7.03 Hz, 2H), 1.77-2.00 (m, 2H), 1.50-1.73 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 169.7, 158.8, 139.3, 129.7, 128.9, 128.6, 54.1, 44.4, 41.7, 29.8, 25.5; ESI-MS: $C_{13}H_{22}N_5O^+$ $[M+H]^+$, 264.1; found 264.2.

3.12 (S)-3-Amino-4-(benzylamino)-4-oxobutanoic acid (12):



12a: (S)-Benzyl 4-(benzylamino)-3-((*tert*-butoxycarbonyl)-amino)-4-oxobutanoate.

A solution of Boc-Asp(OBn)-OH (1 eq., 1275 mg, 3.94 mmol), HOBt (1 eq., 532 mg, 3.94 mmol) and DCC (1 eq., 813 mg, 3.94 mmol) in THF (25 mL) was stirred at 25 ˚C for 15 min. Benzylamine (1.1 eq., 464 mg, 0.474 mL, 4.34 mmol) was then added to the resulting suspension and stirring was maintained for 5 hours. The mixture was filtered and washed with EtOAc (2x10 mL). The solution was concentrated under reduced pressure and DCM (50 mL) was added to the residue. The insoluble material was filtered off and washed with DCM (2x10 mL). The filtrate was concentrated under reduced pressure and the residue was purified by column chromatography (eluent: DCM, then DCM/MeOH, 99:1) to give **12a** (1460 mg, 3.55 mmol, 90 %) as a white solid.

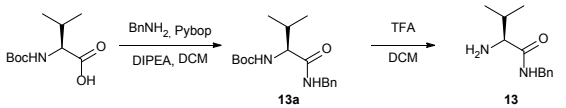
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.19-7.44 (m, 10H), 5.14 (s, 2H), 4.53 (t, *J* = 6.53 Hz, 1H), 4.38 (s, 2H), 2.84-2.98 (m, 1H), 2.70-2.82 (m, 1H), 1.44 (s, 9H); ESI-MS: $C_{23}H_{29}N_2O_5^+$ $[M+H]^+$, 412.2; found 412.2.

12: (S)-3-Amino-4-(benzylamino)-4-oxobutanoic acid.

To a degased solution of **12a** (1 eq., 400 mg, 0.97 mmol) in MeOH (10 mL), Pd/C (40 mg) was added. The resulting mixture was stirred overnight under hydrogen at 25 ˚C. Solution was filtered through Celite 545, evaporated under reduced pressure to give the crude product, which was resolubilised in DCM (10 mL). TFA (20 eq., 1440 µL, 19.4 mmol) was added to the suspension. The obtained solution was stirred for 3 hours at 25 ˚C and then evaporated under reduced pressure to give the crude product which was purified by flash chromatography (reverse phase, H_2O (0.5% TFA):MeOH, gradient from 0% to 100% of MeOH in 20 min) to give **12** (TFA salt, 230 mg, 0.69 mmol, 71 %) as a white solid.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.23-7.48 (m, 5H), 4.34-4.61 (m, 2H), 4.22 (dd, *J* = 4.77, 8.03 Hz, 1H), 2.83-3.08 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 172.9, 169.1, 139.3, 129.7, 128.7, 128.5, 51.2, 44.5, 36.1; ESI-MS: $C_{11}H_{13}N_2O_3^-$ $[M-H]^-$, 221.0; found 221.2.

3.13 (S)-2-Amino-*N*-benzyl-3-methylbutanamide (13):



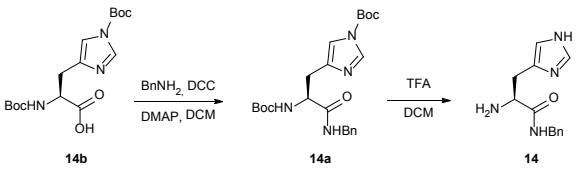
13a: (S)-*tert*-Butyl (1-(benzylamino)-3-methyl-1-oxobutan-2-yl)carbamate.

This compound was synthesised following the same procedure as for the synthesis of **1a**. White solid; yield: 78%.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.14-7.31 (m, 5H), 6.22 (br. s., 1H), 4.97 (br. s., 1H), 4.31-4.46 (m, 2H), 3.84 (dd, *J* = 6.27, 8.78 Hz, 1H), 2.11 (d, *J* = 6.53 Hz, 1H), 1.35 (s, 9H), 0.88 (d, *J* = 11.29 Hz, 3H), 0.72-1.01 (m, 3H); ESI-MS: $C_{17}H_{27}N_2O_3^+$ $[M+H]^+$, 307.2; found 307.2.

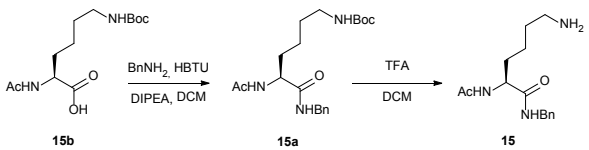
13: (S)-2-Amino-*N*-benzyl-3-methylbutanamide (TFA salt).
This compound was synthesised following the same procedure as for the synthesis of **1**. White solid; yield: 99%.
^{1}H NMR (400 MHz, METHANOL-d$_4$) δ 7.25-7.34 (m, 5H), 4.47 (d, *J* = 14.75 Hz, 1H), 4.39 (d, *J* = 14.75 Hz, 1H), 3.64 (d, *J* = 5.89 Hz, 1H), 2.17 (d hept , *J* = 5.89, 6.96 Hz, 1H), 1.03 (d, *J* = 6,96 Hz, 3H), 1.02 (d, *J* = 6,96 Hz, 3H); ^{13}C NMR (101 MHz, METHANOL-d$_4$) δ 169.3, 139.4, 129.7, 129.0, 128.6, 59.9, 44.4, 31.6, 18.8, 18.0; ESI-MS: $C_{12}H_{19}N_2O^+$ [M+H]$^+$, 207.1; found 207.1.

3.14 (S)-2-Amino-*N*-benzyl-3-(imidazol-4-yl)propanamide (14):



14b: (S)-3-(1-(*tert*-butoxycarbonyl)-*1H*-imidazol-4-yl)-2-((*tert*-butoxycarbonyl)amino)propanoic acid.
To a suspension of L-histidine (1 eq., 1g, 6.45 mmol) and TEA (3 eq., 2.7 mL, 19.35 mmol) in water (4 mL), a solution of Boc_2O (2.5 eq., 3.5 g, 16.11 mmol) in dioxane (4 mL) was slowly added. After the mixture has been stirred overnight at 25 ˚C. Water (6 mL) was added and the aqueous phase was washed with diethyl ether (3x50 mL) and ethyl acetate (3x50 mL). The total volume was increased to 200 mL with ethyl acetate and the solution acidified with 10% HCl until pH 3. The aqueous phase was extracted with ethyl acetate (100 mL). The organic phases were washed with brine and dried. The solvent was evaporated under replaced pressure, **8b** was obtained as a white solid (1.8 g, 79%) and was used without further purification.
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 13.86 (br. s., 1H), 8.14 (s, 1H), 7.14 (s, 1H), 5.29-5.53 (m, 1H), 4.44 (br. s., 1H), 3.18 (m, 2H), 1.53 (s, 9H), 1.35 (m, 9H); ESI-MS: $C_{16}H_{26}N_3O_6^+$ [M+H]$^+$, 355.2; found 355.1.
14a: (S)-3-(1-(*tert*-Butoxycarbonyl)-*1H*-imidazol-4-yl)-2-((*tert*-butoxycarbonyl)amino)propanoic acid.
To a solution of **8b** (1 eq., 975 mg, 2.75 mmol) in DCM, DMAP (0.1 eq., 34 mg, 0.275 mmol) and DCC (1.1 eq., 623 mg, 3.02 mmol) were added. The resulting mixture was stirred for 10 min at 25 ˚C, then benzylamine (1.2 eq., 360 µL, 3.30 mmol) was added. After the mixture has been stirred overnight at 25 ˚C, water (50 µL) was added to hydrolyze the excess of DCC. Stirring continued for another 30 min, The resulting mixture was then filtered, filtrate was dried over $MgSO_4$ and evaporated under reduced pressure to give the crude product, which was purified by flash chromatography (DCM:MeOH, gradient from 100% of DCM to 90% of DCM in 20 min) to give **8a** (623 mg, 51%) as a colorless oil.
^{1}H NMR (400 MHz, CHLOROFORM -d) δ 7.92 (s, 1H), 7.08-7.31 (m, 5H), 7.02 (d, *J* = 6.53 Hz, 2H), 6.08-6.21 (m, 1H), 4.32-4.52 (m, 2H), 4.17-4.32 (m, 1H), 3.07 (dd, *J* = 4.39, 14.81 Hz, 1H), 2.90 (dd, *J* = 5.52, 14.81 Hz, 1H), 1.54 (s, 9H), 1.35 (s, 9H); ESI-MS: $C_{23}H_{33}N_4O_5^+$ [M+H]$^+$, 445.2; found 445.1.
14: (S)-2-Amino-*N*-benzyl-3-(imidazol-4-yl)propanamide.
To a solution of **8a** (1 eq., 500 mg, 1.12 mmol) in DCM (20 mL), TFA (20 eq., 1.7 mL) was added. The resulting mixture was stirred for 1 hour at 25 ˚C, then MeOH was added (10 mL), and solvents were evaporated under reduced pressure to give **8** (di-TFA salt, 420 mg, 98%) as a yellowish oil.
^{1}H NMR (400 MHz, METHANOL-d$_4$) δ 8.77 (s, 1H), 7.19-7.46 (m, 6H), 4.45 (d, *J* = 14.56 Hz, 1H), 4.34 (d, *J* = 14.56 Hz, 1H), 4.19 (t, *J* = 7.15 Hz, 1H), 3.12-3.29 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d$_4$) δ 168.2, 139.2, 135.8, 129.7, 129.0, 128.7, 119.4, 101.4, 53.5, 44.5, 27.8; ESI-MS: $C_{13}H_{17}N_4O^+$ [M+H]$^+$, 245.1; found 245.2.

3.15 (S)-2-Acetamido-6-amino-*N*-benzylhexanamide (15):



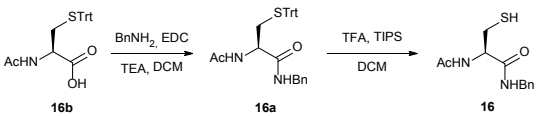
15b: N^2-acetyl-N^6-(*tert*-butoxycarbonyl)-L-lysine.
N^6-(*tert*-butoxycarbonyl)-L-lysine (1 eq., 1230 mg, 5 mmol) was dissolved in 1N NaOH solution (1 eq., 5 mL, 5 mmol) and pH of the obtained solution was adjusted to 10.7 with 4N NaOH if needed. Acetic anhydride (3 eq., 1.4 mL, 15 mmol) and 4N NaOH (6.5 mL, 26 mmol) were added at the same time to maintain the pH at 10.7. The resulting mixture was stirred for another 1 hour after the addition of the reagents, cooled to -5 ˚C, carefully acidified with 12N HCl (2.5 mL, 30 mmol) and extracted with ethyl acetate (3x25 mL). The combined organic layer was washed with brine, dried over Na_2SO_4, evaporated and leaved for 2 hours to crystallise. The obtained white crystals were washed with pentane and dried under reduced pressure.
^{1}H NMR (400 MHz, METHANOL-d$_4$) δ 4.36 (dd, *J* = 5.02, 8.92 Hz, 1H), 3.03 (t, *J* = 6.83 Hz, 2H), 1.99 (s, 3H), 1.79-1.91 (m, 1H), 1.63-1.75 (m, 1H), 1.35-1.57 (m, 4H), 1.43 (s, 9H); ESI-MS: $C_{13}H_{25}N_2O_5^+$ [M+H]$^+$, 289.2; found 289.1.
15a: *tert*-Butyl (S)-(5-acetamido-6-(benzylamino)-6-oxohexyl)carbamate.
To a solution of **15b** (1 eq., 150 mg, 0.52 mmol) in DMF (2.01 mL), DIPEA (1.5 eq., 129 µL, 0.78 mmol) and HBTU (1 eq., 197 mg, 0.52 mmol) were added. Benzylamine (1 eq., 55.7 mg, 54 µL, 0.52 mmol) was added to the resulting mixture, which was stirred for another 2 hours at 25 ˚C. 1N HCl (20 mL) was added to the reaction mixture, the obtained crude product was extracted with ethyl acetate (3x10 mL). United organic fractions were evaporated and the residue was purified by flash chromatography to yield **15a** (166 mg, 0.442 mmol, 85 %) as a colourless oil.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.21-7.36 (m, 5H), 4.39 (s, 2H), 4.29-4.36 (m, 1H), 3.03 (t, *J* = 6.72 Hz, 2H), 2.01 (s, 3H), 1.74-1.88 (m, 1H), 1.58-1.73 (m, 1H), 1.35-1.54 (m, 4H), 1.45 (s, 9H); ESI-MS: $C_{20}H_{32}N_3O_4^+$ $[M+H]^+$, 378.2; found 378.2.

15: (S)-2-Acetamido-6-amino-*N*-benzylhexanamide.

To a solution of **15a** (100 mg, 0.285 mmol) in DCM (2 mL), TFA (50 µL) was added. The resulting mixture was stirred for 30 minutes, evaporated under reduced pressure to give **15** (TFA salt, 102 mg, 96%) which was used without further purification. ^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.10-7.23 (m, 5H), 7.20-7.29 (m, 2H), 2.79 (t, *J* = 7.51, 2H), 1.90 (s, 3H), 1.69-1.80 (m, 1H), 1.48-1.68 (m, 3H), 1.22-1.42 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ174.3, 173.7, 140.0, 129.6, 128.6, 128.3, 54.9, 44.2, 40.7, 32.6, 28.3, 24.0, 22.6; ESI-MS: $C_{15}H_{24}N_3O_2^+$ $[M+H]^+$, 278.2; found 278.2.

3.16 (R)-2-Acetamido-*N*-benzyl-3-mercaptopropanamide (16):



16b 16a 16

16b: (R)-2-Acetamido-3-(tritylthio)propanoic acid.

N-acetyl-L-cysteine (1 eq., 4.6 g, 28.2 mmol) was dissolved in acetic acid (25 mL), followed by the addition triphenylmethanol (1.02 eq., 7.5 g, 28.8 mmol) and boron trifluoride etherate (1.5 eq., 5.36 mL, 42.3 mmol) and the solution was stirred for 3 hours at 25 °C. The solution was treated with water (50 mL) and saturated sodium acetate solution (25 mL), chilled on ice, and filtered to yield white slurry. The slurry was extracted with diethyl ether, concentrated under reduced pressure, and dried overnight to yield **16b** (11.19 g, 27.6 mmol, 98 %) as a white solid.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.40 (d, *J* = 7.78 Hz, 7H), 7.31 (t, *J* = 7.53 Hz, 7H), 7.19-7.27 (m, 4H), 4.31 (dd, *J* = 5.02, 8.03 Hz, 1H), 2.54-2.71 (m, 2H), 1.98 (s, 3H); ESI-MS: $C_{24}H_{22}NO_3S^-$ $[M-H]^-$, 404.1; found 404.1.

16a: (R)-2-Acetamido-*N*-benzyl-3-(tritylthio)propanamide.

To a solution of **16b** (1 eq., 450 mg, 1.11 mmol) in DCM (15 mL), TEA (3 eq., 0.463 mL, 3.33 mmol) and benzylamine (1.1 eq., 0.133 mL, 1.22 mmol) were added and the resulting mixture was cooled to 0 ˚C. EDC (1.1 eq., 234 mg, 1.22 mmol) was added and the solution was stirred overnight at 25 ˚C. DCM was washed with 1M HCl (2x10 mL), saturated $NaHCO_3$ (1x10 mL), dried over Na_2SO_4 and purified by flash chromatography to give **16a** (395 mg, 0.799 mmol, 72 %).

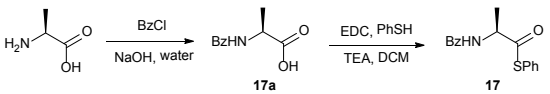
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.39 (d, *J* = 7.53 Hz, 6H), 7.15-7.34 (m, 15H), 4.42 (d, *J* = 15.06 Hz, 1H), 4.23-4.36 (m, 2H), 2.43-2.64 (m, 2H), 1.95 (s, 3H); ESI-MS: $C_{31}H_{31}N_2O_2S^+$ $[M+H]^+$, 495.2; found 495.1.

16: (R)-2-acetamido-*N*-benzyl-3-mercaptopropanamide.

To the solution of **15a** (1 eq., 236 mg, 0.479 mmol) in DCM (4.6 mL), TFA (10 eq., 0.356 mL, 4.79 mmol) was added. Triisopropylsilane (TIPS, 2.5 eq., 0.245 mL, 1.2 mmol) was added to the obtained yellow-colored solution and the reaction mass was stirred until complete disappearance of the yellow color (30-60 min). Solvents were evaporated and the obtained crude was purified by flash chromatography to give **16** (95.4 mg, 0.378 mmol, 79 %).

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.29-7.38 (m, 4H), 7.25 (qd, *J* = 4.2, 8.8 Hz, 1H), 4.51 (t, *J* = 6.4 Hz, 1H), 4.42 (s, 2H), 2.74-2.98 (m, 2H), 2.04 (s, 3H); ^{13}C NMR (101 MHz, METHANOL-d4) δ173.5, 172.4, 139.7, 129.5, 128.5, 128.2, 57.3, 44.2, 26.9, 22.5; ESI-MS: $C_{12}H_{16}N_2O_2SNa^+$ $[M+Na]^+$, 275.1; found 275.1.

3.17 *S*-Phenyl (S)-2-benzamidopropanethioate (17):



17a 17

17a: Benzoyl-L-alanine.

L-Alanine (1 eq., 5.00 g, 56 mmol) was dissolved in water (20 mL) containing NaOH (1.1 eq., 2.25 g, 62 mmol) and the solution was cooled to 0 ˚C. Benzoyl chloride (1.1 eq., 7.18 mL, 62 mmol) and a solution of NaOH (1.1 eq., 2.48 g, 62 mmol) in water (10 mL) were added to the stirred solution of the amino acid in portions during 2 hours. Obtained slurry mixture was stirred for another 2 hours at 25 ˚C, boiled for 20 minutes, then cooled to 0 ˚C and acidified carefully to pH 2-3 with concentrated HCl. The product was filtered, washed with ice-cold water (25 mL) and then recrystallised from 150 mL of boiling water to give **17a** as a white crystalline product.

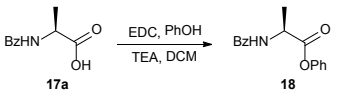
^{1}H NMR (400 MHz, DMSO-d_6) δ 12.53 (s, 1H), 8.65 (d, *J* = 7.28 Hz, 1H), 7.89 (d, *J* = 7.68 Hz, 2H), 7.47-7.54 (m, 3H), 4.42 (quint, *J* = 7.35 Hz, 1H), 1.40 (d, *J* = 7.35 Hz, 3H); ESI-MS: $C_{10}H_{10}NO_3^-$ $[M-H]^-$, 192.1; found 192.0.

17: *S*-Phenyl (S)-2-benzamidopropanethioate.

To a solution of **17a** (1 eq., 295 mg, 1.53 mmol) and EDC·HCl (1 eq., 271 mg, 1.53 mmol) in DCM (5 mL), TEA (1 eq., 213 µL, 1.53 mmol) thiophenol (1.1 eq., 172 µL, 1.68 mmol) were subsequently added. The obtained solution was stirred overnight at 25 ˚C, washed with saturated $NaHCO_3$, dried with Na_2SO_4, and evaporated. The crude product was purified by flash chromatography (20 minutes gradient EtOAc/Cyclohexane) to give **17** (288 mg, 66%) as a yellowish solid.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.78-7.84 (m, 2H), 7.38-7.55 (m, 8H), 6.61 (d, *J* = 8.54 Hz, 1H), 5.07 (quint, *J* = 7.40 Hz, 1H), 1.58 (d, *J* = 7.22 Hz, 3H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ196.1, 167.3, 135.1, 134.2, 132.0, 129.2, 128.8, 127.5, 127.1, 66.2, 18.5 ESI-MS: $C_{16}H_{15}NO_2SNa^+$ $[M+Na]^+$, 308.0; found 308.2.

3.18 Phenyl benzoyl-L-alaninate (18):

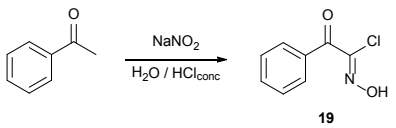


17a 18

Phenol (1.5 eq., 225 mg, 2.40 mmol) was added to a solution of **17a** (1 eq., 285 mg, 1.60 mmol) in DCM (5 mL). EDC·HCl (1 eq., 285 mg, 1.60 mmol) was added, and the reaction mixture was stirred overnight. After completion, the reaction mixture was concentrated, diluted with DCM, washed with 1M HCl, $NaHCO_3$, and dried with anhydrous Na_2SO_4, The crude products was purified *via* flash chromatography (EtOAc/Cyclohexane). The product **18** (300 mg, 70%) was obtained as a white solid.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.70-7.80 (m, 2H), 7.14-7.54 (m, 8H), 6.76 (s, 1H), 5.05 (m, 1H), 3.94-3.95 (d, J = 7.2 Hz, 3H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ172.0, 167.2, 150.5, 133.8, 131.9, 129.6, 128.7, 127.1, 126.3, 121.3, 48.8, 18.5;ESI-MS: $C_{16}H_{16}NO_3^+$ $[M+H]^+$, 270.1; found 270.1.

3.19 *N*-HYDROXY-2-OXO-2-PHENYLACETIMIDOYL (19):

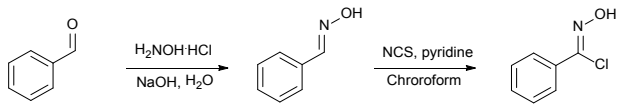


19

This compound was synthesised using slightly modified protocol of Hamersak *et al*.[806]

To a stirring solution of acetophenone (22 eq., 18.7 mL, 160 mmol) in a mixture of water (40 mL) and concentrated HCl (30 mL), sodium nitrite (1 eq., 500 mg, 7.2 mmol) was added at 25 ˚C, followed by dropwise addition of concentrated nitric acid (12 mL) over 1 hour at 70 ˚C. Stirring continued at 70 ˚C for additional 45 minutes. A sticky solid was separated from the mixture by extraction with DCM (3x100 mL). The combined organic layers were dried over Na_2SO_4 and concentrated to approximately 100 mL. Upon addition of hexane (ca. 150 mL), the precipitated product was collected by filtration. On crystallisation from DCM/cyclohexane, pure **19** was obtained (21 g, 55%). Analyses were in accordance with literature.[806]

3.20 *(Z)-N*-HYDROXYBENZIMIDOYL (20):



20a 20

20a: *(E)*-Benzaldehyde oxime.

In a 250 mL conical flask a solution of NaOH (1.77 eq., 14 g, 350 mmol) in water (40 mL) and benzaldehyde (1 eq., 20 mL, 197 mmol) were mixed. Hydroxylamine hydrochloride (1.1 eq., 15 g, 215 mmol) was added in small portions while shaking the mixture continually. Some heat is developed and the benzaldehyde eventually disappears. Upon cooling, a crystalline mass of the sodium derivative separates out. Sufficient amount of water was added to form a clear solution, and carbon dioxide was passed into solution until saturation. A colourless emulsion of *(E)*- and *(Z)*-aldoxime were separated by extraction with diethyl ether. The organic layer was dried over $MgSO_4$, ether was evaporated under reduced pressure and the obtained crude product was distilled to give **20a** (12 g, 50%) which solidified on cooling in ice.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 8.10 (s, 1H), 7.52-7.61 (m, 2H), 7.32-7.38 (m, 3H); ESI-MS: $C_7H_8NO^+$ $[M+H]^+$, 122.1; found 122.0.

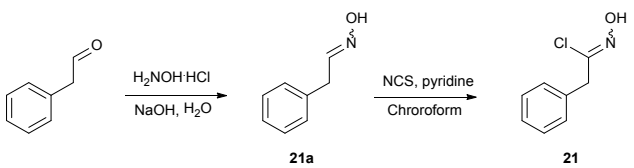
20: *(Z)*-*N*-Hydroxybenzimidoyl.

Modified procedure reported by Castellano *et al*.[840]

To a solution of **20a** (1 eq., 5 g, 41.3 mmol) in chloroform (150 mL), pyridine (300 μL, 3.7 mmol) was added. The reaction mixture was heated at 40 ˚C, and *N*-chlorosuccinimide (1.1 eq., 6.07 g, 45.5 mmol) was added portionwise. After the reaction was complete (1 hour, monitored by TLC), the mixture was diluted with dichloromethane (250 mL) and washed with brine (3x10 mL). The organic phase was dried with Na_2SO_4, filtered, and concentrated *in vacuo* to give crude product which was purified by flash chromatography (20 minutes gradient EtOAc/Cyclohexane) to give **21** (5.84 g, 91%) as a white solid.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 8.32 (s, 1H); 7.36-7.39 (m, 2H); 7.57-7.60 (m, 3H); ESI-MS: $C_7H_5ClNO^-$ $[M-H]^-$, 154.0; found 154.0.

3.21 *N*-Hydroxybenzimidoyl (21, mixture of isomers):



21a: 2-Phenylacetaldehyde oxime.

To a suspension of phenylacetaldehyde (1 eq., 13.9 mL, 141 mmol) and hydroxylamine hydrochloride (1.1 eq., 10.7 g, 155 mmol) in water:methanol mixture (1:1, 150 mL), an aqueous solution of NaOH (1.77 eq., 10 g, 250 mmol) in water (40 mL), was slowly added. The resulting mixture was stirred at 25 ˚C for 3 hours, and methanol was evaporated. The aqueous phase was extracted with diethyl ether (3x30 mL). The combined organic layers were dried over $MgSO_4$, and concentrated *in vacuo* to give a mixture of *(E)*- and *(Z)*-aldoximes **21a** (isomers **a** and **b**, 6.9 g, 37%) after the purification by flash chromatography (20 minutes gradient EtOAc/Cyclohexane).

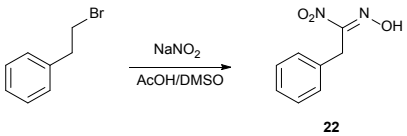
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 8.40 (s, 1H, isomer **a**), 7.88 (s, 1H, isomer **b**), 7.53 (t, J = 6.2 Hz, 1H, isomer **b**) 7.10-7.40 (m, 5H isomer **a** and 5H isomer **b**), 3.75 (d, J = 5.2 Hz, isomer **a**), 3.57 (d, J = 6.2 Hz, 2H, isomer **b**); ESI-MS: $C_8H_{10}NO^+$ $[M+H]^+$, 136.1; found 136.0.

21: *N*-Hydroxy-2-phenylacetimidoyl chloride (mixture of isomers).

The same procedure as for the synthesis of the product **20**. A yellowish solid; yield: 80%.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 8.15 (s, 1H); 7.45-7.13 (m, 5H); 3.81 (m, 2H); ESI-MS: $C_8H_7ClNO^-$ $[M-H]^-$, 168.0; found 168.0.

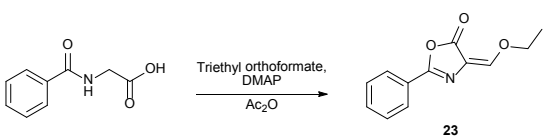
3.22 Nitro(phenyl)methanone oxime (22):



(2-Bromoethyl)benzene (1 eq., 300 mg, 1.62 mmol) and sodium nitrite (3 eq., 336 mg, 4.86 mmol) were stirred in DMSO (2.5 mL) containing glacial acetic acid (1 mL). Water (10 mL) was added to give a slurry emulsion which was extracted with EtOAc (3x3 mL). Organic phases were combined, dried over Na_2SO_4, solvents were evaporated under reduced pressure to give crude product, which was purified by flash chromatography (20 minutes gradient EtOAc/Cyclohexane) to give **22** (53 mg, 18%) as a pale yellow solid.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 9.50-10.05 (br. s, 1H); 7.10-7.19 (m, 5H); 4.19 (s, 2H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 160.8, 136.0, 132.2, 128.5, 127.3, 29.8; IR (neat film, cm^{-1}): 3293.6, 2890.3, 1541.9, 1496.1, 1455.6, 1421.3, 1343.0, 1075.8, 1030.7, 855.6, 736.7, 709.2; ESI-MS: $C_8H_7N_2O_3^-$ $[M-H]^-$, 179.0; found 179.0.

3.23 4-(Ethoxymethylene)-2-phenyloxazol-5(*4H*)-one (23):

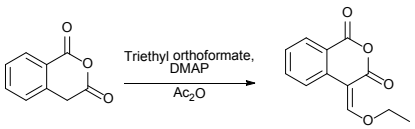


The product was synthesised *via* modified procedure described by Matos *et al.* [841]

To a stirred mixture of hippuric acid (1 eq., 1.00 g, 5.6 mmol), ethyl orthoformate (1 eq., 0.93 mL, 5.6 mmol), 4-(dimethylamino)pyridine (10 mg) at 25 ˚C and under argon atmosphere was added acetic anhydride (2 eq., 1.05 mL, 11.2 mmol) and then refluxed for 60 minutes. The reaction mixture was concentrated under vacuo and the obtained brown oil was purified by flash chromatography (20 minutes gradient EtOAc/Cyclohexane) to give **23** (0.90 g, 74%).

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 8.11-8.25 (m, 2H), 7.51-7.58 (m, 3H), 7.36 (s, 1H), 4.52 (quint, 2H), 1.56 (t, 3H); ESI-MS: $C_{12}H_{11}NO_3Na^+$ [M+Na]$^+$, 240.1; found 240.1.

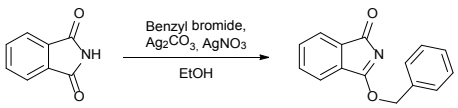
3.24 4-(Ethoxymethylene)isochromane-1,3-dione (24):



24

The product was synthesised using the protocol reported by Matos *et al.*[808] Analyses were in accordance with literature.

3.25 3-(Benzyloxy)-*1H*-isoindol-1-one (25):

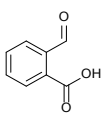


25

Isoindoline-1,3-dione (1 eq., 2.1 g, 14.28 mmol), silver carbonate (1.75 eq., 6.7 g, 25 mmol) and silver nitrate (0.02 eq., 50 mg, 0.29 mmol) were dissolved in absolute ethanol. The reaction mixture was heated to reflux for 48 hours. The solvent was removed *in vacuo*, and the intermediate silver salt was dried under vacuum at 85 ˚C for 12 hours. To the flask containing the dried salt were added reagent grade chloroform and benzyl bromide (2 eq., 4.9 g, 28.6 mmol). The mixture was refluxed for 72 hours. The silver salts were removed by filtration, and the filtrate was concentrated by rotary evaporation. The crude product was purified by flash chromatography (20 minutes gradient EtOAc/Cyclohexane) to give **25** (305 mg, 9%) as a grey solid.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.81-7.89 (m, 2H), 7.62-7.72 (m, 2H), 7.38-7.44 (m, 2H), 7.19-7.38 (m, 4H), 4.87 (s, 2H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 192.6, 170.0, 134.6, 134.3, 133.4, 132.7, 130.1, 129.8, 129.7 129.0, 128.4, 123.6, 65.3; ESI-MS: $C_{15}H_{12}NO_3^+$ [M+H]$^+$, 239.1; found 238.9.

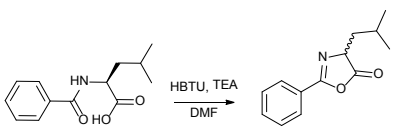
3.26 2-Formylbenzoic acid (26):



26

The product is commercially available.

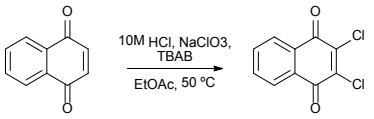
3.27 4-Isobutyl-2-phenyloxazol-5-one (27):



27

The product was a kind gift from Frederic Taran (CEA, Saclay).

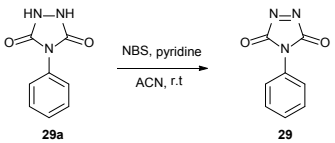
3.28 2,3-Dichloronaphthalene-1,4-dione (28):



28

This compound is synthesised using the protocol reported by Cheng *et al.*[812] Analyses were in accordance with literature.

3.29 4-Phenyl-3*H*-1,2,4-triazole-3,5(4*H*)-dione (29):

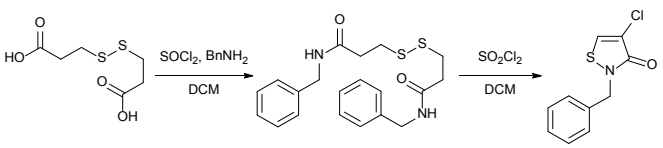


29a 29

To a solution of **29a** (1.05 eq., 10 mg, 0.056 mmol) and pyridine (1 eq., 4.4 µL, 0.054 mmol) in ACN (3 mL) was added *N*-bromosuccinimide (1 eq., 9.6 mg, 0.056 mmol) at 25 ˚C. The resulting solution was stirred at 25 ˚C for 5 min (colour changed from colourless to red). This solution was used in the screening without isolation.

Used similarly as described by Barbas *et al.*[522]

3.30 2-Benzyl-5-chloroisothiazol-3(*2H*)-one (30):



30b 30a 30

30a: 3,3'-Disulfanediylbis(*N*-benzylpropanamide).

A mixture of thionyl chloride (4 eq., 1.5 mL, 20.54 mmol) and **30b** (1 eq., 1.08 g, 5.14 mmol) was heated to reflux until total solubilising of the departure product. Excess of thionyl chloride was evaporated under reduced pressure to give the corresponding chloroanhydride. This was dissolved in DCM (3 mL), and the obtained solution was added dropwise to a cooled to 0 ˚C solution of benzylamine (6 eq., 3.37 mL, 30.8 mmol) in DCM (20 mL). The obtained precipitate was filtered and washed with 1N HCl and water, dried and used in next stages without any further purification (1.99 g, 100%).

^{1}H NMR (400 MHz, DMSO-d6) δ 8.47 (t, J = 6.3 Hz, 2H), 7.19-7.34 (m, 10H), 4.28 (d, J = 6.3 Hz, 4H), 2.94 (t, J = 7.10 Hz, 4H), 2.56 (t, J = 7.10 Hz, 4H); ESI-MS: $C_{20}H_{25}N_2O_2S_2^+$ [M+H]$^+$, 388.1; found 388.2.

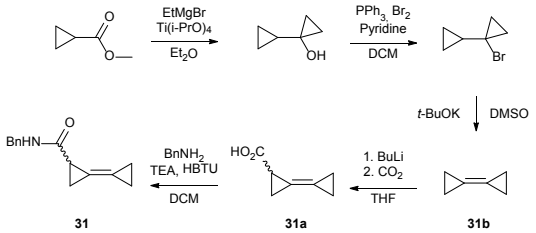
30: 2-Benzyl-5-chloroisothiazol-3(*2H*)-one.

To a suspension of **30a** (1 eq. 1.99 g, 5.14 mmol) in DCM (10 mL), sulfuryl chloride (5 eq., 2.10 mL, 25.7 mmol) was added dropwise. The resulting mixture was stirred overnight still at 25 ˚C, filtrated and the obtained filtrate was washed

twice with water, dried over $MgSO_4$, dried and evaporated. The crude product was purified by flash chromatography (20 minutes gradient EtOAc/Cyclohexane) to give **30** (288 mg, 25%).

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.20-7.34 (m, 5H), 6.22 (s, 1H), 4.81 (s, 2H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 166.9, 146.6, 135.6, 129.0, 128.7, 128.5, 114.5, 47.4; ESI-MS: $C_{10}H_9ClNOS^+$ $[M+H]^+$, 226.0; found 226.0.

3.31 *N*-Benzyl-[1,1'-bi(cyclopropylidene)]-2-carboxamide (31):



31b: 1,1'-Bi(cyclopropylidene).

The compound was synthesised using the protocol described by Meijere *et al.*[842]

31a: [1,1'-Bi(cyclopropylidene)]-2-carboxylic acid.

2.5 M solution of butyllithium in hexane (262 µL, 655.2 µmol) was added into anhydrous THF (1 mL) at -10°C. The solution of **31b** (50 mg, 624 µmol) in anhydrous THF (1 mL) was added to the resulting mixure and was allowed to heat up to 0°C. After 1 h of stirring dry CO_2 was bubbled through the mixture and it was allowed to reach room temperature. After 30 min the reaction was quenched by the addition of 2N NaOH (1 mL) and the mixture was washed with Et_2O (5 mL). Subsequently the aqueous phase was acidified with conc. HCl to reach pH 3, extracted with Et_2O (3x5 mL) and the combined organic phases were dried over Na_2SO_4 and evaporated under reduced pressure to yield the acid **31a** (72.4 mg, 93%) as a white solid.

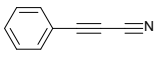
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 2.22-2.36 (m, 1H), 1.86-1.96 (m, 1H), 1.70-1.80 (m, 1H), 1.10-1.24 (m, 4H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 179.5, 113.2, 110.6, 18.1, 12.7, 3.8, 3.5; ESI-MS: $C_7H_7O_2^-$ $[M-H]^-$, 123.1; found 123.0.

31: *N*-Benzyl-[1,1'-bi(cyclopropylidene)]-2-carboxamide.

31a (72.4 mg, 583.2 mmol) was dissolved in anhydrous DCM under argon atmosphere and subsequently TEA (89.2 µL, 64.9 mg, 641.5 µmol) and benzyl amine (70.2 µL, 68.7 mg, 641.5 µmol) were added. The reaction mixture was cooled to 0°C, HBTU (243.3 mg, 641.5 µmol) was added and the mixture was stirred for 10 min before it was allowed to reach RT and stirred for 14 h. The solvent was evaporated under reduced pressure and the crude product was purified by preparative HPLC (H_2O/ACN + 0.05% TFA, 9:1- 5:95, 30 min) to yield the amide **31** (47 mg, 38%) as a white solid.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.19-7.44 (m, 5H), 5.85 (br s, 1H), 4.46 (d, *J* = 5.8 Hz, 2H), 2.22-2.33 (m, 1H), 1.80-1.90 (m, 1H), 1.67-1.77 (m, 1H), 1.15-1.38 (m, 4H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 171.4, 138.4, 128.7, 127.6, 127.4, 113.1, 111.0, 43.6, 20.4, 11.3, 3.5, 3.3; ESI-MS: $C_{14}H_{16}NO^+$ $[M+H]^+$, 214.1; found 214.0.

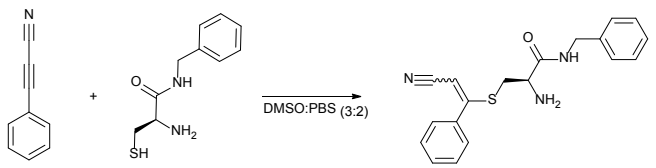
3.32 3-Phenylpropiolonitrile (32):



32

The product is commercially available.

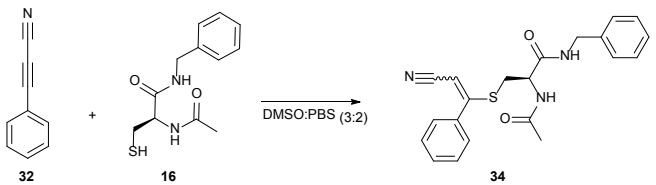
3.33 2-Amino-*N*-benzyl-3-((2-cyano-1-phenylvinyl)thio)pro-panamide (33):



32 (1 eq., 10 mg, 0.0787 mmol) and **4** (1 eq., 26 mg, 0.0802 mmol) were solubilised in (3:2) PBS:DMSO mixture (1 mL). The resulting mixture was stirred 12 hours at 25 ˚C, diluted with water (5 mL), and extracted with DCM to give **33** (26 mg, 0.0771 mmol, 98 %) as a white solid.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.35-7.50 (m, 5H), 7.13-7.24 (m, 5H), 5.79 (*Z*-form) and 5.55 (*E*-form) (s, 1H), 4.41 (d, *J* = 14.85 Hz, 1H), 4.24 (d, *J* = 14.85 Hz, 1H), 3.82 (t, *J* = 6.50 Hz, 1H), 3.06 (d, *J* = 6.50 Hz, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 167.5, 161.3, 138.9, 136.6, 132.3, 130.4, 129.7, 129.5, 129.1, 128.6, 117.4, 100.0, 53.8, 44.8, 34.5; ESI-MS: $C_{19}H_{20}N_3OS^+$ $[M+H]^+$, 338.1; found 338.1.

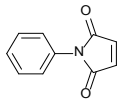
3.34 2-Amino-*N*-benzyl-3-((2-cyano-1-phenylvinyl)thio)pro-panamide (34):



The same protocol as for the synthesis of **33** was used. White solid; yield: 99%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.43-7.51 (m, 2H), 7.31-7.42 (m, 3H), 7.15-7.23 (m, 2H), 7.13 (d, *J* = 7.03 Hz, 3H), 5.59 (*E*-form) and 5.69 (*Z*-form) (s, 1H), 4.14-4.36 (m, 3H), 3.02 (dd, *J* = 8.25, 13.80 Hz, 1H), 2.80 (dd, *J* = 8.25, 13.80 Hz, 1H), 1.89 (*Z*-form) and 1.91 (*E*-form) (s, 3H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 171.1, 161.0, 160.5, 138.9, 132.2, 132.0, 130.2, 129.3, 129.1, 129.0, 128.7, 117.5, 99.8, 54.0, 44.9, 34.5, 22.9; ESI-MS: $C_{21}H_{21}N_3O_2SNa^+$ $[M+H]^+$, 402.1; found 402.1.

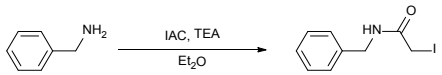
3.35 1-PHENYL-*1H*-PYRROLE-2,5-DIONE (35):



35

The product is commercially available.

3.36 *N*-BENZYL-2-IODOACETAMIDE (36):



36

A solution of iodoacetyl chloride (IAC, 1 eq., 250 µL, 2.79 mmol) in anhydrous diethyl ether (5 mL) at 0 ˚C was treated with benzylamine (2 eq., 610 µL, 5.59 mmol). The mixture was stirred for one hour at 0 ˚C, and filtered. The filtrate was washed with 1N HCl, saturated sodium bicarbonate and brine, dried over $MgSO_4$. The crude product was purified by flash chromatography (20 minutes gradient EtOAc/Cyclohexane) to give **36** (422 mg, 55%) as a white solid.

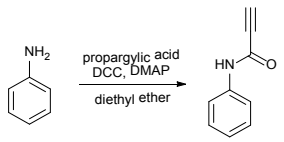
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.23-7.52 (m, 5H), 4.38 (s, 2H), 3.75 (s, 2H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 167.1, 137.7, 129.0, 127.9, 44.6, –0.4; ESI-MS: $C_9H_{11}INO^+$ $[M+H]^+$, 276.0; found 276.0.

3.37 4-VINYLPYRIDINE (37):

37

The product is commercially available.

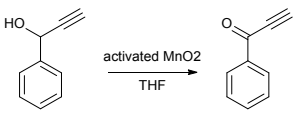
3.38 *N*-PHENYLPROPIOLAMIDE (38):



38

This compound was synthesised following the procedure described by Bio *et al.*[843] Analyses were in accordance.

3.39 1-PHENYLPROP-2-YN-1-ONE (39):

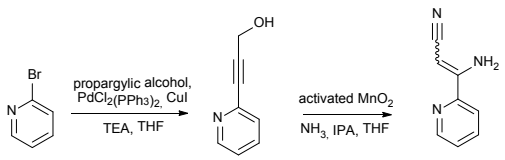


39a 39

To a solution of **39a** (1 eq., 175 µL, 1.41 mmol) in THF (4 mL), activated MnO_2 (5 eq., 612 mg, 7.04 mmol) was added. The resulting mixture was stirred for 1 hour at 25 ˚C (until no starting material was detected by TLC), diluted with DCM (25 mL) and filtered through Celite 545 ®. **39** (174 mg, 95 %) was obtained as a yellowish solid, no further purification was needed.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 8.10 (d, J = 7.25 Hz, 2H), 7.57 (t, J = 7.40 Hz, 1H), 7.44 (dd, J = 7.25, 7.40 Hz, 2H), 3.37 (s, 1H); ESI-MS: $C_9H_9O^+$ $[M+H]^+$, 133.1; found 133.0.

3.40 3-AMINO-3-(PYRIDIN-2-YL)ACRYLONITRILE (40):



40a 40

40a: 2-(3-Bromoprop-1-yn-1-yl).

Synthesised using protocol A for Sonogashira coupling (Section D 2.9, page 77). Brown solid, yield: 77%.

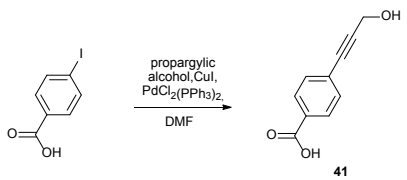
^{1}H NMR (400 MHz, METHANOL-d_4) δ 8.51 (dd, J = 1.72, 5.05 Hz, 1H), 7.82 (td, J = 1.72, 7.60 Hz, 1H), 7.53 (d, J = 7.60 Hz, 1H), 7.39 (dd, J = 7.60, 5.05, 1H), 4.44 (s, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 150.5, 143.8, 138.6, 128.7, 124.8, 90.2, 84.1, 51.0; ESI-MS: $C_8H_8NO^+$ $[M+H]^+$, 134.0; found 134.0.

40: 3-Amino-3-(pyridin-2-yl)acrylonitrile.

Synthesised using McAllister's protocol (Section D 2.10, page 77). Reaction time: 1 hour. White solid, yield: 95%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 8.64 (d, J = 4.80 Hz, 1H), 7.75-7.90 (m, 2H), 7.46 (dd, J = 4.80, 6.55 Hz, 1H), 4.76 (s, 1H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 160.2, 152.0, 150.1, 138.6, 126.4, 121.7, 121.0, 60.5; ESI-MS: $C_8H_8N_3^+$ $[M+H]^+$, 146.2; found 146.2.

3.41 4-(3-HYDROXYPROP-1-YN-1-YL)BENZOIC ACID (41):

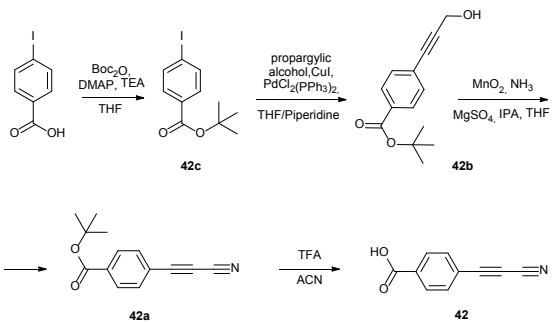


41

Synthesised using protocol A for Sonogashira coupling (Section D 2.9, page 77). White solid, yield: 45%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.94 (d, J = 8.40 Hz, 2H), 7.46 (d, J = 8.40 Hz, 2H), 4.38 (s, 2H), 3.35 (s, 1H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 169.1, 132.5, 131.6, 130.8, 129.0, 92.0, 84.7, 51.2; ESI-MS: $C_{10}H_7O_3^-$ $[M-H]^-$, 175.0; found 175.2.

3.42 3-AMINO-3-(PYRIDIN-2-YL)ACRYLONITRILE (42):



42c: *tert*-Butyl 4-iodobenzoate.

To the solution of 4-iodobenzoic acid (1 eq., 1 g, 4 mmol) in THF (12 mL) was added Boc_2O (2 eq., 1.75 g, 1.72 mL, 8 mmol) and DMAP (30 %, 0.150 g, 1.2 mmol). The mixture was stirred at 25 ˚C for 14 hours and then another portion of Boc_2O (1 eq.) was added, followed by TEA (1 eq., 560 µL, 4 mmol). The mixture was stirred for another 14 hours and then evaporated. The residue was solubilised in EtOAc and washed with 1M HCl, 1M NaOH and brine. The organic layer was dried over $MgSO_4$ and evaporated under reduced pressure to yield 42 (1.15 g, 95%) as a white solid.

^{1}H NMR (300 MHz, CHLOROFORM-d) δ 7.77 (d, J = 8.55 Hz, 2H), 7.69 (d, J = 8.55 Hz, 2H), 1.59 (s, 9H); ESI-MS: $C_{11}H_{14}IO_2^+$ [M+H]$^+$, 305.0; found 305.0.

42b: *tert*-Butyl 4-(3-hydroxyprop-1-yn-1-yl)benzoate.

A solution of **42c** (1 eq., 1.11 g, 3.66 mmol), $PdCl_2(PPh_3)_2$ (1 %, 26 mg, 0.04 mmol) , CuI (2 %, 14 mg, 0.08 mmol) , and propargylic alcohol (2 eq., 434 µL, 7.33 mmol) in THF:piperidine mixture (1:1, 3 mL) were stirred for 12 hours. The mixture was diluted with DCM (2.5 mL), washed with sat NH_4Cl (2x2 mL), and water (2.5 mL). The solvent was removed under reduced pressure and the crude product was used in the next step without purification. Estimated yield (NMR) – 98%.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.86 - 7.99 (m, J = 8.28 Hz, 2H), 7.42 - 7.51 (m, J = 8.03 Hz, 2H), 4.53 (s, 2H), 1.60 (s, 9H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 165.1, 131.7, 131.4, 129.3, 126.6, 89.8, 85.1, 81.4, 51.6, 28.1; ESI-MS: $C_{14}H_{17}O_3^+$ [M+H]$^+$, 233.1; found 233.1.

42a: *tert*-Butyl 4-(cyanoethynyl)benzoate.

Synthesised using McAllister's protocol (Section D 2.10, page 77). Reaction time: 30 minutes. White solid, yield: 47%.

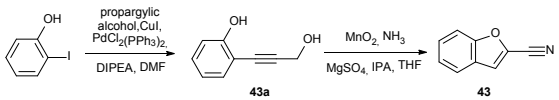
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.97 - 8.10 (m, J = 8.28 Hz, 2H), 7.61 - 7.73 (m, J = 8.28 Hz, 2H), 1.62 (s, 9H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 164.3, 134.8, 133.3, 129.7, 121.3, 105.2, 82.2, 81.9, 64.8, 28.1; ESI-MS: $C_{14}H_{14}NO_2^+$ [M+H]$^+$, 228.1; found 228.0.

42: 4-(Cyanoethynyl)benzoic acid.

To the solution of **42a** (1 eq., 384 mg, 1.69 mmol) in ACN (15 mL) was added TFA (30.6 eq., 3.8 mL, 51 mmol) . The mixture was stirred for 36 hours at 25 ˚C and then filtered and washed with Et_2O (3x2 mL). The precipitate was filtered off to give pure **42** (188 mg, 1.09 mmol, 65 %).

^{1}H NMR (400 MHz, METHANOL-d_4) δ 8.12 (d, J = 8.28 Hz, 3H), 7.83 (d, J = 8.28 Hz, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 166.7, 133.7, 133.4, 129.7, 121.3, 104.3, 81.5, 63.4; ESI-MS: $C_{10}H_4NO_2^-$ [M-H]$^-$, 170.0; found 170.0.

3.43 BENZOFURAN-2-CARBONITRILE (43):



43a: 2-(3-Hydroxyprop-1-yn-1-yl)phenol.

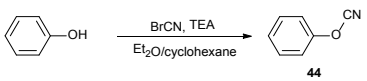
Synthesised using protocol A for Sonogashira coupling (Section D 2.9, page 77). White solid, yield: 77%.

43: Benzofuran-2-carbonitrile.

The compound was obtained as the only product of the standard MnO_2 oxidation protocol (Section D 2.10, page 77). Reaction time: 1 hour. White solid, yield: 91%.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.69 (d, J = 7.80 Hz, 1H), 7.51-7.60 (m, 2H), 7.47 (s, 1H), 7.38 (t, J = 6.80, 1H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 155.7, 128.5, 127.3, 125.5, 124.6, 122.6, 118.5, 112.1, 111.9; ESI-MS: $C_9H_6NO^+$ [M+H]$^+$, 144.0; found 144.1.

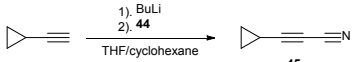
3.44 PHENYL CYANATE (44):



A mixture of cyclohexane (40 mL) and anhydrous ethyl ether (20 mL) was cooled in an ice bath, and treated with phenol (1 eq., 8.47 g, 90 mmol) and cyanogen bromide (1.05 eq., 10 g, 94.9 mmol). While vigorously stirring, TEA (1.05 eq., 13.2 mL, 94.9 mmol) was added dropwise *via* a syringe. The resultant white slurry mass was stirred vigorously for an additional 1 hour at 0 ˚C (ice bath), then the precipitated triethylamine hydrobromide was filtered on a Büchner funnel. The solid was washed with cyclohexane (3x20 mL) and the combined filtrates were distilled (50-70 ˚C at 10 mbar) to give phenyl cyanate (9.22 g, 77.4 mmol, 86%) as a colourless liquid.

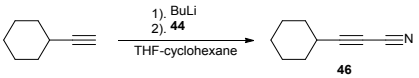
Analyses were in accordance with literature.[826]

3.45 3-CYCLOPROPYLPROPIOLONITRILE (45):



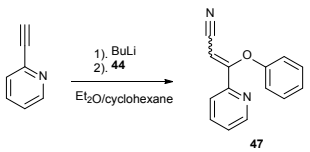
This compound was synthesised following the same procedure as for the compound **46**. Obtained as a colourless oil with the yield of 52%. Analyses were in accordance with literature.[826]

3.46 3-Cyclohexylpropiolonitrile (45):



A solution of cyclohexylacetylene (1 eq., 1.25 mL, 9.57 mmol) in degased anhydrous ethyl ether (2.5 mL) was cooled to -70 ˚C (dry ice/acetone bath), and a hexane solution of BuLi (1.02 eq., 2 M, 4.86 mL, 9.73 mmol) was added dropwise *via* a syringe. Phenyl cyanate (**44**, 1.1 eq., 1.26 g, 10.5 mmol) was added *via* a syringe at a rate such that the reaction temperature could be maintained below -60 ˚C. After stirring for an additional 30 minutes, the resultant solution was brought to 0 ˚C, stirred for 15 min, and then allowed to warm up to room temperature. 10 M solution of sodium hydroxide (10 mL) was added to a flask, stirring continued for 10 minutes (no spot of phenyl cyanate on TLC), organic phase was separated, washed with water (2x10 mL), dried with anhydrous $MgSO_4$. The solvents were removed and the residue was distilled (40-90 ˚C at 5 mbar) to give **46** (955 mg, 7.18 mmol, 75 %) as a colourless liquid. Analyses were in accordance with literature.[826]

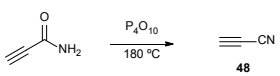
3.47 3-Phenoxy-3-(pyridin-2-yl)acrylonitrile (47):



This compound was the only isolated product while conducting the cyanation with **44** (the same reaction procedure as for **46**). White solid, 90% yield.

^{1}H NMR (400 MHz, ACETONITRILE-d_3) δ 8.94 (br. s., 1H), 8.73 (br. s., 1H), 8.34 (br. s., 1H), 7.73 (br. s., 1H), 7.32-7.50 (m, 2H), 7.08-7.32 (m, 3H), 5.77-6.02 (m, 1H); ^{13}C NMR (101 MHz, ACETONITRILE-d_3) δ 166.2, 156.2, 152.6, 148.8, 135.2, 130.7, 128.8, 125.1, 124.4, 118.7, 115.4, 85.2; ESI-MS: $C_{14}H_{11}N_2O^+$ $[M+H]^+$, 223.0; found 223.0.

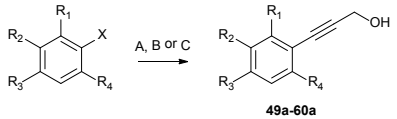
3.48 Propinenitrile (48):



Propiolamide (1 eq., 300 mg, 4.35 mmol) was mixed with seesand (10 g) and phosphorus pentoxide (5 eq., 3.09 g, 21.75 mmol) was added. The obtained mixture was carefully mixed, put under the vacuum and heated to 180 ˚C. The target product **48** (45 mg, 20%) was collected as a distillate in a receiver held in liquid nitrogen.

^{1}H NMR (400 MHz, CHLOROFORM-d) – not informative; ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 105.1, 74.9, 57.5; IR (neat film, cm^{-1}): 3325, 2272, 2075, 862.

3.49 Substituted 3-(aryl)prop-2-yn-1-ols (49a-60a):



	R_1	R_2	R_3	R_4	X	Protocol[a]
49a	OMe	H	H	H	I	B
50a	H	OMe	H	H	Br	C
51a	H	H	OMe	H	I	B
52a	OMe	H	H	OMe	I	C
53a	NH_2	H	H	H	I	B
54a	H	NH_2	H	H	I	A
55a	H	H	NH_2	H	I	A
56a	Me	H	H	H	I	B
57a	Me	H	H	Me	I	A
58a	NO_2	H	H	H	Br	B
59a	H	H	NHAc	H	I	A
60a	H	H	CONHMe	H	I	A

[a] See detailed description in Section D 2.9, page 77.

3-(2-Methoxyphenyl)prop-2-yn-1-ol (49a).
Reaction time: 18 hours; yield: 72%.
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.36 (dd, *J* = 1.5, 7.5 Hz, 1H), 7.26-7.33 (m, 1H), 6.97 (d, *J* = 8.3 Hz, 1H), 6.86-6.93 (m, 1H), 4.43 (s, 2H), 3.84 (s, 3H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 161.6, 134.5, 131.0, 121.5, 113.4, 112.0, 92.7, 82.0, 56.2, 51.5.

3-(3-Methoxyphenyl)prop-2-yn-1-ol (50a).
Reaction time: 16 hours; yield: 87%.
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.15-7.24 (pseudo-t, *J* = 7.5 Hz, 1H), 7.00 (d, *J* = 7.5 Hz, 1H), 6.93-6.98 (m, 1H), 6.87 (dd, *J* = 2.13, 7.5 Hz, 1H), 4.41 (s, 2H), 3.73 (s, 3H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 160.9, 130.6, 125.4, 125.1, 117.8, 115.7, 88.8, 85.6, 55.9, 51.4.

3-(4-Methoxyphenyl)prop-2-yn-1-ol (51a).
Reaction time: 16 hours; yield: 92%.
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.40 (d, *J* = 8.78 Hz, 2H), 6.86 (d, *J* = 8.78 Hz, 2H), 4.51 (d, *J* = 4.9 Hz, 2H), 3.83 (s, 3H), 1.78 (t, *J* = 4.9 Hz, 1H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 159.8, 133.2, 114.6, 114.0, 85.9, 85.7, 55.3, 51.7.

3-(2-Aminophenyl)prop-2-yn-1-ol (52a).
Reaction time: 24 hours; yield: 62%.
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.19 (dd, *J* = 1.25, 7.9 Hz, 1H), 7.03-7.12 (m, 1H), 6.75 (d, *J* = 7.9 Hz, 1H), 6.56-6.65 (m, 1H), 4.47 (s, 2H), 4.26 (s, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 150.3, 133.0, 130.6, 118.2, 115.6, 93.8, 78.9, 69.5, 51.0.

3-(3-Aminophenyl)prop-2-yn-1-ol (53a).
Reaction time: 18 hours; yield: 77%.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.08 (t, *J* = 7.8 Hz, 1H), 6.83 (d, *J* = 7.8 Hz, 1H), 6.76 (s, 1H), 6.64 (dd, *J* = 1.5, 7.8 Hz, 1H), 4.46 (s, 2H), 2.17 (s, 1H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 146.3, 129.2, 123.4, 122.0, 118.0, 115.5, 87.0, 85.6, 51.4.

3-(4-Aminophenyl)prop-2-yn-1-ol (54a).

Reaction time: 18 hours; yield: 42%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.11-7.21 (d, *J* = 8.5 Hz, 2H), 6.53-6.68 (d, *J* = 8.5 Hz, 2H), 4.37 (s, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 149.7, 133.8, 115.7, 112.5, 86.7, 85.9, 51.4;

3-(2-Nitrophenyl)prop-2-yn-1-ol (55a).

Reaction time: 15 hours; yield: 35%.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.97 (d, *J* = 8.0 Hz, 1H), 7.54-7.60 (d, *J* = 8.0 Hz, 1H), 7.46-7.54 (t, *J* = 8.0 Hz, 1H), 7.36-7.44 (t, *J* = 8.0 Hz, 1H), 4.49 (s, 2H), 1.68 (br. s., 1H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 149.9, 134.8, 132.8, 128.9, 124.6, 118.0, 95.2, 80.9, 51.7.

4-(3-Hydroxyprop-1-yn-1-yl)-*N*-methylbenzamide (56a).

Reaction time: 12 hours; yield: 91%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.72-7.82 (m, *J* = 8.28 Hz, 2H), 7.41-7.53 (m, *J* = 8.28 Hz, 2H), 4.43 (s, 2H), 2.92 (s, 3H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 169.9, 135.2, 132.7, 128.3, 127.6, 91.5, 84.7, 51.3, 27.1.

***N*-(4-(3-Hydroxyprop-1-yn-1-yl)phenyl)acetamide (57a).**

Reaction time: 18 hours; yield: 85%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.56-7.64 (d, *J* = 8.8 Hz, 2H), 7.47-7.56 (d, *J* = 8.8 Hz, 2H), 4.74 (s, 2H), 2.04 (s, 3H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 171.9, 143.8, 135.7, 120.7, 112.8, 106.2, 84.5, 62.7, 24.1.

3-(*o*-Tolyl)prop-2-yn-1-ol (58a).

Reaction time: 5 hours; yield: 70%.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.40 (d, *J* = 7.5 Hz, 1H), 7.11-7.24 (m, 3H), 4.54 (s, 2H), 2.43 (s., 3H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 138.2, 131.2, 128.4, 128.0, 119.3, 115.3, 86.5, 85.2, 51.2, 21.2.

3-(2,6-Dimethylphenyl)prop-2-yn-1-ol (59a).

Reaction time: 24 hours; yield: 25%.

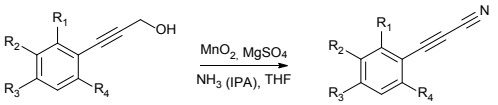
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.03 (t, *J* = 7.5 Hz, 1H), 6.95 (d, *J* = 7.5 Hz, 2H), 4.50 (s, 2H), 2.34 (s, 7H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 140.5, 127.9, 126.7, 122.3, 95.6, 83.3, 51.9, 21.1.

3-(2,6-Dimethoxyphenyl)prop-2-yn-1-ol (60a).

Reaction conditions: 30 ˚C, propylamine, 16 hours; yield: 38%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.25 (t, *J* = 8.4 Hz, 1H), 6.62 (d, *J* = 8.4 Hz, 2H), 4.46 (s, 2H), 3.84 (s, 6H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 163.0, 131.0, 104.7, 102.5, 97.0, 78.1, 56.4, 51.7.

3.50 Substituted 3-aryl-propiolonitriles (49-60):



	R_1	R_2	R_3	R_4	Time, h	Yield[a], %
49	OMe	H	H	H	3	45
50	H	OMe	H	H	2	85
51	H	H	OMe	H	3	95
52	OMe	H	H	OMe	4	60
53	NH_2	H	H	H	1	47
54	H	NH_2	H	H	2	71
55	H	H	NH_2	H	9	94
56	Me	H	H	H	1.5	70
57	Me	H	H	Me	2	55
58	NO_2	H	H	H	4	21
59	H	H	NHAc	H	2	92
60	H	H	CONHMe	H	2	61

[a] See detailed reaction procedures in Section D 2.10, page 77.

3-(2-Methoxyphenyl)propiolonitrile (49, APN-*o*-OMe).

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.51-7.65 (m, 2H), 7.12 (d, *J* = 8.3 Hz, 1H), 7.01 (t, *J* = 7.7 Hz, 1H), 3.94 (s, 4H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 164.7, 136.4, 135.3, 122.0, 112.6, 107.7, 106.4, 81.8, 66.7, 56.7; IR (neat film, cm^{-1}): 2946, 2264, 2142 1596, 1490, 1245, 1164, 1122, 1047, 1021, 752, 498; GC-ESI-HRMS: 157.05276; found 157.05044.

3-(3-Methoxyphenyl)propiolonitrile (50, APN-*m*-OMe).

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.38 (t, *J* = 7.8 Hz, 1H), 7.27 (d, *J* = 7.8 Hz, 1H), 7.20-7.24 (m, 1H), 7.12-7.20 (m, 1H), 3.83 (s, 3H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 161.2, 131.4, 127.1, 120.0, 119.4, 119.2, 106.0, 84.1, 62.7, 56.1; IR (neat film, cm^{-1}): 2491, 2264, 2144, 1595, 1573, 1488, 1464, 1420, 1324, 1294, 1207, 1178, 1045, 783, 681, 494; GC-ESI-HRMS: 157.05276; found 157.05298.

3-(4-Methoxyphenyl)propiolonitrile (51, APN-*p*-OMe).

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.46-7.70 (m, *J* = 8.8 Hz, 2H), 6.86-6.96 (m, *J* = 8.8 Hz, 2H), 3.86 (s, 3H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 161.4, 134.4, 113.7, 108.2, 104.8, 82.7, 61.5, 54.5; IR (neat film, cm^{-1}): 2985, 2358, 2342, 2263, 2178, 2149, 1603, 1514, 1307, 1270, 1180, 1028, 835, 808, 669, 424; GC-ESI-HRMS: 157.05276; found 157.05337.

3-(2,6-Dimethoxyphenyl)propiolonitrile (52, APN-*o,o'*-diOMe):

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.38 (t, *J* = 8.5 Hz, 1H), 6.53 (d, *J* = 8.5 Hz, 2H), 3.88 (s, 6H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 164.4, 133.8, 106.2, 103.4, 96.5, 77.7, 70.5, 56.2; IR (neat film, cm^{-1}): 2847, 2359, 2259, 2201, 2139, 1926, 1586, 1574, 1478, 1432, 1302, 1255, 1188, 1109, 1025, 778, 727, 648, 632, 545, 506, 488, 420; GC-ESI-HRMS: 187.06333; found 184.06465.

3-(2-Aminophenyl)propiolonitrile (53, APN-*o*-NH_2).
^{1}H NMR (500 MHz, METHANOL-d_4) δ 6.81 (d, *J* = 7.88 Hz, 1H), 6.65-6.76 (m, 1H), 6.08-6.19 (m, 2H), 3.85 (br. s., 1H); ^{13}C NMR (126 MHz, CHLOROFORM-d) δ 151.4, 134.0, 133.4, 118.2, 115.0, 105.8, 101.0, 81.6, 68.5; IR (neat film, cm^{-1}): 3413, 3332, 3211, 2925, 2853, 2250, 2136, 1632, 1600, 1563, 1486, 1452, 1312, 1273, 1252, 1161, 740, 673, 493; GC-ESI-HRMS: 142.05310; found 142.05458.

3-(3-Aminophenyl)propiolonitrile (54, APN-*m*-NH_2).
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.17 (t, *J* = 7.6 Hz, 1H), 6.99 (d, *J* = 7.6 Hz, 1H), 6.74-6.89 (m, 2H), 3.85 (br. s., 2H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 146.8, 129.8, 123.6, 118.7, 118.7, 118.0, 105.7, 83.7, 62.3; IR (neat film, cm^{-1}): 3426, 3340, 2923, 2852, 2265, 2142, 1630, 1594, 1579, 1513, 1448, 1326, 1313, 1300, 1220, 1164, 993, 882, 862, 784, 680, 534, 456; GC-ESI-HRMS: 142.05310; found 142.05197.

3-(4-Aminophenyl)propiolonitrile (55, APN-*p*-NH_2).
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.26 (d, *J* = 8.6 Hz, 2H), 6.51 (d, *J* = 8.6 Hz, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 152.5, 135.1, 113.6, 105.6, 102.3, 86.3, 60.2; IR (neat film, cm^{-1}): 3431, 3333, 3211, 2250, 2132, 1632, 1599, 1513, 1438, 1303, 1178, 1043, 949, 826, 814, 526, 495, 452; GC-ESI-HRMS: 142.05310; found 142.05464.

3-(*o*-Tolyl)propiolonitrile (56, APN-*o*-Me).
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.47 (d, *J* = 7.78 Hz, 1H), 7.28-7.36 (m, 1H), 7.18 (d, *J* = 8.03 Hz, 1H), 7.08-7.16 (m, 1H), 2.39 (s, 3H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 143.4, 134.1, 131.8, 130.1, 126.1, 117.4, 105.6, 82.4, 66.4, 20.5; IR (neat film, cm^{-1}): 2295, 2257, 2141, 1599, 1484, 1456, 1383, 1291, 1199, 1162, 1116, 1039, 757, 711, 672, 548, 490, 452; GC-ESI-HRMS: 141.05785; found 141.05926.

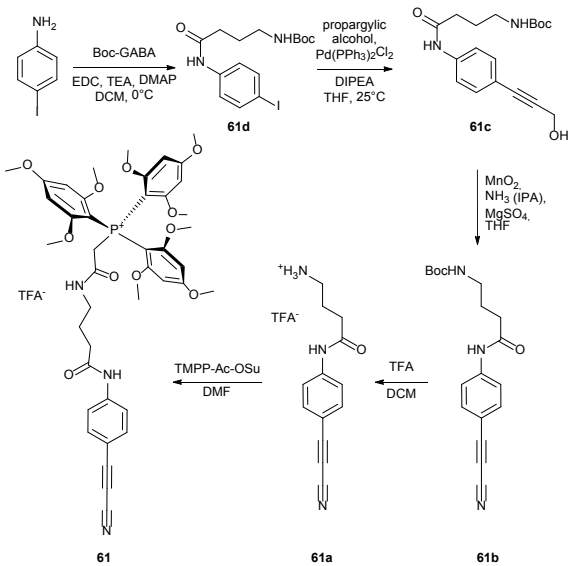
3-(2,6-Dimethylphenyl)propiolonitrile (57, APN-o,o'-diMe).
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.12-7.27 (t, *J* = 7.5 Hz, 1H), 7.01 (d, *J* = 7.5 Hz, 2H), 2.38 (s, 6H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 143.8, 131.2, 127.3, 117.6, 105.6, 81.5, 70.2, 20.8; IR (neat film, cm^{-1}): 2923, 2856, 2261, 2138, 1732, 1595, 1468, 1381, 1265, 1168, 1033, 774, 728, 490; GC-ESI-HRMS: 155.07350; found 155.07507.

3-(2-Nitrophenyl)propiolonitrile (58, APN-*o*-NO_2).
^{1}H NMR (400 MHz, METHANOL-d_4) δ 8.28 – 8.35 (m, 1H), 7.96 – 8.06 (m, 1H), 7.81 – 7.90 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 151.9, 138.3, 135.2, 134.1, 126.6, 114.2, 105.7, 79.0, 68.6; IR (neat film, cm^{-1}): 2268, 1604, 1567, 1528, 1502, 1480, 1345, 851, 787, 744, 709, 687, 537, 491; GC-ESI-HRMS: 172.02728; found 172.02869.

***N*-(4-(Cyanoethynyl)phenyl)acetamide (59, APN-*p*-NHAc).**
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.56-7.63 (m, *J* = 8.8 Hz, 2H), 7.49-7.56 (m, *J* = 8.8 Hz, 2H), 2.04 (s, 3H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 171.9, 143.8, 135.7, 120.7, 112.8, 106.2, 84.5, 62.7, 24.1; IR (neat film, cm^{-1}): 3303, 3174, 3098, 2278, 2262, 2139, 1670, 1594, 1535, 1407, 1364, 1321, 1263, 1177, 834, 534; GC-ESI-HRMS: 184.06366; found 184.06212.

4-(Cyanoethynyl)-*N*-methylbenzamide (60, APN-*p*-CONHMe):
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.96-8.05 (m, *J* = 7.78 Hz, 2H), 7.85-7.93 (m, *J* = 7.78 Hz, 2H), 3.03 (s, 3H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 169.1, 138.3, 134.9, 129.0, 121.6, 105.9, 83.0, 64.6, 28.8; IR (neat film, cm^{-1}): 3348, 2270, 1641, 1549, 1502, 1408, 1392, 1327, 1303, 1283, 1162, 854, 760, 617, 488; GC-ESI-HRMS: 184.06366; found 184.06465.

3.51 (2-((4-((4-(Cyanoethynyl)phenyl)amino)-4-oxobutyl)-amino)-2-oxoethyl)tris(2,4,6-trimethoxyphenyl)phospho-nium trifluoroacetate (61):



61d: *tert*-Butyl (4-((4-iodophenyl)amino)-4-oxobutyl)carbamate.
To the cooled to 0˚C solution of Boc-GABA (1 eq., 0.928 g, 4.57 mmol), TEA (3 eq., 1.39 g, 1.9 mL, 13.7 mmol) and DMAP (0.05 eq., 0.0279 g, 0.228 mmol) in DCM (11.7 mL), EDC (1 eq., 0.875 g, 4.57 mmol) was added. The resulting mixture was stirred for another 10 minutes at 0˚C, an ice bath was removed, and *p*-iodoaniline (1 eq., 1 g, 4.57 mmol) was added and the reaction was left overnight at 25˚C. The resulting mixture was washed with 1M HCl (2x20 mL), water (1x20 mL), and dried over Na_2SO_4 to give **61d** (1125 mg, 2.79 mmol, 61 %), which was used without further purification..
^{1}H NMR (400 MHz, CHLOROFORM-d) δ 9.04 (br. s., 1H), 7.58 - 7.72 (m, *J* = 8.50 Hz, 2H), 7.37 - 7.51 (m, *J* = 8.50 Hz, 2H), 4.81 (br. s., 1H), 3.27 (m, 2H), 2.30 - 2.50 (m, 2H), 1.88 (m, 2H), 1.49 (s, 9H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 174.2, 157.4, 137.8, 120.9, 120.0, 87.3, 77.0, 33.1, 32.8, 28.4, 26.0; ESI-MS: $C_{15}H_{22}N_2O_3^+$ $[M+H]^+$, 405.0; found 405.1.

61c: *tert*-Butyl (4-((4-(3-hydroxyprop-1-yn-1-yl)phenyl)-amino)-4-oxobutyl)carbamate.

Synthesised following the protocol B for Sonogashira coupling (Section D 2.9, page 77). Yellowish solid, yield: 79%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ7.54 - 7.58 (m, *J* = 8.50 Hz, 2H), 7.34 - 7.38 (m, *J* = 8.50 Hz, 2H), 4.40 (s, 2H), 3.13 (t, *J* = 6.90 Hz, 2H), 2.41 (t, *J* = 7.40 Hz, 2H), 1.81-1.89 (m, 2H), 1.44 (s, 9H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 174.0, 158.6, 140.1, 133.2, 120.8, 119.5, 88.3, 85.3, 80.1, 51.3, 40.9, 35.3, 28.8, 27.1; ESI-MS: $C_{18}H_{25}N_2O_4^+$ $[M+H]^+$, 332.1; found 332.0.

61b: *tert*-Butyl (4-((4-(cyanoethynyl)phenyl)amino)-4-oxobutyl)carbamate.

Synthesised using standard protocol of MnO_2 oxidation (Section D 2.10, page 77). Reaction time: 1 hour. White solid, yield: 85%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ7.61 - 7.65 (m, *J* = 8.80 Hz, 2H), 7.54 - 7.59 (m, *J* = 8.50 Hz, 2H), 3.04 (t, *J* = 6.85 Hz, 2H), 2.34 (t, *J* = 7.40 Hz, 2H), 1.74-1.81 (m, 2H), 1.34 (s, 9H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 174.3, 159.1, 143.8, 135.7, 120.8, 112.8, 106.3, 84.6, 62.7, 40.8, 35.3, 34.8, 28.8, 26.9; ESI-MS: $C_{18}H_{22}N_3O_3^+$ $[M+H]^+$, 328.1; found 328.1.

61a: 4-((4-(Cyanoethynyl)phenyl)amino)-4-oxobutan-1-aminium trifluoroacetate.

To a suspension of **61b** (1 eq., 62.8 mg, 0.192 mmol) in DCM (1 mL), TFA (20 eq., 285 µL, 3.83 mmol) was added and the obtained solution was stirred at 25˚C for 30 minutes. The target product **61a** (TFA salt, 65.0 mg, 0.19 mmol, 99 %) was obtained after the evaporation of the reaction mass and was used without further purification in the next step.

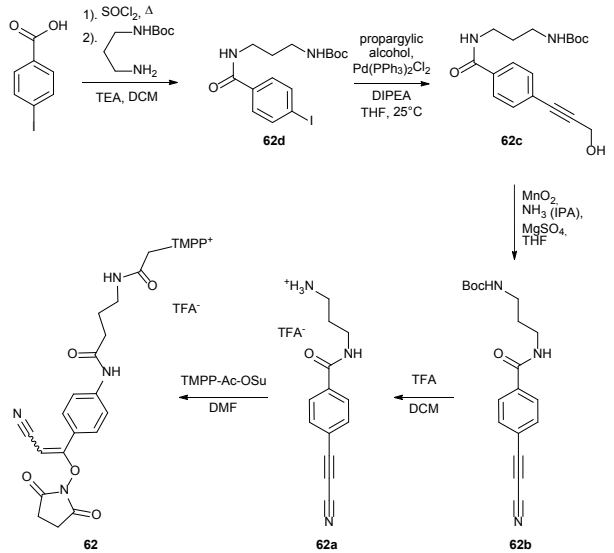
^{1}H NMR (400 MHz, METHANOL-d_4) δ7.71 - 7.79 (m, *J* = 9.15 Hz, 2H), 7.63 - 7.70 (m, *J* = 9.15 Hz, 2H), 3.04 (t, *J* = 6.80 Hz, 2H), 2.6 (t, *J* = 7.05 Hz, 2H), 1.98-2.08 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 173.1, 143.6, 135.7, 120.7, 112.9, 106.2, 84.5, 62.7, 40.4, 34.5, 24.0; ESI-MS: $C_{13}H_{14}N_3O^+$ $[M+H]^+$, 228.1; found 228.1.

61: (2-((4-((4-(Cyanoethynyl)phenyl)amino)-4-oxobutyl)amino)-2-oxoethyl)tris(2,4,6-trimethoxyphenyl)phosphonium trifluoroacetate.

To the solution of **61a** (1 eq., 10.1 mg, 0.0296 mmol) in DMF (250 µL), TEA (1 eq., 4 µL, 0.0296 mmol) was added. TMPP-Ac-OSu (1 eq., 22.7 mg, 0.0296 mmol) was added to the obtained solution and the reaction mass was for 15 minutes at 25 ˚C. The crude product was purified by HPLC to isolate **61** (9.9 mg, 0.0126 mmol, 42 %) as a main product.

^{1}H NMR (400 MHz, METHANOL-d_4) δ7.44 - 7.59 (m, 4H), 6.13 (d, *J* = 4.52 Hz, 6H), 3.75 (s, 9H), 3.50 (s, 18H), 3.00 (td, *J* = 7.91, 15.31 Hz, 2H), 2.26 (t, *J* = 6.90 Hz, 2H), 1.64-1.75 (m, 2H), 1.25-1.43 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ174.2, 167.4, 167.4, 165.3, 143.6, 135.8, 120.6, 112.8, 106.3, 92.2 (d, *J* = 8 Hz), 84.5, 62.7, 56.5, 56.2, 37.5, 29.4 (d, *J* = 64 Hz), 27.9, 27.7, 24.9; ESI-HRMS: $C_{42}H_{47}N_3O_{11}P^+$ $[M]^+$, 800.29427; found 800.29401.

3.52 (2-((4-((4-(2-Cyano-1-((2,5-dioxocyclopentyl)oxy)vinyl)phenyl)amino)-4-oxobutyl)amino)-2-oxoethyl)tris(2,4,6-trimethoxyphenyl)phosphonium trifluoroacetate (62):



This compound was obtained as the only product during the synthesis of the corresponding APN.

62d: *tert*-Butyl (3-(4-iodobenzamido)propyl)carbamate.

4-iodobenzoic acid (1 eq., 760 mg, 3.07 mmol) was heated at 110 ˚C in $SOCl_2$ (9 eq., 2 mL, 27.63 mmol) until complete dissolving (about 15 min). Excess of $SOCl_2$ was removed *in vacuo* and obtained solid was poured into DCM (10 mL), then the obtained solution was cooled to -78 ˚C and TEA (2.1 eq., 900 µL, 6.45 mmol) was added under vigorous stirring; *tert*-butyl (3-aminopropyl)carbamate (1 eq., 534 mg, 3.07 mmol) was added to the resulting mixture still at -78 ˚C, left stirred for another 5 min, then heated up to room temperature, and stirred for another 20 minutes. Ethyl acteate (50 mL) was added to the reaction mixture followed by the addition of 1N HCl (5 mL). The obtained precipitate was filtered, washed with water and dried to yield **62d** (1.06 g, 86%) as a yellowish solid.

^{1}H NMR (400 MHz, METHANOL-d_4) δ7.84 - 7.88 (m, *J* = 8.50 Hz, 2H), 7.57 - 7.62 (m, *J* = 8.50 Hz, 2H), 3.42 (t, *J* = 7.10 Hz, 2H), 3.14 (t, *J* = 6.65 Hz, 2H), 1.71-1.82 (m, 2H), 1.45 (s, 9H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 167.5, 157.1, 134.8, 132.3, 129.0, 127.6, 79.8, 41.0, 39.8, 27.8; ESI-MS: $C_{15}H_{22}N_2O_3^+$ $[M+H]^+$, 405.1; found 405.0.

62c: *tert*-Butyl (3-(4-(3-hydroxyprop-1-yn-1-yl)benzamido)propyl)carbamate.

Synthesised following the protocol B for Sonogashira coupling (Section D 2.9, page 77). White solid, yield: 43%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ7.75 (d, *J* = 8.30 Hz, 2H), 7.44 (d, *J* = 8.30 Hz, 2H), 4.38 (s, 2H), 3.37 (t, *J* = 6.75 Hz, 2H), 3.10 (t, *J* = 6.75 Hz, 2H), 1.97 (s, 1H), 1.72 (quint., 2H), 1.39 (s, 9H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 169.4, 158.6, 135.3, 132.7, 128.4, 127.6, 91.5, 84.8, 80.1, 51.3, 38.9, 38.5, 30.8, 28.9; ESI-MS: $C_{18}H_{25}N_2O_4^+$ $[M+H]^+$, 333.2; found 333.2.

62b: *tert*-Butyl (3-(4-(cyanoethynyl)benzamido)propyl)carbamate.

Synthesised using standard protocol of MnO_2 oxidation (Section D 2.10, page 77). Reaction time: 2 hour. White solid, yield: 73%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ7.92 (d, *J* = 8.20 Hz, 2H), 7.81 (d, *J* = 8.20 Hz, 2H), 3.44 (t, *J* = 6.80 Hz, 2H), 3.15 (t, *J* = 6.70 Hz, 2H), 1.73-1.83 (m, 2H), 1.45 (s, 9H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 169.3, 155.6, 134.9, 133.0, 128.2, 126.8, 112.1, 103.4, 91.4, 80.2, 39.2, 38.3, 30.5, 28.7; ESI-MS: $C_{18}H_{22}N_3O_3^+$ $[M+H]^+$, 328.2; found 328.2.

62a: 3-(4-(Cyanoethynyl)benzamido)propan-1-aminium trifluoroacetate.

To the solution of **62b** (1 eq., 174 mg, 0.534 mmol) in DCM (3 mL), TFA (10 eq., 608 mg, 0.396 mL, 5.34 mmol) was added and the resulting mixture was stirred until complete disappearance of the departure material (control by TLC). The reaction was quenched with methanol and evaporated under reduced pressure to yield **62a** (99 mg, 0.438 mmol, 82 %) as a colourless sticky oil.

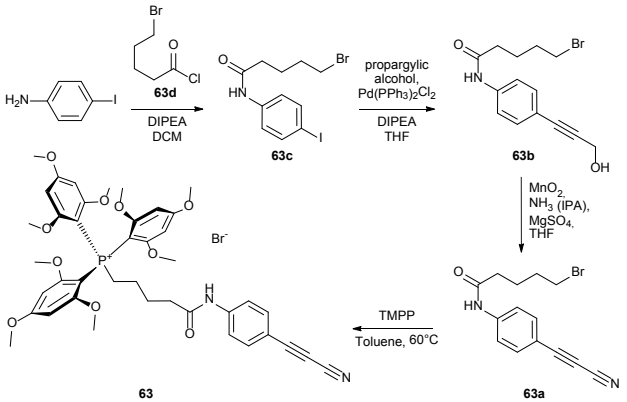
^{1}H NMR (400 MHz, METHANOL-d_4) δ7.52 (d, *J* = 8.70 Hz, 2H), 7.30 (d, *J* = 8.70 Hz, 2H), 3.44 (t, *J* = 6.80 Hz, 2H), 2.94 (t, *J* = 7.80 Hz, 2H), 2.46 (t, *J* = 7.00 Hz, 2H), 1.85-1.96 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 172.8, 146.1, 138.9, 123.0, 114.6, 112.4, 103.5, 87.8, 40.4, 34.4, 24.1; ESI-MS: $C_{13}H_{14}N_3O^+$ $[M+H]^+$, 228.1; found 228.0.

62: (2 ((4 ((4 (2 Cyano 1 ((2,5 dioxocyclopentyl)oxy)vinyl)phenyl)amino)-4-oxobutyl)amino)-2-oxoethyl)tris(2,4,6-trimethoxyphenyl)phosphonium trifluoroacetate.

To the solution of **62a** (1 eq., 9.0 mg, 0.0264 mmol) in DMF (200 µL), TEA (1.2 eq., 4.5 µL, 0.0316 mmol) was added. TMPP-Ac-OSu (1 eq., 20.3 mg, 0.0264 mmol) was added to the obtained solution and obtained reaction mass was stirred until complete disappearance of the starting material. The target **62** (18 mg, 0.0197 mmol, 75 %) was obtained after HPLC purification of the crude product.

^{1}H NMR (400 MHz, METHANOL-d_4) δ7.76 - 7.82 (m, *J* = 8.50 Hz, 2H), 7.70 - 7.75 (m, *J* = 8.50 Hz, 2H), 6.13 (d, *J* = 5.00 Hz, 6H), 5.63 (s, 1H), 4.28 (d, *J* = 16.06 Hz, 2H), 3.73 (s, 9H), 3.53 (s, 18H), 3.12 (t, *J* = 6.65 Hz, 2H), 3.00 - 3.07 (m, 2H), 2.76 (s, 4H), 1.49 (t, *J* = 6.53 Hz, 2H); ESI-HRMS: $C_{46}H_{52}N_4O_{14}P^+$ $[M]^+$, 915.32122; found 915.31585.

3.53 (5-((4-(Cyanoethynyl)phenyl)amino)-5-oxopentyl)tris(2,4,6-trimethoxyphenyl)phosphonium bromide (63):



63d: 5-Bromopentanoyl chloride.

Degassed solution of 5-bromopentanoic acid (1 eq., 2.85 g, 15.7 mmol) and $SOCl_2$ (1 eq., 1.87 g, 1.14 mL, 15.7 mmol) in DCM (50 mL) was refluxed for 3 hours. The resulting mixture was evaporated under reduced pressure to give **63d** (3.11 g, 100%) as a yellowish oil. The crude product was used in the next step without purification.

63c: 5-Bromo-*N*-(4-iodophenyl)pentanamide.

Solution of **63d** (1 eq., 3.11 g, 15.7 mmol) in DCM (50 mL) was poured into a cooled to -78°C solution of 4-iodoaniline (1 eq., 3.45 g, 15.7 mmol) and DIPEA (1 eq., 2.03 g, 2.6 mL, 15.7 mmol) in DCM (50 mL). Obtained reaction mass was allowed to warm to room temperature, stirred for another 30 min, washed with 1N HCl (2x25 mL), water (1x25 mL), dried over Na_2SO_4 and evaporated under reduced pressure to give **63c** (5.60 g, 14.66 mmol, 93%) as brown solid.

^{1}H NMR (400 MHz, METHANOL-d$_4$) δ7.62 - 7.66 (m, 2H), 7.38 - 7.43 (m, 2H), 3.50 (t, *J* = 6.53 Hz, 2H), 2.42 (t, *J* = 7.28 Hz, 2H), 1.81 - 1.98 (m, 4H), 1.37 - 1.42 (m, 1H); ^{13}C NMR (101 MHz, METHANOL-d$_4$) δ174.0, 140.5, 138.9, 123.1, 87.6, 36.9, 33.8, 33.4, 25.3; ESI-MS: $C_{11}H_{14}BrINO^+$ $[M+H]^+$, 381.9; found 381.8.

63b: 5-Bromo-*N*-(4-(3-hydroxyprop-1-yn-1-yl)phenyl)pentanamide:

Synthesised following the protocol A for Sonogashira coupling (Section D 2.9, page 77). Brown solid, yield: 92%.

^{1}H NMR (400 MHz, METHANOL-d$_4$) δ7.49 - 7.63 (m, *J* = 8.53 Hz, 2H), 7.32 - 7.43 (m, *J* = 8.53 Hz, 2H), 4.40 (s, 2H), 4.26 (s, 1H), 3.50 (t, *J* = 6.53 Hz, 2H), 2.43 (t, *J* = 7.15 Hz, 2H), 1.90 - 2.04 (m, 2H), 1.74 - 1.90 (m, 2H), 1.32 (s, 1H); ^{13}C NMR (101 MHz, METHANOL-d$_4$) δ173.9, 140.0, 133.1, 120.7, 119.6, 88.2, 85.2, 51.2, 36.8, 33.7, 33.3, 25.3; ESI-MS: $C_{14}H_{17}BrNO^+$ $[M+H]^+$, 310.0; found 310.0.

63a: 5-Bromo-*N*-(4-(cyanoethynyl)phenyl)pentanamide.

2M solution of NH_3 (4 eq., 94.8 mg, 5.56 mmol) in IPA and anhydrous $MgSO_4$ (15 eq., 2511 mg, 20.9 mmol) were added to a stirred solution of **63b** (1 eq., 431 mg, 1.39 mmol) in THF (3.42 mL). Activated MnO_2 (15 eq., 1814 mg, 20.9 mmol) was added to the solution and the resulting mixture was stirred at 25 °C for 4 hours (controlled by TLC, no more starting alcohol; ***NB:* too long reaction time gives hydrolysis product**), diluted with DCM (13 mL). The mixture was filtered, washed thoroughly with DCM and the combined filtrates were concentrated under reduced pressure. The solid residue was purified by flash chromatography (EtOAc-cyclohexane, 20 min gradient from 0 to 100% of EtOAc) to give **63a** as a white solid (288 mg, 0.946 mmol, 68 %).

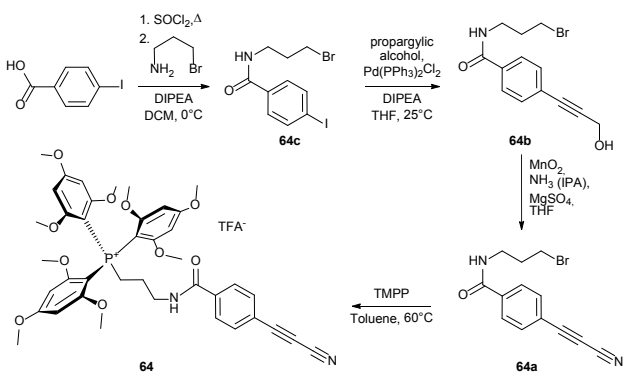
^{1}H NMR (400 MHz, METHANOL-d$_4$) δ7.69 - 7.79 (m, *J* = 8.78 Hz, 2H), 7.59 - 7.69 (m, *J* = 8.78 Hz, 2H), 3.50 (t, *J* = 6.53 Hz, 2H), 1.79 - 1.99 (m, 4H), 1.26 (t, *J* = 7.15 Hz, 2H); ^{13}C NMR (101 MHz, METHANOL-d$_4$) δ174.3, 143.8, 135.7, 120.7, 112.8, 106.2, 101.4, 84.6, 37.0, 33.7, 33.4, 25.2. ESI-MS: $C_{14}H_{14}BrN_2O^+$ $[M+H]^+$, 304.0; found 304.0.

63: (5-((4-(Cyanoethynyl)phenyl)amino)-5-oxopentyl)tris(2,4,6-trimethoxy–phenyl)phosphonium bromide.

63a (1 eq., 20 mg, 0.0655 mmol) and tris(2,4,6-trimethoxyphenyl)phosphane (TMPP, 1.2 eq., 41.9 mg, 0.0786 mmol) were dissolved in dry toluene (1 mL) and stirred overnight at 25 °C. **63** (TFA salt, 22 mg, 39%) was obtained after reverse-phase HPLC as a white solid.

^{1}H NMR (400 MHz, METHANOL-*d*$_4$) δ7.51-7.55 (m, 4H), 6.13 (d, *J* = 4.77 Hz, 6H), 3.75 (s, 9H), 3.50 (s, 18H), 3.00 (td, *J* = 6.90, 15.31 Hz, 2H), 2.26 (t, *J* = 6.90 Hz, 2H), 1.73 (m, 2H), 1.22 - 1.45 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-*d*$_4$) δ 174.2, 167.4, 165.3, 143.6, 135.8, 120.6, 112.8, 106.2, 93.6, 92.3, 92.2, 84.5, 62.7, 56.3, 37.5, 29.7, 27.7, 24.9; ESI-HRMS: $C_{41}H_{46}N_2O_{10}P^+$ $[M]^+$, 757.28846; found 757.29552.

3.54 (4-(4-(Cyanoethynyl)benzamido)butyl)tris(2,4,6-trimethoxyphenyl)phosphonium trifluoroacetate (64):



64c: *N*-(4-Bromobutyl)-4-iodobenzamide.

4-iodobenzoic acid (1 eq., 1.45 g, 5.85 mmol) was heated at 110 ˚C in $SOCl_2$ (9 eq., 3.8 mL, 52.6 mmol) until complete dissolving (around 15 min). Excess of $SOCl_2$ was removed *in vacuo* and obtained solid was poured into DCM (15 mL), cooled to -78 ˚C and DIPEA (3.1 eq., 3 mL, 18.2 mmol) was added under vigorous stirring. 3-Bromopropylamine hydrobromide (1.5 eq., 1.90 g, 8.77 mmol) was added to the resulting mixture was left stirring for 5 minutes still at -78 ˚C, let to warm up to room temperature, while stirring for another 20 minutes. Ethyl acetate (100 mL) was added with 1M HCl (5 mL), obtained solid was filtered (product), washed with water and dried to yield **64c** (2.09 g, 5.67 mmol, 97 %) as a white solid.

^{1}H NMR (400 MHz, METHANOL-d_4) δ7.85 (m, *J* = 8.40 Hz, 2H), 7.58 (m, *J* = 8.40 Hz, 2H), 3.48-3.56 (m, 4H), 2.12-2.23 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 160.1, 139.2, 136.0, 132.8, 102.4, 50.1, 43.2, 23.0; ESI-MS: $C_{10}H_{12}BrINO^+$ $[M+H]^+$, 367.9; found 368.0.

64b: *N*-(4-Bromobutyl)-4-(3-hydroxyprop-1-yn-1-yl)benzamide.

Synthesised following the protocol B for Sonogashira coupling (Section D 2.9, page 77). Brown solid, yield: 81%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ7.73 (m, *J* = 8.40 Hz, 2H), 7.38 (m, *J* = 8.40 Hz, 2H), 4.37 (t, *J* = 5.30 Hz, 2H), 4.34 (s, 2H), 3.51 (t, *J* = 5.80 Hz, 2H), 1.92-1.98 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 159.2, 134.2, 132.4, 128.2, 127.1, 91.2, 84.8, 67.2, 51.2, 43.3, 22.5; ESI-MS: $C_{13}H_{15}BrNO_2^+$ $[M+H]^+$, 295.0; found 295.0.

64a: *N*-(3-Bromopropyl)-4-(cyanoethynyl)benzamide.

Synthesised using standard protocol of MnO_2 oxidation (Section D 2.10, page 77). Reaction time: 45 minutes. Brown solid, yield: 52%.

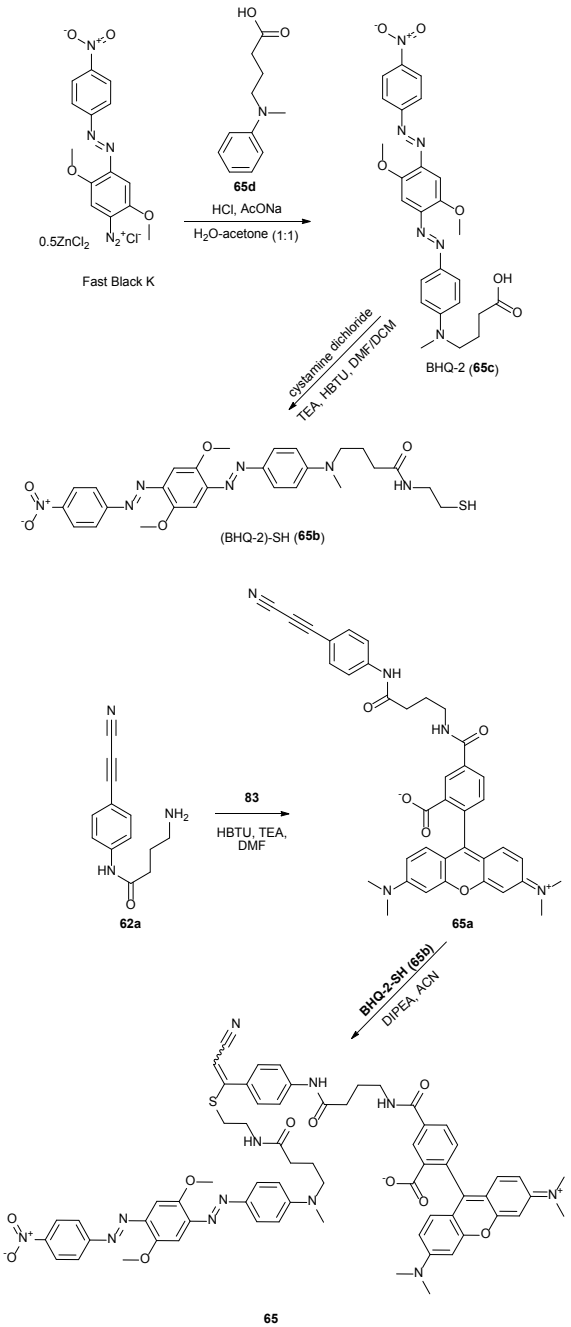
^{1}H NMR (400 MHz, METHANOL-d_4) δ7.90 (m, *J* = 8.50 Hz, 2H), 7.80 (m, *J* = 8.50 Hz, 2H), 3.42-3.55 (m, 2H), 3.25-3.35 (m, 2H), 2.13-2.23 (m, 2H), 1.92-1.98 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 168.7, 138.8, 134.9, 128.9, 121.4, 105.8, 83.0, 67.3, 41.9, 39.8, 22.8; ESI-MS: $C_{13}H_{12}BrN_2O^+$ $[M+H]^+$, 291.0; found 291.2.

64: (4-(4-(Cyanoethynyl)benzamido)butyl)tris(2,4,6-trimethoxyphenyl)phosphonium trifluoroacetate.

64a (1 eq., 30 mg, 0.103 mmol) and tris(2,4,6-trimethoxyphenyl)phosphane (TMPP, 1 eq., 54.9 mg, 0.103 mmol) were dissolved in dry toluene (2 mL). The obtained solution was left overnight at 25 ˚C. The precipitate was filtered, resolubilised in DMSO, and purified by HPLC to give **64** (35 mg, 0.0409 mmol, 40 %) as a white solid.

^{1}H NMR (400 MHz, METHANOL-d_4) δ7.75 (d, *J* = 8.50 Hz, 2H), 7.69 (d, *J* = 8.50 Hz, 2H), 6.16 (d, *J* = 4.70 Hz, 2H), 3.76 (s, 9H), 3.51 (s, 18H), 3.35 (t, *J* = 7.10 Hz, 2H), 2.98-3.10 (m, 2H), 1.53-1.64 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 168.5, 167.5, 165.3, 138.7, 134.9, 128.8, 121.2, 105.8, 94.0, 92.9, 92.3, 82.9, 64.5, 56.5, 41.7, 27.8, 25.7; ESI-HRMS: $C_{40}H_{44}N_2O_{10}P^+$ $[M]^+$, 743.22728; found 743.23946.

3.55 5-((4-((4-(2-Cyano-1-((2-(4-((4-((*E*)-(2,5-dimethoxy-4-((*E*)-(4-nitrophenyl)diazenyl)phenyl)diazenylphenyl)-(methyl)amino)butanamido)ethyl)thio)vinyl)phenyl)amino)-4-oxobutyl)carbamoyl)-2-(6-(dimethylamino)-3-(dimethyliminio)-*3H*-xanthen-9-yl)benzoate (65):



65d: 4-(Methyl(phenyl)amino)butanoic acid.

This compound was synthesised following a reported protocol.[844]

65c: 4-((4-((*(E)*-(2,5-Dimethoxy-4-((*(E)*-(4-nitrophenyl)diazenyl)phenyl)diazenyl)phenyl)(methyl)amino)-butanoic acid (BHQ-2).

Fast Black K hemi (zinc chloride) salt (practical grade, ≈30% dye content) (7.76 g) was suspended in cold water (150.0 mL, 0°C) and stirred for 20 minutes. The suspension was filtered, and the red solution was added dropwise to a cold (0 °C) mixture of **65d** (1.33 g, 6.88 mmol), concentrated hydrochloric acid (3.1 mL) and sodium acetate (3.6 g, 43.90 mmol) in water-acetone mixture (1:1) (150.0 mL). The reaction mixture was stirred at 10 ˚C for 15 minutes and at 25 ˚C for 2 hours. Then the reaction crude was extracted with ethyl acetate (3x150 mL) and the combined organic layers were dried over Na_2SO_4. The crude product was purified by column chromatography on silica gel (100% EtOAc, then 100% DCM to DCM/MeOH (95:5)). **BHQ-2** (1.36 g, 39%) was obtained as a dark violet solid.

1H NMR (400 MHz, METHANOL-d_4) δ8.31 (d, *J* = 9.0 Hz, 2H), 8.00 (d, *J* = 9.0 Hz, 2H), 7.86 (d, *J* = 9.0 Hz, 2H), 7.45 (s, 1H), 7.40 (s, 1H), 6.77 (d, *J* = 9.0 Hz, 2H), 4.05 (s, 3H), 4.00 (s, 3H), 3.5 (t, *J* = 7.1 Hz, 2H), 2.36 (t, *J* = 7.1 Hz, 2H), 1.98-1.90 (m, 2H); 13C NMR (101 MHz, METHANOL-d_4) δ176.2, 157.1, 154.3, 153.0, 151.4, 149.0, 147.4, 145.0, 142.6, 126.9, 125.3, 124.2, 112.1, 101.7, 100.7, 57.2, 52.3, 39.0, 31.6, 22.9.

65b: 4-((4-((*(E)*-(2,5-Dimethoxy-4-((*(E)*-(4-nitrophenyl)diazenyl)phenyl)diazenyl)phenyl)(methyl)amino)-*N*-(2-mercaptoethyl)butanamide ((BHQ-2)-SH).

BHQ-2 (1 eq., 92.2 mg, 0.182 mmol) was dissolved in a mixture of DMF (5 mL) and DCM (10 mL). TEA (6 eq., 152 µL, 1.09 mmol) and cystamine dichloride (5 eq., 204 mg, 0.91 mmol) were added. The mixture was cooled to 0 ˚C and HBTU (1 eq., 69 mg, 0.182 mmol) was added. The solution was allowed to reach room temperature and stirred for 15 hours. When total conversion was reached, DTT (6 eq., 168 mg, 0.162 mL, 1.09 mmol) was added. After the resulting mixture has been stirred for 10 minutes at 25 ˚C, the crude was diluted with saturated $NaHCO_3$ solution (75 mL) and extracted with EtOAc (2x50 mL). The organic layers were combined, washed with water (50 mL), brine (50 mL) and dried over Na_2SO_4. The crude product was purified by column chromatography on silica gel (DCM/MeOH from 100:0 to 95:5) to yield **(BHQ-2)-SH** (60.7 mg, 0.107 mmol, 59 %) as a dark violet solid.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ8.33 (d, *J* = 9.0 Hz, 2H), 8.0 (d, *J* = 9.1 Hz, 2H), 7.9 (d, *J* = 9.1 Hz, 2H), 7.42 (s, 1H), 7.42 (s, 1H), 6.75 (d, *J* = 9.00 Hz, 2H), 5.90 (t, *J* = 5.6 Hz, 1H), 4.06 (s, 3H), 4.01 (s, 3H), 3.49 (t, *J* = 7.4 Hz, 2H), 3.41 (dt, *J* = 6.2, 6.4 Hz, 2H), 2.64 (td, *J* = 6.4, 8.47 Hz, 2H), 2.24 (t, *J* = 7.4 Hz, 2H), 2.01-1.94 (m, 2H); 13C NMR (101 MHz, CHLOROFORM-d) δ172.2, 156.6, 153.8, 152.4, 151.1, 148.5, 147.0, 144.7, 142.3, 126.4, 124.9, 123.7, 111.6, 101.2, 100.3, 57.0, 56.9, 51.8, 42.5, 38.7, 33.4, 24.9, 23.0; HR-ESI-MS: $C_{27}H_{31}N_7O_5S$, 565.2107; found 565.2105.

65a: 5-((4-((4-(Cyanoethynyl)phenyl)amino)-4-oxobutyl)carbamoyl)-2-(6-(dimethylamino)-3-(dimethyliminio)-*3H*-xanthen-9-yl)benzoate.

To a cooled to 0 ˚C degased solution of 62a (1 eq., 17.3 mg, 0.0507 mmol) and TAMRA-5'-COOH (**83**, 1 eq., 21.8 mg, 0.0507 mmol) in DMF (1.4 mL), HBTU (1 eq., 19.2 mg) was added at 0 ˚C. Obtained reaction mass was stirred for 5 minutes and TEA was added. The resulting mixture was stirred for 1 hour at 25 °C, evaporated and purified by HPLC to yield 65a (22 mg, 68%) as a dark-violet solid.

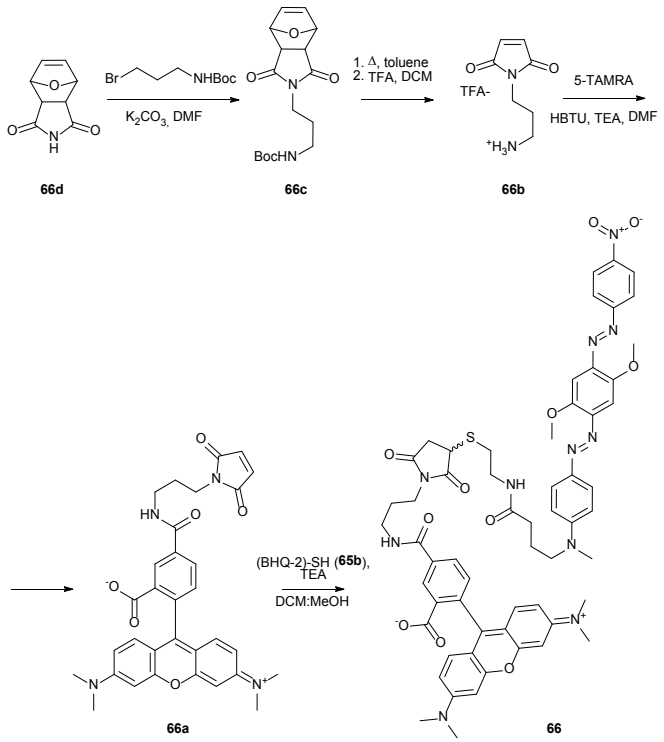
^{1}H NMR (400 MHz, METHANOL-d_4) δ8.8 (br. s, 1H), 8.7 (s, 1H), 8.08 - 8.16 (d, *J* = 8.2 Hz, 1H), 7.60 - 7.70 (d, *J* = 8.9 Hz, 2H), 7.49 - 7.58 (d, *J* = 8.9 Hz, 2H), 7.32-7.39 (d, *J* = 8.2, 1H), 7.01 (s, 4H), 6.93 (s, 2H), 3.48-3.58 (m, 2H), 3.26 (s, 12H), 2.44-2.54 (t, *J* = 7.17 Hz, 2H), 1.98-2.12 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) – not informative; HR-ESI-MS: 639.24817; found 639.24310.

65: 5-((4-((4-(2-Cyano-1-((2-(4-((4-((*(E)*-(2,5-dimethoxy-4-((*(E)*-(4-nitrophe-nyl)diazenyl)phenyl)diazenylphenyl)-(methyl)amino)butanamido)ethyl)thio)vinyl)phenyl)amino)-4-oxobutyl)carbamoyl)-2-(6-(dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)benzoate.

To a degased solution of BHQ-SH (**65b**) (1.13 eq., 2 mg, 0.00354 mmol) in DCM (0.5 mL), a degased solution of TAMRA-APN (**24**) (1 eq., 2 mg, 0.00313 mmol) in methanol (0.5 mL) was added. TEA (4.6 eq., 2 µL, 0.0144 mmol) was added and the resulting mixture was left overnight at 25 °C. Solvents were evaporated; the crude product was solubilised in DMSO (0.5 mL) and purified by HPLC to give BHQ-APN-TAMRA (**65**, 2.7 mg, 0.00225 mmol, 72%) as dark-violet solid.

^{1}H NMR (400 MHz, METHANOL-d4)δ8.63 (d, *J* = 2.0 Hz, 1H), 8.34 (d, *J* = 8.8 Hz, 2H), 8.22 (dd, *J* = 7.8, 2.0 Hz, 1H), 8.00 (d, *J* = 8.8 Hz, 2H), 7.75 (d, *J* = 9.0 Hz, 2H), 7.69 (d, *J* = 8.5 Hz, 2H), 7.53 (d, *J* = 8.0 Hz, 1H), 7.39 (d, *J* = 8.5 Hz, 2H), 7.34 (d, *J* = 9.0 Hz, 2H), 6.75-6.86 (m, 8H), 5.48 (s, 1H), 4.02 (s, 3H), 3.92 (s, 3H), 3.58-3.63 (m, 2H), 3.45-3.52 (m, 2H), 3.21 (s, 12H), 3.16 (t, *J* = 6.9 Hz, 2H), 3.08 (s, 3H), 2.71 (t, *J* = 6.5 Hz, 2H), 2.55 (t, *J* = 6.2 Hz, 2H), 2.22 (t, *J* = 6.9 Hz, 2H), 2.08-2.16 (m, 2H), 1.88-1.96 (m, 2H), 1.61 (br.s, 1H); ^{13}C NMR (101 MHz, METHANOL-d_4) – not informative; HR-ESI-MS: $C_{65}H_{65}N_{12}O_{10}S^+$ $[M+H]^+$, 1205.46618; found 1205.46748.

3.56 5-((3-(3-((2-(4-((4-((*(E)*-(2,5-DIMETHOXY-4-(*(E)*-(4-NITROPHENYL)DIAZENYL)PHENYL)DIAZENYL)PHENYL)(METHYL)AMINO)BUTANAMIDO)ETHYL)THIO)-2,5-DIOXOPYRROLIDIN-1-YL)PROPYL)CARBAMOYL)-2-(6-(DIMETHYLAMINO)-3-(DIMETHYLIMINIO)-*3H*-XANTHEN-9-YL)BENZOATE (66):



66d: 3a,4,7,7a-Tetrahydro-*1H*-4,7-epoxyisoindole-1,3(*2H*)-dione.

This compound was synthesised according to the previously described procedure. [845]

66c: *tert*-Butyl (3-((3aR,7aS)-1,3-dioxo-3a,4,7,7a-tetrahydro-*1H*-4,7-epoxyisoindol-2(*3H*)-yl)propyl)carbamate.

To the solution of **66d** (1 eq., 1.76 g, 10.7 mmol) and *tert*-butyl N-(3-bromopropyl)carbamate (2 eq., 5.07 g, 21.3 mmol) in DMF (20 mL), K_2CO_3 (1.2 eq., 1.77 g, 12.8 mmol) was added. The resulting mixture was heated at 50 ˚C for 18 hours. The solution was left to cool down; a solid residue was filtered and washed with DMF. United organic fractions were evaporated, hexane (50 mL) was added to the obtained slurry mass. Obtained suspensions were stirred for another hour, filtered and washed with hexane to give **66c** (3.36 g, 10.4 mmol, 98 %) as white solid.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ6.49 (s, 2H), 5.23 (s, 2H), 3.52 (t, *J* = 6.5 Hz, 2H), 2.96 – 3.09 (m, 2H), 2.82 (s, 2H), 1.66-1.75 (m, 2H), 1.41 (s, 9H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ176.5, 155.9, 136.5, 81.0, 79.3, 47.5, 37.1, 36.0, 28.4, 27.8.

66b: 1-(3-Aminopropyl)-*1H*-pyrrole-2,5-dione (TFA salt).

A solution of **66c** (1 eq., 243 mg, 0.754 mmol) in toluene (25 mL) was refluxed for 3 hours. Toluene was evaporated under replaced pressure; the obtained white crude product was resolubilised in DCM (5 mL), TFA (0.5 mL) was added. Stirring was continued for 2 hours until complete disappearance of a starting material (controlled by TLC). Solvent were evaporated after the reaction was quenched by methanol (3 mL). Obtained 1-(3-aminopropyl)-*1H*-pyrrole-2,5-dione (**66b**, TFA salt, 190 mg, 94%) was used without further purification.

^{1}H NMR (400 MHz, METHANOL-d$_4$) δ6.76 (s, 2H), 3.52 (t, *J* = 6.7 Hz, 2H), 2.80-2.88 (m, 2H), 1.76-1.88 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d4) δ170.6, 135.6, 38.5, 35.4, 28.0.

66a: 2-(6-(Dimethylamino)-3-(dimethyliminio)-*3H*-xanthen-9-yl)-5-((3-(2,5-dioxo-2,5-dihydro-*1H*-pyrrol-1-yl)propyl)carbamoyl)benzoate (TAMRA-maleimide):

To a solution of TAMRA-5'-COOH (1 eq., 71.5 mg, 0.166 mmol) in DMF (3.21 mL), TEA (2.5 eq., 57.7 µL, 0.415 mmol) and HATU (1.12 eq., 70.7 mg, 0.186 mmol) here added. Obtained reaction mass was stirred for another 5 minutes and **66b** (1 eq., 71.5 mg, 0.166 mmol) was added. Stirring continued for 25 minutes and the reaction mass was evaporated under reduced pressure to the volume of about 1 mL, and the reaction mass was purified by preparative HPLC to give TAMRA-Maleimide (66a, 34.8 mg, 0.0615 mmol, 37 %) as a pink solid.

^{1}H NMR (400 MHz, METHANOL-d$_4$) δ8.66 (d, *J* = 1.8 Hz, 1H), 8.14 (dd, *J* = 1.8, 8.0 Hz, 2H), 7.41 (d, *J* = 8.0 Hz, 2H), 7.02 (d, *J* = 9.5 Hz, 1H), 6.92 (dd, *J* = 9.5, 2.2 Hz, 2H), 6.81 (d, *J* = 2.2 Hz, 2H), 6.72 (s, 2H), 3.52 (t, *J* = 6.8 Hz, 2H), 3.34 (t, *J* = 7.0 Hz, 2H), 3.17 (s, 12H), 1.80-1.90 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d$_4$) δ172.5, 168.2, 167.4, 160.6, 159.0, 158.9, 138.1, 137.7, 137.6, 135.5, 132.9, 132.3, 132.0, 131.4, 115.6, 114.8, 97.5, 82.4, 41.0, 38.6, 36.4, 29.3. HR-ESI-MS: $C_{32}H_{30}N_4O_6$, 566.21653; found 566.21654.

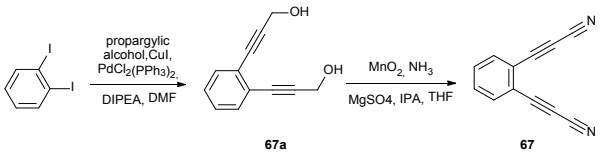
66: 5-((3-(3-((2-(4-((4-((*(E)*-(2,5-Dimethoxy-4-(*(E)*-(4-nitrophenyl)diazenyl)phenyl)diazenyl)phenyl)(methyl)amino)butanamido)ethyl)thio)-2,5-dioxopyrrolidin-1-yl)propyl)carbamoyl)-2-(6-(dimethylamino)-3-(dimethyliminio)-*3H*-xanthen-9-yl)benzoate.

To a degased solution of BHQ-SH (**65b**) (1.15 eq., 4.6 mg, 0.00812 mmol) in DCM (0.5 mL), a degased solution of TAMRA-Maleimide (**66a**) (1 eq., 4 mg, 0.00313 mmol) in methanol (0.5 mL) was added. TEA (5 eq., 5 µL, 0.0353 mmol) was added and the resulting mixture was left overnight at 25 °C. Solvents were evaporated; the crude product was solubilised in DMSO (0.5 mL) and purified by HPLC to give **66** (7 mg, 0.00621 mmol, 88%) as dark-violet solid.

^{1}H NMR (400 MHz, DMSO-d$_6$) δ8.91 (t, *J* = 6.1 Hz, 1H), 8.68 (s, 1H), 8.43 (d, *J* = 9.1 Hz, 2H), 8.31 (d, *J* = 8.9 Hz, 1H), 8.01-8.10 (m, 3H), 7.77 (d, *J* = 9.1 Hz, 2H), 7.60 (d, *J* = 7.8 Hz, 1H), 7.39 (s, 1H), 7.33 (s, 1H), 6.99 (s, 3H), 6.89 (s, 1H), 6.85 (d, *J* = 9.1 Hz, 2H), 4.04 (dd, *J* = 3.9, 8.9 Hz, 1H), 3.98 (s, 3H), 3.92 (s, 3H), 3.41-3.45 (m, 2H), 3.27-3.38 (m, 6H), 3.23 (m, 12H),

3.06 (s, 3H), 2.85-2.95 (m, 1H), 2.72-2.81 (m, 1H), 2.52-2.56 (m, 2H), 2.17 (t, *J* = 7.3 Hz, 2H), 1.74-1.88 (m, 4H). HR-ESI-MS: $C_{59}H_{62}N_{11}O_{11}S^+$ $[M+H]^+$, 1132.43455; found 1132.43384.

3.57 3,3'-(1,2-PHENYLENE)DIPROPIOLONITRILE (67):



67a 67

67a: 3,3'-(1,2-Phenylene)bis(prop-2-yn-1-ol).

To the degased solution of 1,2-diiodobenzene (1 eq., 661 mg, 0.262 mL, 2 mmol) and propargylic alcohol (2.3 eq., 272 µL, 4.61 mmol) in butyl amine (15.8 mL), $Pd(PPh_3)_4$ (4%, 92.6 mg, 0.0801 mmol) was added and the resulting mixture was refluxed overnight. Solvents were evaporated and the obtained crude product was purified by flash chromatography (20 minutes gradient EtOAc/Cyclohexane) to yield **67a** (150 mg, 0.8 mmol, 40 %) as a brownish solid.

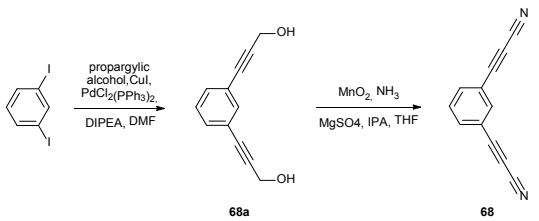
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.38 - 7.53 (m, 2H), 7.25 - 7.38 (m, 2H), 4.48 (s, 4H); ^{13}C NMR (101 MHz, METHANOL-d_4) d 135.6, 131.9, 129.2, 95.6, 86.6, 53.9; ESI-MS: $C_{12}H_{11}O_2^+$ $[M+H]^+$, 187.1; found 187.1.

67: 3,3'-(1,2-Phenylene)dipropiolonitrile.

The compound was obtained as the only product of the standard MnO_2 oxidation protocol (Section D 2.10, page 77). Reaction time: 75 minutes. Brown solid, yield: 42%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.89 (dd, *J* = 3.30, 5.80 Hz, 2H), 7.73 (dd, *J* = 3.30, 5.80 Hz, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 136.0, 133.5, 126.5, 105.5, 80.2, 67.2; GC-ESI-MS: $C_{12}H_5N_2^+$ $[M+H]^+$, 177.0; found 177.0.

3.58 3,3'-(1,3-PHENYLENE)DIPROPIOLONITRILE (68):



68a 68

68a: 3,3'-(1,3-Phenylene)bis(prop-2-yn-1-ol).

Same procedure as for the synthesis of **67a**. Brownish solid, yield: 55%.

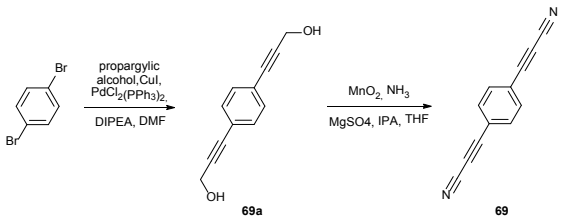
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.47 (s, 1H), 7.36 - 7.43 (m, 2H), 7.29 - 7.36 (m, 1H), 4.41 (s, 4H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 135.3, 132.5, 129.8, 124.8, 89.7, 84.5, 51.2; ESI-MS: $C_{12}H_{11}O_2^+$ $[M+H]^+$, 187.1; found 187.0.

68: 3,3'-(1,3-Phenylene)dipropiolonitrile.

The compound was obtained as the only product of the standard MnO_2 oxidation protocol (Section D 2.10, page 77). Reaction time: 2 hours. Brown solid, yield: 35%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 8.10 (d, *J* = 1.50 Hz, 1H), 7.93 (dd, *J* = 1.50, 8.00 Hz, 1H), 7.63 (t, *J* = 8.00 Hz, 2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 139.3, 137.8, 131.2, 120.0, 105.7, 81.7, 64.2; GC-ESI-MS: $C_{12}H_5N_2^+$ $[M+H]^+$, 177.0; found 177.1.

3.59 3,3'-(1,4-PHENYLENE)DIPROPIOLONITRILE (69):



69a 69

69a: 3,3'-(1,4-Phenylene)bis(prop-2-yn-1-ol).

Same procedure as for the synthesis of **67a**, but refluxed for 72 hours. Brownish solid, yield: 35%.

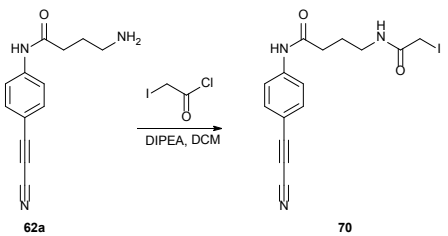
^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.39 (s, 4H), 4.41 (s, 4H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 132.6, 124.3, 101.4, 90.8, 84.9, 51.2; ESI-MS: $C_{12}H_{11}O_2^+$ $[M+H]^+$, 187.1; found 187.1.

69: 3,3'-(1,4-Phenylene)dipropiolonitrile.

The compound was obtained as the only product of the standard MnO_2 oxidation protocol (Section D 2.10, page 77). Reaction time: 2 hours. Brown solid, yield: 19%.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.94 (s, 4H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 135.0, 121.6, 105.5, 82.0, 65.9; GC-ESI-MS: $C_{12}H_5N_2^+$ $[M+H]^+$, 177.0; found 177.0.

3.60 N-(4-(CYANOETHYNYL)PHENYL)-4-(2-IODOACETAMIDO)-BUTANAMIDE (70):



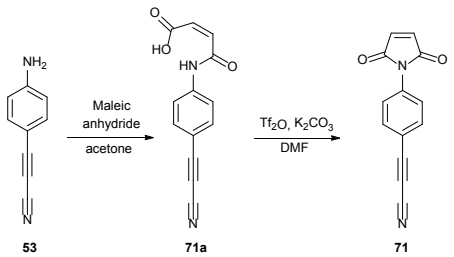
62a 70

To the suspension of **62a** (TFA salt, 1 eq., 21.4 mg, 0.0629 mmol) in DCM (4 mL), DIPEA (4.33 eq., 45 µL, 0.272 mmol) was added. Obtained solution was cooled to -78 °C and iodoacetyl chloride (1.5 eq., 8.5 µL, 0.0943 mmol) was added to give slightly yellow solution, which was left to warm up to room temperature and stirred for another 20 minutes. DCM (10 mL) was added to the reaction mass, which was then washed with 1M HCl (2x2 mL), water (5 mL), dried over Na_2SO_4, and evaporated under reduced pressure to give crude product, which was purified by flash chromatography to give **70** (15.9 mg, 0.0402 mmol, 64 %) as a white solid.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.58 - 7.64 (m, *J* = 8.55 Hz, 2H), 7.49 - 7.57 (m, *J* = 8.55 Hz, 2H), 3.58 (s, 2H), 3.16 (t, *J* = 6.78 Hz, 2H), 2.35 (t, *J* = 7.40 Hz, 2H), 1.78 (m,

2H); ^{13}C NMR (101 MHz, METHANOL-d_4) δ 172.5, 170.1, 142.3, 134.3, 119.3, 111.4, 104.8, 87.0, 83.2, 61.3, 38.9, 33.7, 24.6; ESI-MS: $C_{15}H_{15}N_3O_2^+$ [M+H]$^+$, 396.0; found 396.0.

3.61 3-(4-(2,5-Dioxo-2,5-dihydro-*1H*-pyrrol-1-yl)phenyl)propiolonitrile (71):



71a: *(Z)*-4-((4-(Cyanoethynyl)phenyl)amino)-4-oxobut-2-enoic acid.

To the solution of **53** (1 eq., 76.8 mg, 0.541 mmol) in acetone (2 mL), maleic anhydride (2 eq., 106 mg, 1.08 mmol) was added. A yellowish solid was obtained after about 7 hours of stirring. The resulting mixture was evaporated under replaced pressure, an excess of maleic anhydride and maleic acid was washed with methanol. **71a** (127 mg, 0.53 mmol, 98 %) was obtained as yellowish solid, no further purification was needed.

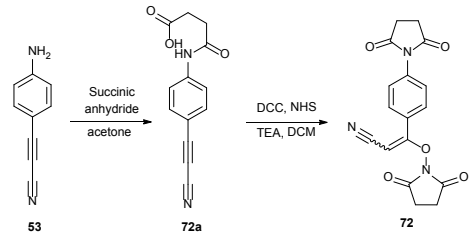
^{1}H NMR (400 MHz, DMSO-d_6) δ 12.90 (br. s., 1H), 10.70 (s, 1H), 7.62 - 7.90 (m, 4H), 6.50 (d, *J* = 11.90 Hz, 1H), 6.34 (d, *J* = 11.90 Hz, 1H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 166.8, 163.8, 142.4, 135.0, 131.7, 130.1, 119.3, 110.2, 105.6, 84.3, 61.9; ESI-MS: $C_{13}H_7N_2O_3^-$ [M-H]$^-$, 239.0; found 239.0.

71: 3-(4-(2,5-Dioxo-2,5-dihydro-*1H*-pyrrol-1-yl)phenyl)-propiolonitrile.

To the solution of **71a** (1 eq., 75 mg, 0.312 mmol) in dry DMF (1.21 mL) trifluoroacetic anhydride (2 eq., 86.9 µL, 0.624 mmol) was added. Stirring continued for another 5 minutes at 25 ˚C and K_2CO_3 (3 eq., 129 mg, 0.937 mmol) was added. The resulting mixture stirred for another 60 minutes, then directly purified by HPLC to give **71** (65.9 mg, 0.297 mmol, 95%) as a slightly yellow solid.

^{1}H NMR (400 MHz, DMSO-d_6) δ 7.81 (m, *J* = 8.50 Hz, 2H), 7.52 (m, *J* = 8.50 Hz, 2H), 6.96 (s, 2H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 169.0, 134.4, 134.0, 126.0, 117.3, 117.0, 82.2, 78.5, 62.3; ESI-MS: $C_{13}H_7N_2O_2^+$ [M+H]$^+$, 223.0; found 229.9.

3.62 3-((2,5-Dioxopyrrolidin-1-yl)oxy)-3-(4-(2,5-dioxopyrrolidin-1-yl)phenyl)acrylonitrile (72):



Compound **72** was obtained as the only product during the synthesis of the target APN-NHS probe.

72a: 4-((4-(Cyanoethynyl)phenyl)amino)-4-oxobutanoic acid.

Same procedure as for the synthesis of **71a**. White solid, yield: 99%.

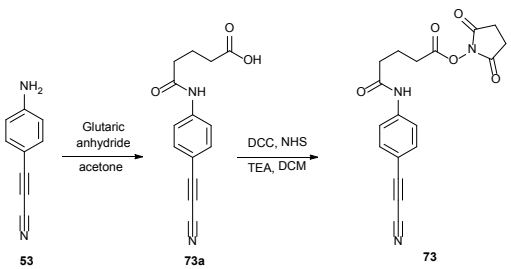
^{1}H NMR (400 MHz, DMSO-d_6) δ 10.41 (s, 1H), 7.65 - 7.83 (m, 4H), 2.61 (t, *J* = 6.30 Hz, 2H), 2.53 (t, *J* = 6.30 Hz, 2H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 174.2, 171.4, 143.5, 135.5, 119.3, 110.0, 106.2, 84.9, 62.2, 31.7, 29.1; ESI-MS: $C_{13}H_9N_2O_3^-$ [M-H]$^-$, 241.0; found 241.0.

72: 3-(4-(2,5-Dioxopyrrolidin-1-yl)phenyl)propiolonitrile.

To a cooled to 0˚C solution of **72a** (1 eq., 15.6 mg, 0.0644 mmol) in DCM (2 mL), DCC (1.02 eq., 13.6 mg, 0.0657 mmol) and TEA (1 eq., 6.52 mg, 0.00895 mL, 0.0644 mmol) were added. The resulting mixture was stirred for another 5 minutes at 0˚C, left to warm up to room temperature and NHS (1 eq., 7.41 mg, 0.0644 mmol) was added. The crude product was purified by flash chromatography to yield **72** (17.5 mg, 0.0515 mmol, 80%) as a white solid.

^{1}H NMR (400 MHz, ACETONE-d_6) δ 7.86 - 8.03 (m, *J* = 8.53 Hz, 2H), 7.54 - 7.66 (m, *J* = 8.53 Hz, 2H), 5.62 (s, 1H), 2.94 (s, 4H), 2.93 (s, 4H); ^{13}C NMR (101 MHz, ACETONE-d_6) δ 177.3, 170.9, 168.2, 137.0, 130.1, 128.2, 127.9, 114.7, 75.5, 28.9, 25.6; ESI-MS: $C_{17}H_{14}N_3O_5^+$ [M+H]$^+$, 340.1; found 340.1.

3.63 2,5-Dioxopyrrolidin-1-yl 5-((4-(cyanoethynyl)-phenyl)-amino)-5-oxopentanoate (73):



73a: 5-((4-(Cyanoethynyl)phenyl)amino)-5-oxopentanoic acid.

To a solution of **53** (1 eq., 200 mg, 1.41 mmol) in acetone (1 mL), glutaric anhydride (2 eq., 321 mg, 2.81 mmol) was added. The obtained solution was stirred for 24 hours at 25 ˚C. Acetone was evaporated under replaced pressure, the crude product was recrystallised from IPA-cyclohexane to give **73a** (324 mg, 1.27 mmol, 90 %) as a grey solid.

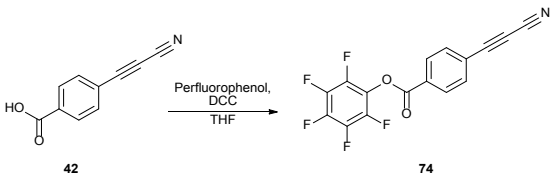
^{1}H NMR (400 MHz, DMSO-d_6) δ 10.21 (s, 1H), 8.15 (br. s., 1H), 7.60 (d, *J* = 8.72 Hz, 2H), 7.52 (d, *J* = 8.72 Hz, 2H), 2.52-2.62 (m, 4H), 2.22-2.32 (m, 2H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 170.0, 168.5, 140.9, 134.2, 119.0, 111.9, 105.4, 84.1, 63.3, 30.1, 29.0, 21.2; ESI-MS: $C_{14}H_{11}N_2O_3^-$ [M-H]$^-$, 255.1; found 255.1.

73: 2,5-Dioxopyrrolidin-1-yl 5-((4-(cyanoethynyl)phenyl)-amino)-5-oxopentanoate.

To a solution of **73a** (1 eq., 18 mg, 0.0702 mmol) in DCM (1 mL), DCC (1.02 eq., 14.8 mg, 0.0716 mmol) and TEA (1 eq., 6.52 mg, 0.00895 mL, 0.0644 mmol) were added. The resulting mixture was stirred for 5 minutes, NHS (1 eq., 8.08 mg, 0.0702 mmol) was added. The resulting solution stirred for another 2 hours at 25 ˚C. The crude product was purified by flash chromatography (cyclohexane-EtOAc) to give **73** (6.45 mg, 0.0183 mmol, 26%) as a white solid.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 8.27 (br. s., 1H), 7.64 (d, *J* = 8.78 Hz, 2H), 7.57 (d, *J* = 8.78 Hz, 2H), 2.94 (s, 4H), 2.74 (t, *J* = 6.53 Hz, 2H), 2.52 (t, *J* = 6.90 Hz, 2H), 2.23 (m, 2H); ^{13}C NMR (101 MHz, CHLOROFORM-d) d 170.3, 169.5, 168.2, 141.3, 134.6, 119.4, 112.4, 105.7, 83.2, 62.9, 35.6, 29.9, 25.7, 21.2; ESI-MS: $C_{18}H_{16}N_3O_5^+$ [M+H]$^+$, 353.1; found 353.2.

3.64 Perfluorophenyl 4-(cyanoethynyl)benzoate (74):

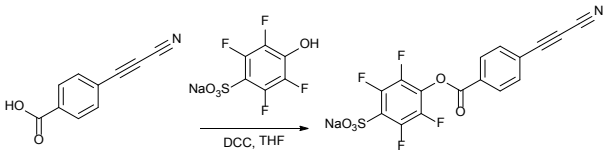


42 74

The solution of perfluorophenol (1 eq., 89.2 mg, 0.484 mmol) and **42** (1 eq., 82.9 mg, 0.484 mmol) in THF (5 mL) was cooled to 0 ˚C and DCC (1 eq., 100 mg, 0.484 mmol) was added to the mixture. The resulting solution was stirred at 25 ˚C for 14 hours, then filtered and washed with Et_2O. The filtrate was evaporated under reduced pressure to give **74** (98 mg, 0.290 mmol, 66%) as a white solid.

^{1}H NMR (400 MHz, DMSO-d_6) δ 8.20 - 8.35 (m, *J* = 8.28 Hz, 2H), 8.04 - 8.15 (m, *J* = 8.28 Hz, 2H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 161.7, 135.1, 134.5, 133.9, 131.2, 130.1, 129.3, 128.0, 123.3, 105.5, 82.1, 65.4; ESI-MS: $C_{16}H_5F_5NO_2$ [M+H]$^+$, 338.0; found 338.0.

3.65 Sodium 4-((4-(cyanoethynyl)benzoyl)oxy)-2,3,5,6-tetrafluorobenzenesulfonate (75):

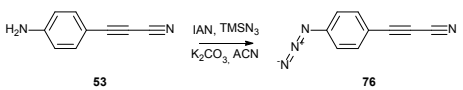


42 75

To the solution of **42** (1 eq., 54.2 mg, 0.317 mmol) and sodium 2,3,5,6-tetrafluoro-4-hydroxybenzene-1-sulfonate (1 eq., 84.9 mg, 0.317 mmol) in dry THF (0.792 mL) was added DCC (1 eq., 65.3 mg, 0.317 mmol). The resulting mixture was stirred at r.t. for 36h, then cooled to 0 ˚C, stirred for 1 hour, filtered and washed with 1 mL of dry DMF. The filtrate was diluted with 16 mL of Et_2O, stirred for 15 min for complete crystallisation and the precipitate was filtered to give **75** (72.5 mg, 0.172 mmol, 54 %) as a white solid.

^{1}H NMR (400 MHz, DMSO-d_6) δ 8.23 - 8.39 (m, 2H), 8.01 - 8.15 (m, 2H); ^{13}C NMR (101 MHz, DMSO-d_6) – not informative (even with ^{19}F decoupling, signals are too broad); ESI-MS: $C_{16}H_4F_4NO_5S^-$ [M-Na]$^-$, 397.9; found 398.0.

3.66 Sodium 4-((4-(cyanoethynyl)benzoyl)oxy)-2,3,5,6-tetrafluorobenzenesulfonate (76):

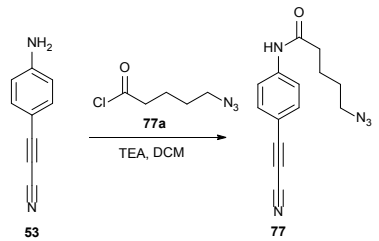


53 76

53 (1 eq., 151 mg, 1.07 mmol) was dissolved in acetonitrile (2.34 mL) in a 25 mL roundbottomed flask and cooled to 0°C in an ice bath. To this stirred mixture was added isoamyl nitrite (IAN, 1.5 eq., 215 µL, 1.6 mmol) followed by trimethylsilyl azide (1.2 eq., 147 mg, 0.168 mL, 1.28 mmol) dropwise. The resulting solution was stirred at 25 ˚C for 45 minutes. The reaction mixture was concentrated under vacuum and the crude product was resolubilised in EtOAc, washed with water, dried and evaporated under reduced pressure to give **76** (177 mg, 1.06 mmol, 99%).

^{1}H NMR (400 MHz, ACETONITRILE-d_3) δ 7.58 - 7.81 (m, *J* = 8.78 Hz, 2H), 7.11 - 7.26 (m, *J* = 8.78 Hz, 2H); ^{13}C NMR (101 MHz, ACETONITRILE-d_3) δ 144.9, 136.0, 120.4, 113.6, 105.9, 83.4, 62.9; GC-ESI-MS: $C_9H_5N_4^+$ [M+H]$^+$, 169.0; found 169.0.

3.67 5-Azido-*N*-(4-(cyanoethynyl)phenyl)pentanamide (77):



77a: 5-Azidopentanoyl chloride.

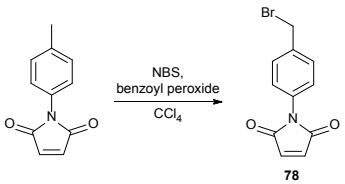
5-azidopentanoic acid (1 eq., 1.1 g, 6.99 mmol) was refluxed in $SOCl_2$ (10 eq., 5.1 mL, 69.9 mmol) for 30 minutes. Excess of $SOCl_2$ was removed *in vacuo* and the obtained crude solid was used in the next step without purification.

77: 5-Azido-*N*-(4-(cyanoethynyl)phenyl)pentanamide.

53 (1 eq., 16.1 mg, 0.113 mmol) and TEA (1.5 eq., 24 µL, 0.17 mmol) were dissolved in DCM (3 mL), cooled to -78 ˚C, and **77a** (1.1 eq., 20.1 mg, 0.125 mmol) was added to the reaction mixture that was then left to warm to room temperature while stirring for another 1 hour. The resulting mixture was washed with 1M HCl (2x1 mL), water (2 mL), dried over Na_2SO_4, and evaporated under reduced pressure to give crude product, which was purified by flash chromatography to give **77** (25.5 mg, 0.101 mmol, 89 %) as a grey solid.

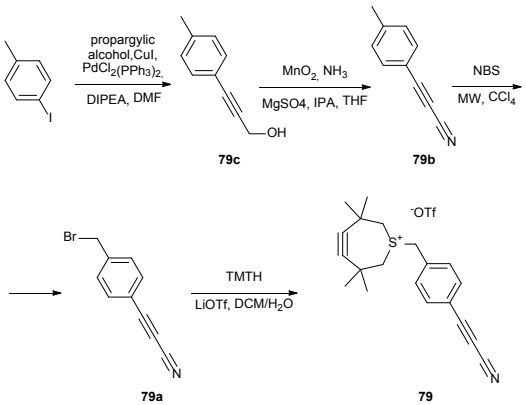
^{1}H NMR (400 MHz, METHANOL-d$_4$) δ 7.58 - 7.67 (m, *J* = 8.70 Hz, 2H), 7.40 - 7.58 (m, *J* = 8.70 Hz, 2H), 3.23 - 3.28 (m, 2H), 2.34 (t, *J* = 7.28 Hz, 2H), 1.61 - 1.72 (m, 2H), 1.49 - 1.61 (m, 2H); ^{13}C NMR (101 MHz, METHANOL-d$_4$) δ173.0, 144.0, 135.7, 134.3, 120.7, 113.0, 106.2, 84.5, 40.4, 34.4, 28.8, 24.0; ESI-MS: $C_{14}H_{14}N_5O^+$ [M+H]$^+$, 268.1; found 268.1.

3.68 1-(4-(Bromomethyl)phenyl)-*1H*-pyrrole-2,5-dione (78):



This compound was synthesised following the previously described procedure.[846]

3.69 1-[4-(Cyanoethynyl)benzyl]-3,3,6,6-tetramethyl-4,5-didehydro-2,3,6,7-tetrahydrothiepinium triflate (79):



79c: 3-(*p*-Tolyl)prop-2-yn-1-ol.

Synthesised using protocol A for Sonogashira coupling (Section D 2.9, page 77). Yellowish solid, yield: 88%.

^{1}H NMR (400 MHz, ACETONITRILE-d$_3$) δ 7.25 - 7.49 (m, *J* = 8.03 Hz, 2H), 7.04 - 7.25 (m, *J* = 8.03 Hz, 2H), 4.34 (d, *J* = 6.02 Hz, 2H), 3.31 (t, *J* = 6.02 Hz, 1H), 2.32 (s, 3H); ^{13}C NMR (101 MHz, ACETONITRILE-d$_3$) δ 139.8, 132.4, 130.3, 120.8, 88.8, 85.1, 51.2, 21.5; ESI-MS: $C_{10}H_{11}O^+$ [M+H]$^+$, 146.1; found 146.0.

79b: 3-(*p*-Tolyl)propiolonitrile.

The compound was obtained as the only product of the standard MnO_2 oxidation protocol (Section D 2.10, page 77). Reaction time: 3 hours. White solid, yield: 67%.

^{1}H NMR (400 MHz, METHANOL-d$_4$) δ 7.37 - 7.59 (m, *J* = 8.03 Hz, 2H), 7.02 - 7.31 (m, *J* = 8.03 Hz, 2H), 2.29 (s, 3H); ^{13}C NMR (101 MHz, METHANOL-d$_4$) δ 143.2, 133.3, 129.5, 114.0, 104.8, 83.2, 61.3, 20.4; ESI-MS: $C_{10}H_8N^+$ [M+H]$^+$, 141.1; found 141.0.

79a: 3-(4-(Bromomethyl)phenyl)propiolonitrile.

Degased solution of **79b** (1 eq., 68 mg, 0.482 mmol) in DCM (1 mL) was MW-irradiated (100˚C) for 5 minutes. The reaction mixture was evaporated under replaced pressure, the crude was purified by preparative HPLC to give **79a** (42.4 mg, 0.193 mmol, 40 %) as a yellowish solid.

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.56 - 7.71 (m, *J* = 8.28 Hz, 2H), 7.40 - 7.49 (m, *J* = 8.28 Hz, 2H), 4.48 (s, 2H); ^{13}C NMR (101 MHz, CHLOROFORM-d) δ 141.8, 133.9, 129.5, 117.5, 105.3, 82.3, 63.7, 31.8; GC-ESI-MS: $C_{10}H_7BrN^+$ [M+H]$^+$, 219.0; found 219.0.

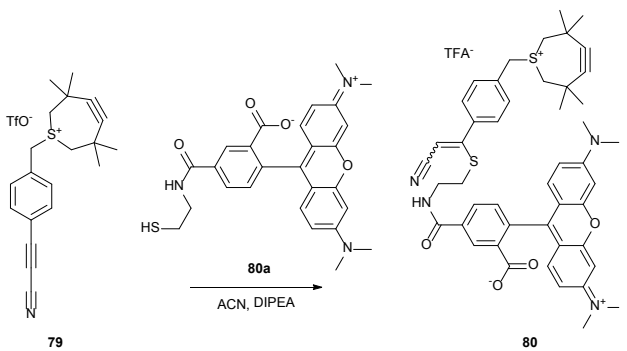
79: 1-[4-(Cyanoethynyl)benzyl]-3,3,6,6-tetramethyl-4,5-didehydro-2,3,6,7-tetrahydrothiepinium triflate.

To a degassed solution of **79a** (1 eq., 43.7 mg, 0.199 mmol) and TMTH (1.29 eq., 43 mg, 0.255 mmol; synthesised following previously described procedure[831]) in DCM (1.34 mL), a solution of LiOTf (11.6 eq., 360 mg, 2.31 mmol) in distilled and degased H_2O (0.668 mL) was added. The obtained biphasic

mixture was vigorously stirred for 5 days at 25 °C (degasing once per day). Two phases were separated, the organic one was washed with DCM (5x2 mL). United organic fractions were evaporated and the crude was purified by HPLC to give **79** (46.9 mg, 0.111 mmol, 56 %) as colourless oil (crystallizes slowly at 0 °C to yield a white solid).

^{1}H NMR (400 MHz, CHLOROFORM-d) δ 7.65 - 7.73 (m, *J* = 8.03 Hz, 2H), 7.56 - 7.65 (m, *J* = 8.03 Hz, 2H), 5.07 (s, 2H), 4.12 (d, *J* = 12.30 Hz, 2H), 3.72 (d, *J* = 12.30 Hz, 2H), 1.36 (s, 6H), 1.30 (s, 6H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 135.1, 133.6, 131.9, 117.8, 106.4, 105.8, 83.3, 63.6, 60.1, 43.2, 34.6, 26.4, 25.4; HR-ESI-MS: $C_{20}H_{22}NS^+$ $[M]^+$, 308.1; found 308.1.

3.70 1-({4-[1-{[2-({3-Carboxylato-4-[6-(dimethylamino)-3-(dimethyliminiumyl)-3H-xanthen-9-yl]phenyl}formamido)ethyl]sulfanyl}-2-cyanoeth-1-en-1-yl]phenyl}methyl)-3,3,6,6-tetramethyl-1-thiacyclohept-4-yn-1-ium trifluoroacetate (80, TAMRA-APN-TMTI):



80a: 2-(6-(Dimethylamino)-3-(dimethyliminio)-*3H*-xanthen-9-yl)-5-((2-mercaptoethyl)carbamoyl)benzoate.

To a solution of TAMRA-5'-COOH (**83**, 1 eq., 68.3 mg, 0.159 mmol) in DMF (0.228 mL), HATU (1 eq., 60.3 mg, 0.159 mmol), DIPEA (6 eq., 123 mg, 0.157 mL, 0.952 mmol) and cystamine dichloride (5 eq., 178 mg, 0.793 mmol) were subsequently added; the obtained solution mass was stirred overnight. A solution of DTT (5 eq., 122 mg, 0.118 mL, 0.793 mmol) in DCM (0.911 mL) was added to the reaction mass, the stirring continued for 2 hours. Solvents were evaporated; the obtained crude mass was purified by HPLC to yield **80a** (33.5 mg, 0.0555 mmol, 35 %) as a dark-violet solid.

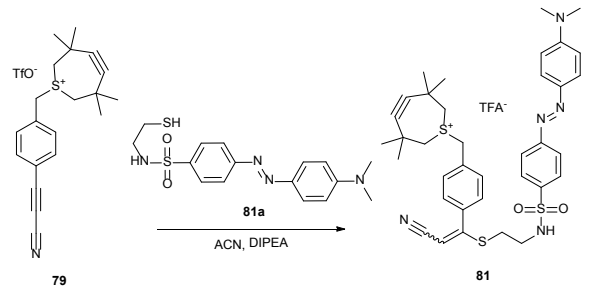
^{1}H NMR (400 MHz, DMSO-d_6) δ 9.06 (t, *J* = 5.4 Hz, 1H, 8.70 (d, *J* = 1.8 Hz, 1H), 8.30 (dd, *J* = 1.8, 8.0 Hz, 1H), 7.59 (d, *J* = 8.0 Hz, 1H), 7.08-7.02 (m, 4H), 6.95 (s, 2H), 3.52-3.42 (m, 2H), 3.26 (s, 12H), 2.72 (dt, *J* = 6.8, 8.0 Hz, 2H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 166.0, 164.7, 156.8, 156.6, 135.9, 131.2, 130.6, 114.6, 96.3, 42.9, 40.5, 23.3; HR-ESI-MS: $C_{27}H_{27}N_3O_4S$: 489.1722; found 489.1723.

80: 1-({4-[(1Z)-1-{[2-({3-Carboxylato-4-[6-(dimethylamino)-3-(dimethyliminiumyl)-3H-xanthen-9-yl]phenyl}formamido)ethyl]sulfanyl}-2-cyanoeth-1-en-1-yl]phenyl}methyl)-3,3,6,6-tetramethyl-1-thiacyclohept-4-yn-1-ium trifluoroacetate (TAMRA-APN-TMTI).

A solution of **80a** (1 eq., 6.74 mg, 0.016 mmol) in ACN (1 mL) was mixed with a solution of **80a** (1 eq., 9.64 mg, 0.016 mmol) in DMF (1 mL). DIPEA (5 eq., 132 µL, 0.08 mmol) was then added and obtained reaction mass was injected into HPLC after 5 minutes of reaction to yield **80** (11.9 mg, 0.0149 mmol, 93 %) as a dark-violet solid.

^{1}H NMR (400 MHz, DMSO-d_6) δ 8.98 (t, *J* = 5.40 Hz, 1H), 8.30 (d, *J* = 8.28 Hz, 1H), 8.10 - 8.20 (m, 1H), 7.85 (s, 1H), 7.57 - 7.69 (m, 4H), 6.97 - 7.15 (m, 5H), 6.07 (s, 1H), 4.85 (s, 2H), 2.81 - 2.90 (m, 4H), 3.28 (br. s., 16H), 1.25 (s, 6H), 1.05 (s, 6H); ^{13}C NMR (101 MHz, DMSO-d_6) – not informative (low resolved signals); HR-ESI-MS: $C_{47}H_{49}N_4O_4S_2^+$, 797.31897; found 797.32739.

3.71 1-({4-[2-Cyano-1-[(2-{4-[*(E)*-2-[4-(dimethylamino)phenyl]diazen-1-yl]benzenesulfonamido}ethyl)sulfanyl]eth-1-en-1-yl]phenyl}methyl)-3,3,6,6-tetramethyl-1-thiacyclohept-4-yn-1-ium (81, BHQ2-APN-TMTI):



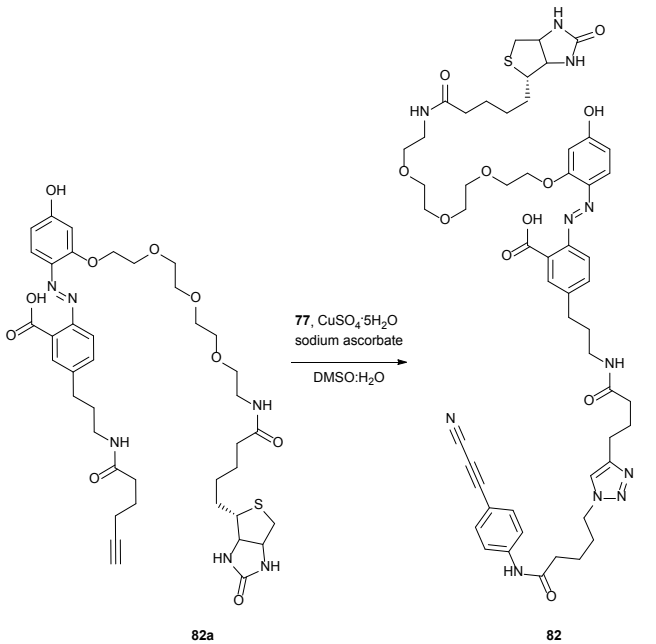
81a: *(E)*-4-((4-(Dimethylamino)phenyl)diazenyl)-*N*-(2-mercaptoethyl)benzenesulfonamide.

To a cooled to 0°C solution of Dabsyl chloride (1 eq., 100 mg, 0.309 mmol) in dry ACN (3 mL), TEA (7 eq., 218 mg, 0.3 mL, 2.16 mmol) and cystamine dihydrochloride (5 eq., 347 mg, 1.54 mmol) were subsequently added. After 2 hours of stirring, DTT (6 eq., 285 mg, 0.275 mL, 1.85 mmol) was added to the reaction mass. The obtained solution was stirred for another 2 hours, evaporated and the obtained crude product was purified by flash chromatography (cyclohexane-EtOAc) to yield **81a** (105.9 mg, 94%) as an orange solid.

81: 1-({4-[(2-Cyano-1-[(2-{4-[*(E)*-2-[4-(dimethylamino)phenyl]diazen-1-yl]benzenesulfonamido}ethyl)sulfanyl]eth-1-en-1-yl]phenyl}methyl)-3,3,6,6-tetramethyl-1-thiacyclohept-4-yn-1-ium (BHQ2-APN-TMTI).

The same procedure as for the synthesis of the **80**. Yield: 94%. ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.03 (t, *J* = 4.89 Hz, 1H), 7.91 (d, *J* = 8.53 Hz, 2H), 7.80 - 7.87 (m, *J* = 9.04 Hz, 2H), 7.72 - 7.79 (m, *J* = 8.53 Hz, 2H), 7.68 (s, 4H), 6.87 (d, *J* = 9.04 Hz, 2H), 6.08 (s, 1H), 4.86 (s, 2H), 3.92 (d, *J* = 12.05 Hz, 2H), 3.84 (d, *J* = 12.30 Hz, 2H), 3.10 (s, 6H), 2.71 - 2.87 (m, 4H), 1.32 (s, 6H), 1.17 (s, 6H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 160.2, 158.6, 158.3, 155.1, 153.7, 143.1, 140.3, 136.8, 131.9, 131.3, 129.5, 128.2, 125.9, 122.8, 117.2, 112.1, 106.4, 99.4, 60.0, 43.3, 42.9, 34.5, 26.4, 25.3; HR-ESI-MS: $C_{36}H_{42}N_5O_2S_3^+$, 672.24951; found 672.25042.

3.72 5-(3-{4-[1-(4-{[4-(2-Cyanoeth-1-yn-1-yl)phenyl]carbamoyl}butyl)-*1H*-1,2,3-triazol-4-yl]butanamido}propyl)-2-[*(E)*-2-[4-hydroxy-2-(2-{2-[2-(2-{5-[(4S)-2-oxo-hexahydro-*1H*-thieno[3,4-d]imidazolidin-4-yl]pentanamido}ethoxy)ethoxy]ethoxy}ethoxy)phenyl]diazen-1-yl]benzoic (82, APN-HAZA-biotin):



82a: 2-(6-(Dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)-5-((2-mercaptoethyl)carbamoyl)benzoate.

This compound was synthesised following the previously reported protocol.[847]

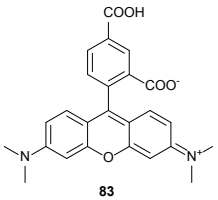
82: 5-(3-{4-[1-(4-{[4-(2-Cyanoeth-1-yn-1-yl)phenyl]carbamoyl}butyl)-*1H*-1,2,3-triazol-4-yl]butanamido}propyl)-2-[*(E)*-2-[4-hydroxy-2-(2-{2-[2-(2-{5-[(4S)-2-oxo-hexahydro-*1H*-thieno[3,4-d]imidazolidin-4-yl]pentanamido}ethoxy)ethoxy]ethoxy}ethoxy)phenyl]diazen-1-yl]benzoic (APN-HAZA-biotin).

To a solution of **82a** (1 eq., 10 mg, 0.0123 mmol) and **77** (1 eq., 3.12 mg, 0.0123 mmol) in DMSO (0.472 mL), solution of sodium ascorbate (10 eq., 24.4 mg, 0.123 mmol) and $CuSO_4 \cdot 5H_2O$ (5 eq., 15.4 mg, 0.0617 mmol) in water was added. The resulting mixture was degassed and stirred overnight at 25 °C. The resulting mixture was directly purified by HPLC to give **82** (8.3 mg, 0.0078 mmol, 63 %) as a yellow solid.

^{1}H NMR (400 MHz, METHANOL-d_4) δ 7.78 (br. s., 2H), 7.72 (d, *J* = 8.28 Hz, 1H), 7.67 (s, 1H), 7.58 (d, *J* = 8.78 Hz, 2H), 7.50 (d, *J* = 8.78 Hz, 2H), 7.29 (d, *J* = 8.28 Hz, 1H), 7.25 (d, *J* = 8.53 Hz, 1H), 6.29 (d, *J* = 7.78 Hz, 1H), 3.70 - 3.81 (m, 8H), 3.62 (d, *J* = 4.77 Hz, 2H), 3.58 (d, *J* = 5.02 Hz, 2H), 3.48 - 3.53 (m, 2H), 3.41 - 3.48 (m, 2H), 3.01 - 3.07 (m, 6H), 2.89 - 3.00 (m, 10H), 2.78 (dd, *J* = 4.89, 12.93 Hz, 1H), 2.52 - 2.67 (m,

8H), 2.33 (t, *J* = 7.28 Hz, 2H), 2.09 - 2.19 (m, 2H), 1.81 - 1.91 (m, 4H), 1.65 - 1.78 (m, 2H), 1.52 - 1.63 (m, 2H), 1.42 - 1.52 (m, 1H), 1.23 - 1.31 (m, 1H); HR-ESI-MS: $C_{56}H_{67}N_{11}O_{11}S$, 1077.47422; found 1077.45931.

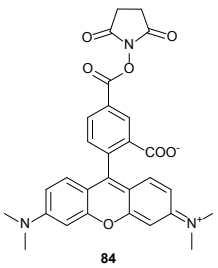
3.73 5-Carboxytetramethylrhodamine (83, TAMRA-5'-COOH):



83

This compound was synthesised following a reported protocol.[848]

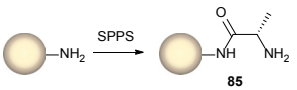
3.74 5-Carboxytetramethylrhodamine (84, TAMRA-5'-NHS Ester):



84

This compound is commercially available.

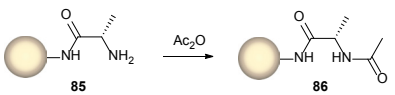
3.75 TentaGel®-Alanine beads (85):



85

Synthesised using standard SPPS protocol (see Section D 2.7.1, page 76).

3.76 TentaGel®-Ac-Alanine beads (86):



85 **86**

Synthesised from **85** *via* acetylation (see Section D 2.7.1, page 76).

E. REFERENCES

1. Zalipsky, S. Functionalized poly (ethylene glycols) for preparation of biologically relevant conjugates. *Bioconjugate Chem.* **6,** 150–165 (1995).
2. Harris, J. M. & Chess, R. B. Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug. Discov.* **2,** 214–221 (2003).
3. Baker, D. P. *et al.* N-terminally PEGylated human interferon-beta-1a with improved pharmacokinetic properties and in vivo efficacy in a melanoma angiogenesis model. *Bioconjugate Chem.* **17,** 179–188 (2006).
4. Bell, S. J. *et al.* Enhanced circulating half-life and antitumor activity of a site-specific pegylated interferon-alpha protein therapeutic. *Bioconjugate Chem.* **19,** 299–305 (2008).
5. Griffin, B. A. Specific Covalent Labeling of Recombinant Protein Molecules Inside Live Cells. *Science* **281,** 269–272 (1998).
6. O'Hare, H. M., Johnsson, K. & Gautier, A. Chemical probes shed light on protein function. *Curr. Opin. Struct. Biol.* **17,** 488–494 (2007).
7. Michalet, X., Weiss, S. & Jäger, M. Single-Molecule Fluorescence Studies of Protein Folding and Conformational Dynamics. *Macromolecules* **106,** 1785–1813 (2006).
8. Royer, C. A. Probing protein folding and conformational transitions with fluorescence. *Chem. Rev.* **106,** 1769–1784 (2006).
9. Bagshaw, C. R. & Cherny, D. Blinking fluorophores: what do they tell us about protein dynamics? *Biochem. Soc. Trans.* **34,** 979–982 (2006).
10. Cravatt, B. F., Wright, A. T. & Kozarich, J. W. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* **77,** 383–414 (2008).
11. Fernández-Suárez, M. & Ting, A. Y. Fluorescent probes for super-resolution imaging in living cells. *Nat. Rev. Mol. Cell Biol.* **9,** 929–943 (2008).
12. Wang, Q., Lin, T., Tang, L., Johnson, J. E. & Finn, M. G. Icosahedral Virus Particles as Addressable Nanoscale Building Blocks. *Angew. Chem. Int. Ed. (English)* **41,** 459–462 (2002).
13. Smith, J. C. *et al.* Nanopatterning the chemospecific immobilisation of cowpea mosaic virus capsid. *Nano Lett.* **3,** 883–886 (2003).
14. Löwik, D. W. P. M. *et al.* A Highly Ordered Material from Magnetically Aligned Peptide Amphiphile Nanofiber Assemblies. *Adv. Mater.* **19,** 1191–1195 (2007).
15. Christman, K. L., Broyer, R. M., Tolstyka, Z. P. & Maynard, H. D. Site-specific protein immobilisation through N-terminal oxime linkages. *J. Mater. Chem.* **17,** 2021–2027 (2007).
16. Abedin, M. J., Liepold, L., Suci, P., Young, M. & Douglas, T. Synthesis of a Cross-Linked Branched Polymer Network in the Interior of a Protein Cage. *J. Am. Chem. Soc.* **131,** 4346–4354 (2009).
17. Witus, L. S. & Francis, M. B. Using synthetically modified proteins to make new materials. *Acc. Chem. Res.* (2011).
18. Chen, M.-L. *et al.* Fabrication of an Oriented Fc-Fused Lectin Microarray through Boronate Formation. *Angew. Chem. Int. Ed. (English)* **47,** 8627–8630 (2008).
19. Hucknall, A. *et al.* Simple fabrication of antibody microarrays on nonfouling polymer brushes with femtomolar sensitivity for protein analytes in serum and blood. *Adv. Mater.* **21,** 1968–1971 (2009).
20. Huang, Y.-L. & Wu, C.-Y. Carbohydrate-based vaccines: challenges and opportunities. *Expert Rev. Vaccines* **9,** 1257–1274 (2010).
21. Aggarwal, S. What's fueling the biotech engine—2009–2010. *Nat. Biotechnol.* **28,** 1165–1171 (2010).
22. Mullard, A. Maturing antibody–drug conjugate pipeline hits 30. *Nat. Rev. Drug. Discov.* **12,** 329–332 (2013).
23. Huang, Y.-L. *et al.* Carbohydrate-based vaccines with a glycolipid adjuvant for breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 2517–2522 (2013).
24. Adamo, R. *et al.* Synthetically defined glycoprotein vaccines: current status and future directions. *Chem. Sci.* **4,** 2995–3008 (2013).
25. Gruttner, C., Muller, K. & Teller, J. Comparison of Strain-Promoted Alkyne-Azide Cycloaddition With Established Methods for Conjugation of Biomolecules to Magnetic Nanoparticles. *Magnetics, IEEE Transactions on* **49,** 172–176 (2013).
26. Sinz, A. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J. Mass Spectrom.* **38,** 1225–1237 (2003).
27. Hinner, M. J. & Johnsson, K. How to obtain labeled proteins and what to do with them. *Curr. Opin. Chem. Biol.* **21,** 766–776 (2010).
28. Hermanson, G. T. Bioconjugate Techniques, Second Edition. (Academic Press 2008)
29. Westheimer, F. H. & Schmidt, D. E. pK of the lysine amino group at the active site of acetoacetate decarboxylase. *Biochemistry* **10,** 1249–1253 (1971).
30. Anderson, G. W., Zimmerman, J. E. & Callahan, F. M. The use of esters of N-hydroxysuccinimide in peptide synthesis. *J. Am. Chem. Soc.* **86,** 1839–1842 (1964).
31. Lowe, A. The chemistry of isocyanates. *Proc. R. Soc. Med.* **63,** 367–368 (1970).
32. Coons, A. H., Creech, H. J., Jones, R. N. & Berliner, E. The demonstration of pneumococcal antigen in tissues by the use of fluorescent antibody. *J. Immunol.* **45,** 159–170 (1942).
33. Annunziato, M. E., Patel, U. S., Ranade, M. & Palumbo, P. S. p-Maleimidophenyl isocyanate: a novel heterobifunctional linker for hydroxyl to thiol coupling. *Bioconjugate Chem.* **4,** 212–218 (1993).
34. Creech, H. J. & Jones, R. N. *J. Am. Chem. Soc.* **63,** 1670–1673 (1941).
35. Goodfriend, L. & Sehon, A. H. Preparation of an estrone-protein conjugate. *Can. J. Biochem.* **36,** 1177–1184 (1958).
36. Todrick, A. & Walker, E. A note on the combination of cysteine with allyl isothiocyanate. *Biochem. J.* **31,** 297–298 (1937).
37. Edman, P. Method for determination of the amino acid sequence in peptides. *Acta Chem. Scand.* **4,** 283–293 (1950).
38. Podhradský, D., Drobnica, Ľ. & Kristian, P. Reactions of cysteine, its derivatives, glutathione, coenzyme A, and dihydrolipoic acid with isothiocyanates. *Experientia* **35,** 154–155 (1979).
39. Seiwald, R. J., Riggs, J. L., Burckhalter, J. H., Downs, C. M. & Metcalf, T. G. Isothiocyanate compounds as fluorescent labeling agents for immune serum. *Am. J. Pathol.* **34,** 1081–1097 (1958).
40. Haugland, R. P., Spence, M. T. & Johnson, I. D. Handbook of fluorescent probes and research chemicals. **6,** (1996).
41. Tuls, J., Geren, L. & Millett, F. Fluorescein isothiocyanate specifically modifies lysine 338 of cytochrome P-450scc and inhibits adrenodoxin binding. *J. Biol. Chem.* **264,** 16421–16425 (1989).
42. Burtnick, L. D. Modification of actin with fluorescein isothiocyanate. *Biochim. Biophys. Acta* **791,** 57–62 (1984).
43. Miki, M. The recovery of the polymerizability of Lys-61-labelled actin by the addition of phalloidin. *Eur. J. Biochem.* **164,** 229–235 (1987).
44. Miki, M. Interaction of Lys-61 labeled actin with myosin subfragment-1 and the regulatory proteins. *J. Biochem.* **106,** 651–655 (1989).
45. Bellelli, A. *et al.* Binding and internalisation of ricin labelled with fluorescein isothiocyanate. *Biochem. Biophys. Res. Commun.* **169,** 602–609 (1990).
46. Dedmon, R. E., Holmes, A. W. & Deinhardt, F. Preparation of fluorescein isothiocyanate-labeled γ-globulin by dialysis, gel filtration, and ionexchange chromatography in combination. *J. Bacteriol.* **89,** 734–739 (1965).
47. Amante, L., Ancona, A. & Forni, L. The conjugation of immunoglobulins with tetramethylrhodamine isothiocyanate a comparison between the amorphous and the crystalline fluorochrome. *J.*

Immunol. Methods. **1,** 289–301 (1972).
48. Barbet, A. F. & McGuire, T. C. Crossreacting determinants in variant-specific surface antigens of African trypanosomes. *Proc. Natl. Acad. Sci. U. S. A.* **75,** 1989–1993 (1978).
49. Schultzberg, M. *et al.* Enkephalin-like immunoreactivity in gland cells and nerve terminals of the adrenal medulla. *Neuroscience* **3,** 1169–1186 (1978).
50. Carney, W. P. *et al.* Analysis of T lymphocyte subsets in cytomegalovirus mononucleosis. *J. Immunol.* **126,** 2114–2116 (1981).
51. Varadarajan, A., Sharkey, R. M., Goldenberg, D. M. & Hawthorne, M. F. Conjugation of phenyl isothiocyanate derivatives of carborane to antitumor antibody and in vivo localisation of conjugates in nude mice. *Bioconjugate Chem.* **2,** 102–110 (1991).
52. Pei, R. *et al.* A monospecific HLA-B27 fluorescein isothiocyanate-conjugated monoclonal antibody for rapid, simple and accurate HLA-B27 typing. *Tissue Antigens* **41,** 200–203 (1993).
53. Edidin, M., Zagyansky, Y. & Lardner, T. Measurement of membrane protein lateral diffusion in single cells. *Science* **191,** 466–468 (2003).
54. Huang, K. K. & Cantor, C. R. C. Surface topography of the 30 s Escherichia coli ribosomal subunit: reactivity towards fluorescein isothiocyanate. *J. Mol. Biol.* **67,** 265–275 (1972).
55. Blose, S. H. Distribution of flourescently labeled α-actinin in living and fixed fibroblasts. *J. Cell Biol.* **86,** 608–615 (1980).
56. Ohkuma, S., Moriyama, Y. & Takano, T. Identification and characterisation of a proton pump on lysosomes by fluorescein-isothiocyanate-dextran fluorescence. *Proc. Natl. Acad. Sci. U. S. A.* **79,** 2758–2762 (1982).
57. Morrison, S. J., Hemmati, H. D., Wandycz, A. M. & Weissman, I. L. The purification and characterisation of fetal liver hematopoietic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 10302–10306 (1995).
58. Gabel, C. A., Eddy, E. M. & Shapiro, B. M. Regional differentiation of the sperm surface as studied with 125I-diiodofluorescein isothiocyanate, an impermeant reagent that allows isolation of the labeled components. *J. Cell Biol.* **82,** 742–754 (1979).
59. Rosa-Neto, P. *et al.* [18F]SiFA-isothiocyanate: A New Highly Effective Radioactive Labeling Agent for Lysine-Containing Proteins. *Chem. Eur. J. of Chem. Bio.* **10,** 1321–1324 (2009).
60. Linder, K. E. *et al.* Technetium labeling of monoclonal antibodies with functionalized BATOs. 1. TcCl (DMG) 3PITC [phenyl isothiocyanate]. *Bioconjugate Chem.* **2,** 160–170 (1991).
61. Meares, C. F. *et al.* Conjugation of antibodies with bifunctional chelating agents: Isothiocyanate and bromoacetamide reagents, methods of analysis, and subsequent addition of metal ions. *Anal. Biochem.* **142,** 68–78 (1984).
62. Mirzadeh, S. S., Brechbiel, M. W. M., Atcher, R. W. R. & Gansow, O. A. O. Radiometal labeling of immunoproteins: covalent linkage of 2-(4-isothiocyanatobenzyl)diethylenetriaminepentaacetic acid ligands to immunoglobulin. *Bioconjugate Chem.* **1,** 59–65 (1989).
63. Wilbur, D. S. *et al.* Trifunctional Conjugation Reagents. Reagents That Contain a Biotin and a Radiometal Chelation Moiety for Application to Extracorporeal Affinity Adsorption of Radiolabeled Antibodies. *Bioconjugate Chem.* **13,** 1079–1092 (2002).
64. Shi, H. *et al.* Rhodamine B isothiocyanate doped silica-coated fluorescent nanoparticles (RBITC-DSFNPs)–based bioprobes conjugated to Annexin V for apoptosis detection and imaging. *Nanomed. Nanotechnol.* **3,** 266–272 (2007).
65. Roederer, M. Conjugation of monoclonal antibodies (August, 2004). at <http://www.drmr.com/abcon/index.html>
66. Butcher, E. C. E. & Weissman, I. L. I. Direct fluorescent labeling of cells with fluorescein or rhodamine isothiocyanate. I. Technical aspects. *J. Immunol. Methods.* **37,** 97–108 (1979).
67. Landel, A. M. Stability studies on fluorescein isothiocyanate--bovine serum albumin conjugate. *Anal. Biochem.* **73,** 280–289 (1976).
68. Wilbur, D. S. *et al.* Reagents for astatination of Biomolecules. 6. An intact antibody conjugated with a maleimido-closo-decaborate (2-) reagent *via* sulfhydryl groups had considerably higher kidney concentrations than the same antibody conjugated with an isothiocyanato-closo-decaborate (2-) reagent *via* lysine amines. *Bioconjugate Chem.* **23,** 409–420 (2012).
69. Banks, P. R. & Paquette, D. M. Comparison of Three Common Amine Reactive Fluorescent Probes Used for Conjugation to Biomolecules by Capillary Zone Electrophoresis. *Bioconjugate Chem.* **6,** 447–458 (1995).
70. Smith, M. B. & March, J. *March's Advanced Organic Chemistry*. (John Wiley & Sons, 2007).
71. Smith, G. P. Kinetics of Amine Modification of Proteins. *Bioconjugate Chem.* **17,** 501–506 (2006).
72. Mäkelä, O. Assay of anti-hapten antibody with the aid of hapten-coupled bacteriophage. *Immunology* **10,** 81–86 (1966).
73. Becker, J. M., Wilchek, M. & Katchalski, E. Irreversible inhibition of biotin transport in yeast by biotinyl-p-nitrophenyl ester. *Proc. Natl. Acad. Sci. U. S. A.* **68,** 2604–2607 (1971).
74. Brinkley, M. A brief survey of methods for preparing protein conjugates with dyes, haptens, and cross-linking reagents. *Bioconjugate Chem.* **3,** 2–13 (1991).
75. Anderson, G. W., Zimmerman, J. E. & Callahan, F. M. N-Hydroxysuccinimide Esters in Peptide Synthesis. *J. Am. Chem. Soc.* **85,** 3039–3039 (1963).
76. Anderson, G. W., Callahan, F. M. & Zimmerman, J. E. Synthesis of N-hydroxysuccinimide esters of acyl peptides by the mixed anhydride method. *J. Am. Chem. Soc.* **89,** 178–178 (1967).
77. Yang, W.-C., Mirzaei, H., Liu, X. & Regnier, F. E. Enhancement of Amino Acid Detection and Quantification by Electrospray Ionisation Mass Spectrometry. *Anal. Chem.* **78,** 4702–4708 (2006).
78. Regnier, F. E. & Julka, S. Primary amine coding as a path to comparative proteomics. *Proteomics* **6,** 3968–3979 (2006).
79. Ishikawa, Y., Yamamoto, Y., Otsubo, M., Theg, S. M. & Tamura, N. Chemical modification of amine groups on PS II protein(s) retards photoassembly of the photosynthetic water-oxidizing complex. *Biochemistry* **41,** 1972–1980 (2002).
80. Holmquist, B., Blumberg, S. & Vallee, B. L. Superactivation of neutral proteases: acylation with N-hydroxysuccinimide esters. *Biochemistry* **15,** 4675–4680 (1976).
81. Blumberg, S. & Vallee, B. L. Superactivation of thermolysin by acylation with amino acid N-hydroxysuccinimide esters. *Biochemistry* **14,** 2410–2419 (1975).
82. Miller, B. T., Collins, T. J., Nagle, G. T. & Kurosky, A. The occurrence of O-acylation during biotinylation of gonadotropin-releasing hormone and analogs. Evidence for a reactive serine. *J. Biol. Chem.* **267,** 5060–5069 (1992).
83. Miller, B. T. & Kurosky, A. Elevated Intrinsic Reactivity of Seryl Hydroxyl Groups within the Linear Peptide Triads His-Xaa-Ser or Ser-Xaa-His. *Biochem. Biophys. Res. Commun.* **196,** 461–467 (1993).
84. Miller, B. T. Acylation of peptide hydroxyl groups with the Bolton-Hunter reagent. *Biochem. Biophys. Res. Commun.* **218,** 377–382 (1996).
85. Miller, B. T., Collins, T. J., Rogers, M. E. & Kurosky, A. Peptide Biotinylation with Amine-Reactive Esters: Differential Side Chain Reactivity. *Peptides* **18,** 1585–1595 (1997).
86. Mädler, S., Bich, C., Touboul, D. & Zenobi, R. Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. *J. Mass Spectrom.* **44,** 694–706 (2009).
87. Smith, J. S., Miller, B. T., Knock, S. L. & Kurosky, A. Biotinylated peptides/proteins. *Anal. Biochem.* **197,** 247–253 (1991).
88. Cuatrecasas, P. & Parikh, I. Adsorbents for affinity chromatography. Use of N-hydroxysuccinimide esters of agarose. *Biochemistry* **11,** 2291–2299 (1972).
89. Lomant, A. J. & Fairbanks, G. Chemical probes of extended biological structures: Synthesis and properties of the cleavable protein cross-linking reagent [35S]dithiobis(succinimidyl propionate). *J. Mol. Biol.* **104,** 243–261 (1976).

90. Staros, J. V., Wright, R. W. & Swingle, D. M. Enhancement by N-hydroxysulfosuccinimide of water-soluble carbodiimide-mediated coupling reactions. *Anal. Biochem.* **156,** 220–222 (1986).
91. Abello, N., Kerstjens, H. A. M., Postma, D. S. & Bischoff, R. Selective Acylation of Primary Amines in Peptides and Proteins. *J. Proteome Res.* **6,** 4770–4776 (2007).
92. Yem, A. W. *et al.* Biotinylation of reactive amino groups in native recombinant human interleukin-1 beta. *J. Biol. Chem.* **264,** 17691–17697 (1989).
93. Lombardi, V. C. & Schooley, D. A. A method for selective conjugation of an analyte to enzymes without unwanted enzyme-enzyme cross-linking. *Anal. Biochem.* **331,** 40–45 (2004).
94. Morpurgo, M., Bayer, E. A. & Wilchek, M. N-hydroxysuccinimide carbonates and carbamates are useful reactive reagents for coupling ligands to lysines on proteins. *J. Biochem. Biophys. Methods* **38,** 17–28 (1999).
95. Cooper, M. *et al.* Cy3B: improving the performance of cyanine dyes. *J. Fluoresc.* **14,** 145–150 (2004).
96. Deng, Y. *et al.* Role of Lysine 411 in Substrate Carboxyl Group Binding to the Human Reduced Folate Carrier, as Determined by Site-Directed Mutagenesis and Affinity Inhibition. *Mol. Pharmacol.* **73,** 1274–1281 (2008).
97. Kahne, D. & Still, W. C. Hydrolysis of a peptide bond in neutral water. *J. Am. Chem. Soc.* **110,** 7529–7534 (1988).
98. Becker, J. M. & Wilchek, M. Inactivation by avidin of biotin-modified bacteriophage. *Biochim. Biophys. Acta* **264,** 165–170 (1972).
99. Trester-Zedlitz, M. *et al.* A modular cross-linking approach for exploring protein interactions. *J. Am. Chem. Soc.* **125,** 2416–2425 (2003).
100. Kluger, R. & Alagic, A. Chemical cross-linking and protein–protein interactions—a review with illustrative protocols. *Bioinorg. Chem.* **32,** 451–472 (2004).
101. Trakselis, M. A., Alley, S. C. & Ishmael, F. T. Identification and Mapping of Protein−Protein Interactions by a Combination of Cross-Linking, Cleavage, and Proteomics. *Bioconjugate Chem.* **16,** 741–750 (2005).
102. Tang, X., Munske, G. R., Siems, W. F. & Bruce, J. E. Mass Spectrometry Identifiable Cross-Linking Strategy for Studying Protein−Protein Interactions. *Anal. Chem.* **77,** 311–318 (2005).
103. Chavez, J. D., Liu, N. L. & Bruce, J. E. Quantification of Protein−Protein Interactions with Chemical Cross-Linking and Mass Spectrometry. *J. Proteome Res.* **10,** 1528–1537 (2011).
104. Müller, M. Q., Dreiocker, F., Ihling, C. H., Schäfer, M. & Sinz, A. Cleavable Cross-Linker for Protein Structure Analysis: Reliable Identification of Cross-Linking Products by Tandem MS. *Anal. Chem.* **82,** 6958–6968 (2010).
105. Back, J. W., de Jong, L., Muijsers, A. O. & de Koster, C. G. Chemical Cross-linking and Mass Spectrometry for Protein Structural Modeling. *J. Mol. Biol.* **331,** 303–313 (2003).
106. Young, M. M. *et al.* High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 5802–5806 (2000).
107. Avrameas, S. Coupling of enzymes to proteins with glutaraldehyde. Use of the conjugates for the detection of antigens and antibodies. *Immunochemistry* **6,** 43–52 (1969).
108. Erlanger, B. F. Principles and methods for the preparation of drug protein conjugates for immunological studies. *Pharmacol. Rev.* **25,** 271–280 (1973).
109. Garnett, M. C. Targeted drug conjugates: principles and progress. *Adv. Drug Deliv. Rev.* **53,** 171–216 (2001).
110. Schulz, D. M., Ihling, C., Clore, G. M. & Sinz, A. Mapping the topology and determination of a low-resolution three-dimensional structure of the calmodulin-melittin complex by chemical cross-linking and high-resolution FTICRMS: direct demonstration of multiple binding modes. *Biochemistry* **43,** 4703–4715 (2004).
111. Takeda, S., Tsukiji, S., Ueda, H. & Nagamune, T. Covalent split protein fragment–DNA hybrids generated through N-terminus-specific modification of proteins by oligonucleotides. *Org. Biomol. Chem.* **6,** 2187–2194 (2008).
112. Damle, N. K. Antibody-drug conjugates ace the tolerability test. *Nat. Biotechnol.* **26,** 884–885 (2008).
113. Alley, S. C., Okeley, N. M. & Senter, P. D. Antibody–drug conjugates: targeted drug delivery for cancer. *Curr. Opin. Chem. Biol.* **14,** 529–537 (2010).
114. Webb, S. Pharma interest surges in antibody drug conjugates. *Nat. Biotechnol.* **29,** 297–298 (2011).
115. Janthur, W.-D., Cantoni, N. & Mamot, C. Drug Conjugates Such as Antibody Drug Conjugates (ADCs), Immunotoxins and Immunoliposomes Challenge Daily Clinical Practice. *Int J Mol Sci* **13,** 16020–16045 (2012).
116. Zolot, R. S., Basu, S. & Million, R. P. Antibody–drug conjugates. *Nat. Rev. Drug. Discov.* **12,** 259–260 (2013).
117. Abdella, P. M., Smith, P. K. & Royer, G. P. A new cleavable reagent for cross-linking and reversible immobilisation of proteins. *Biochem. Biophys. Res. Commun.* **87,** 734–742 (1979).
118. Zarling, D. A., Watson, A. & Bach, F. H. Mapping of lymphocyte surface polypeptide antigens by chemical cross-linking with BSOCOES. *J. Immunol.* **124,** 913–920 (1980).
119. Smith, R. J., Capaldi, R. A., Muchmore, D. & Dahlquist, F. Cross-linking of ubiquinone cytochrome c reductase (complex III) with periodate-cleavable bifunctional reagents. *Biochemistry* **17,** 3719–3723 (1978).
120. Reents, R., Jeyaraj, D. A. & Waldmann, H. Enzymatically cleavable linker groups in polymer-supported synthesis. *Drug Discov. Today* **7,** 71–76 (2002).
121. Leriche, G., Chisholm, L. & Wagner, A. Cleavable linkers in chemical biology. *Bioorg. Med. Chem.* **20,** 571–582 (2012).
122. Jin Lee, Y. Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes. *Mol. BioSyst.* **4,** 816–823 (2008).
123. Sinz, A., Kalkhof, S. & Ihling, C. Mapping protein interfaces by a trifunctional cross-linker combined with MALDI-TOF and ESI-FTICR mass spectrometry. *J. Am. Soc. Mass Spectrom.* **16,** 1921–1931 (2005).
124. Wilchek, M. & Bayer, E. A. Biotin-containing reagents. *Methods Enzymol.* **184,** 123–138 (1990).
125. Bayer, E. A. & Wilchek, M. Protein biotinylation. *Methods Enzymol.* **184,** 138–160 (1990).
126. Chen, X., Muthoosamy, K., Pfisterer, A., Neumann, B. & Weil, T. Site-selective lysine modification of native proteins and peptides *via* kinetically controlled labeling. *Bioconjugate Chem.* **23,** 500–508 (2012).
127. Goldman, E. R. *et al.* Avidin: a natural bridge for quantum dot-antibody conjugates. *J. Am. Chem. Soc.* **124,** 6378–6382 (2002).
128. Lockett, M. R., Phillips, M. F., Jarecki, J. L., Peelen, D. & Smith, L. M. A Tetrafluorophenyl Activated Ester Self-Assembled Monolayer for the Immobilisation of Amine-Modified Oligonucleotides. *Langmuir* **24,** 69–75 (2008).
129. Yang, M. *et al.* One-step photochemical attachment of NHS-terminated monolayers onto silicon surfaces and subsequent functionalisation. *Langmuir* **24,** 7931–7938 (2008).
130. Bolton, A. E. & Hunter, W. M. The labelling of proteins to high specific radioactivities by conjugation to a 125 I-containing acylating agent. Application to the radioimmunoassay. *Biochem. J.* **133,** 529–539 (1973).
131. Langone, J. J. [13] Radioiodination by use of the Bolton-Hunter and related reagents. *Methods Enzymol.* **70,** 221–247 (1980).
132. Anderson, P. [39] Radiolabeling of human immune interferon with 125 I-labeled Bolton-Hunter reagent. *Methods Enzymol.* **119,** 281–283 (1986).
133. Glaser, M. *et al.* Preparation of the iodine-124 derivative of the Bolton-Hunter reagent ([124I]I-SHPP) and its use for labelling a VEGF antibody as a PET tracer. *J. Labelled Compd. Radiopharm.* **45,** 1077–1090 (2002).
134. Russell, J. *et al.* Iodination of Annexin V for imaging apoptosis. *J. Nucl. Med.* **43,** 671–677 (2002).

135. Collingridge, D. R., Carroll, V. A., Glaser, M. & Aboagye, E. O. The Development of [124I] Iodinated-VG76e a novel tracer for imaging vascular endothelial growth factor in vivo using positron emission tomography. *Cancer Res.* **62,** 5912–5919 (2002).
136. Park, J. J., Lee, T. S., Kang, J. H., Song, R. & Cheon, G. J. Radioiodination and biodistribution of quantum dots using Bolton-Hunter reagent. *Appl. Radiat. Isot.* **69,** 56–62 (2011).
137. Ji, T. H. & Ji, I. Macromolecular photoaffinity labeling with radioactive photoactivable heterobifunctional reagents. *Anal. Biochem.* **121,** 286–289 (1982).
138. Ross, P. L. Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Mol. Cell. Proteom.* **3,** 1154–1169 (2004).
139. Zieske, L. R. A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J. Exp. Bot.* **57,** 1501–1508 (2006).
140. Aggarwal, K., Choe, L. H. & Lee, K. H. Shotgun proteomics using the iTRAQ isobaric tags. *Brief. Funct. Genom. Proteom.* **5,** 112–120 (2006).
141. Mirzaei, H. & Regnier, F. Enhancing Electrospray Ionisation Efficiency of Peptides by Derivatisation. *Anal. Chem.* **78,** 4175–4183 (2006).
142. Borch, R. F., Bernstein, M. D. & Durst, H. D. Cyanohydridoborate anion as a selective reducing agent. *J. Am. Chem. Soc.* **93,** 2897–2904 (1971).
143. Gray, G. R. The direct coupling of oligosaccharides to proteins and derivatized gels. *Arch. Biochem. Biophys.* **163,** 426–428 (1974).
144. Gildersleeve, J. C., Oyelaran, O., Simpson, J. T. & Allred, B. Improved procedure for direct coupling of carbohydrates to proteins *via* reductive amination. *Bioconjugate Chem.* **19,** 1485–1490 (2008).
145. Schwartz, B. A. & Gray, G. R. Proteins containing reductively aminated disaccharides. *Arch. Biochem. Biophys.* **181,** 542–549 (1977).
146. Lee, R. T. & Lee, Y. C. Preparation and some biochemical properties of neoglycoproteins produced by reductive amination of thioglycosides containing an. omega.-aldehydoaglycon. *Biochemistry* **19,** 156–163 (1980).
147. Freire, T., Bay, S., Vichier-Guerre, S., Lo-Man, R. & Leclerc, C. Carbohydrate Antigens: Synthesis Aspects and Immunological Applications in Cancer. *MRMC* **6,** 1357–1373 (2006).
148. Wong, S. Y. Neoglycoconjugates and their applications in glycobiology. *Curr. Opin. Struct. Biol.* **5,** 599–604 (1995).
149. Seeberger, P. H. & Werz, D. B. Synthesis and medical applications of oligosaccharides. *Nature* **446,** 1046–1051 (2007).
150. Cipolla, L., Peri, F. & Airoldi, C. Glycoconjugates in Cancer Therapy. *ACAMC* **8,** 92–121 (2008).
151. Oppenheimer, S. B., Alvarez, M. & Nnoli, J. Carbohydrate-based experimental therapeutics for cancer, HIV/AIDS and other diseases. *Acta Histochem.* **110,** 6–13 (2008).
152. Tahmassebi, D. C. & Sasaki, T. Synthesis of a Three-Helix Bundle Protein by Reductive Amination. *J. Org. Chem.* **63,** 728–731 (1998).
153. Bentley, M. D., Roberts, M. J. & Harris, J. M. Reductive amination using poly(ethylene glycol) acetaldehyde hydrate generated in situ: Applications to chitosan and lysozyme. *J. Pharm. Sci.* **87,** 1446–1449 (1998).
154. Tao, L., Mantovani, G., Lecolley, F. & Haddleton, D. M. Alpha-aldehyde terminally functional methacrylic polymers from living radical polymerisation: application in protein conjugation PEGylation. *J. Am. Chem. Soc.* **126,** 13220–13221 (2004).
155. Mallik, R., Jiang, T. & Hage, D. S. High-Performance Affinity Monolith Chromatography: Development and Evaluation of Human Serum Albumin Columns. *Anal. Chem.* **76,** 7013–7022 (2004).
156. Tripathi, R., Verma, S., Pandey, J. & Tiwari, V. Recent Development on Catalytic Reductive Amination and Applications. *Curr. Org. Chem.* **12,** 1093–1115 (2008).
157. McFarland, J. & Francis, M. Reductive alkylation of proteins using iridium catalyzed transfer hydrogenation. *J. Am. Chem. Soc.* **127,** 13490–13491 (2005).
158. Weber, G. Polarisation of the fluorescence of macromolecules. 2. Fluorescent conjugates of ovalbumin and bovine serum albumin. *Biochem. J.* **51,** 155–167 (1952).
159. Hartley, B. S. & Massey, V. The active centre of chymotrypsin. I. Labelling with a fluorescent dye. *Biochim. Biophys. Acta* **21,** 58–70 (1956).
160. Gray, W. R. [52] Sequential degradation plus dansylation. *Methods Enzymol.* **11,** 469–475 (1967).
161. Lefevre, C. *et al.* Texas Red-X and Rhodamine Red-X, New Derivatives of Sulforhodamine 101 and Lissamine Rhodamine B with Improved Labeling and Fluorescence Properties. *Bioconjugate Chem.* **7,** 482–489 (1996).
162. Titus, J. A. J., Haugland, R. R., Sharrow, S. O. S. & Segal, D. M. D. Texas Red, a hydrophilic, red-emitting fluorophore for use with fluorescein in dual parameter flow microfluorometric and fluorescence microscopic studies. *J. Immunol. Methods.* **50,** 193–204 (1981).
163. Seiler, N. Use of the dansyl reaction in biochemical analysis. *Methods of Biochemical Analysis, Volume 18* 259–337 (1970).
164. Mao, S.-Y. & Mullins, J. M. in *Theorien in der Entwicklungspsychologie* (Ahnert, L.) **588,** 43–48 (Humana Press, 2009).
165. Gray, W. R. [12] Dansyl chloride procedure. *Methods Enzymol.* **11,** 139–151 (1967).
166. Gros, C. & Labouesse, B. Study of the dansylation reaction of amino acids, peptides and proteins. *Eur. J. Biochem.* **7,** 463–470 (1969).
167. Boyle, R. E. The reaction of dimethyl sulfoxide and 5-dimethylaminonaphthalene-1-sulfonyl chloride. *J. Org. Chem.* **31,** 3880–3882 (1966).
168. Lin, J. K. & Chang, J. Y. Chromophoric labeling of amino acids with 4-dimethylaminoazobenzene-4'-sulfonyl chloride. *Anal. Chem.* **47,** 1634–1638 (1975).
169. Evangelista, R. A. *et al.* A new europium chelate for protein labelling and time-resolved fluorometric applications. *Clin. Biochem.* **21,** 173–178 (1988).
170. Andreoni, A. *et al.* Fluorochromes with long-lived fluorescence as potential labels for pulsed laser immunocytofluorometry: photophysical characterisation of pyrene derivatives. *J. Biochem. Biophys. Methods* **29,** 157–172 (1994).
171. Gorovits, B. M. & Horowitz, P. M. The Molecular Chaperonin cpn60 Displays Local Flexibility That Is Reduced after Binding with an Unfolded Protein. *J. Biol. Chem.* **270,** 13057–13062 (1995).
172. Herzig, D. J., Rees, A. W. & Day, R. A. Bifunctional reagents and protein structure determination. The reaction of phenolic disulfonyl chlorides with lysozyme. *Biopolymers* **2,** 349–360 (1964).
173. Gold, A. M. Sulfonyl Fluorides as Inhibitors of Esterases. III. Identification of Serine as the Site of Sulfonylation in Phenylmethanesulfonyl α-Chymotrypsin. *Biochemistry* **4,** 897–901 (1965).
174. Tipson, R. S. On esters of p-toluenesulfonic acid. *J. Org. Chem.* **9,** 235–241 (1944).
175. Crossland, R. K., Wells, W. E. & Shiner, V. J. Sulfonate leaving groups, structure and reactivity. 2,2,2-Trifluoroethanesulfonate. *J. Am. Chem. Soc.* **93,** 4217–4219 (1971).
176. Nilsson, K. & Mosbach, K. Immobilisation of enzymes and affinity ligands to various hydroxyl group carrying supports using highly reactive sulfonyl chlorides. *Biochem. Biophys. Res. Commun.* **102,** 449–457 (1981).
177. Ma, Z., Gao, C., Ji, J. & Shen, J. Protein immobilisation on the surface of poly-L-lactic acid films for improvement of cellular interactions. *Eur. Polym. J.* **38,** 2279–2284 (2002).
178. Pugdee, K. *et al.* Gene expression of MC3T3-E1 cells on fibronectin-immobilized titanium using tresyl chloride activation technique. *Dent. Mat. J.* **26,** 647–655 (2007).
179. Delgado, C., Patel, J. N., Francis, G. E. & Fisher, D. Coupling of poly (ethylene glycol) to albumin under very mild conditions by activation with tresyl chloride: characterisation of the conjugate

by partitioning in aqueous two-phase systems. *Biotechnol. Bioprocess Eng.* **12,** 119–128 (1990).
180. Francis, G. E. *et al.* PEGylation of cytokines and other therapeutic proteins and peptides: the importance of biological optimisation of coupling techniques. *Int. J. Hematol.* **68,** 1–18 (1998).
181. Roberts, M. J., Bentley, M. D. & Harris, J. M. Chemistry for peptide and protein PEGylation. *Adv. Drug Deliv. Rev.* **64,** 116–127 (2012).
182. Gais, H.-J. & Ruppert, S. Modification and immobilisation of proteins with polyethylene glycol tresylates and polysaccharide tresylates: Evidence suggesting a revision of the coupling mechanism and the structure of the polymer-polymer linkage. *Tetrahedron Lett.* **36,** 3837–3838 (1995).
183. F Sanger, H. T. The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. *Biochem. J.* **49,** 481–490 (1951).
184. Eisen, H. N., Belman, S. & Carsten, M. E. The reaction of 2, 4-dinitrobenzenesulfonic acid with free amino groups of proteins. *J. Am. Chem. Soc.* **75,** 4583–4585 (1953).
185. Shaltiel, S. Thiolysis of some dinitrophenyl derivatives of amino acids. *Biochem. Biophys. Res. Commun.* **29,** 178–183 (1967).
186. Imai, K. & Watanabe, Y. Fluorimetric determination of secondary amino acids by 7-fluoro-4-nitrobenzo-2-oxa-1, 3-diazole. *Anal. Chim. Acta* **130,** 377–383 (1981).
187. Elbashir, A. A. & Suliman, F. The application of 7-chloro-4-nitrobenzoxadiazole (NBD-CL) for the analysis of pharmaceutical-bearing amine group using spectrophotometry and spectrofluorimetry …. *Appl. Spectrosc.* **46,** 222–241 (2011).
188. Uchiyama, S., Santa, T., Okiyama, N., Fukushima, T. & Imai, K. Fluorogenic and fluorescent labeling reagents with a benzofurazan skeleton. *Biomed. Chromatogr.* **15,** 295–318 (2001).
189. Santa, T., Fukushima, T., Ichibangase, T. & Imai, K. Recent progress in the development of derivatisation reagents having a benzofurazan structure. *Biomed. Chromatogr.* **22,** 343–353 (2008).
190. Protein crystallizability. **609,** 385–400 (2010).
191. Sutton, D. A., Drewes, S. E. & Welz, U. Evaluation of 1-fluoro-2-nitro-4-trimethylammoniobenzene iodide, a protein-solubilizing reagent. *Biochem. J.* **130,** 589–595 (1972).
192. Lesniak, M. *et al.* Receptor binding studies on A1-(2-nitro-4-trimethylammoniophenyl) insulin. *H.-S. Z. Physiol. Chem.* **360,** 467–472 (1979).
193. Sutton, D. A., Easter, B. R. & Drewes, S. E. Preparation of a crystalline A1-(2-nitro-4-trimethylammoniophenyl) derivative of bovine insulin. *H.-S. Z. Physiol. Chem.* **357,** 971–976 (1976).
194. Losse, G. & Mützler, G. Funktionelle Substitutionen am Insulinmolekül. *Z. Chem.* **21,** 394–403 (1981).
195. Ladd, D. L. & Snow, R. A. Reagents for the preparation of chromophorically labeled polyethylene glycol-protein conjugates. *Anal. Biochem.* **210,** 258–261 (1993).
196. Hunter, M. J. & Ludwig, M. L. *J. Am. Chem. Soc.* **84,** 3491–3504 (1962).
197. Traut, R. R. *et al.* Methyl 4-mercaptobutyrimidate as a cleavable crosslinking reagent and its application to the Escherichia coli 30S ribosome. *Biochemistry* **12,** 3266–3273 (1973).
198. Nomura, M. Assembly of bacterial ribosomes. *Science* **179,** 864–873 (1973).
199. Wofsy, L. & Singer, S. J. Effects of the Amidination Reaction on Antibody Activity and on the Physical Properties of Some Proteins. *Biochemistry* **2,** 104–116 (1963).
200. Schramm, H. J. H. The synthesis of mono- and bifunctional nitriles and imidoesters carrying a fluorescent group. *H.-S. Z. Physiol. Chem.* **356,** 1375–1379 (1975).
201. Bozler, H., Jany, K. D. & Pfleiderer, G. Synthesis and application of a fluorescent imido ester for specific labelling of amino groups in proteins. *Biochim. Biophys. Acta* **749,** 238–243 (1983).
202. Lee, Y. C., Stowell, C. P. & Krantz, M. J. 2-Imino-2-methoxyethyl 1-thioglycosides: new reagents for attaching sugars to proteins. *Biochemistry* **15,** 3956–3963 (1976).
203. Mehlhorn, R. R., Swanson, M. M., Packer, L. L. & Smith, P. P. An imidoester spin probe of membrane protein interactions: application to cytochrome c. *Arch. Biochem. Biophys.* **204,** 471–476 (1980).
204. Denny, J. B. J. & Roberts, R. M. R. A new immunoreactive probe for the isolation and analysis of plasma membrane polypeptides. Synthesis and properties of isethionyl-3-(N-2,4-dinitrophenyl)-aminopropioimidate. *J. Biol. Chem.* **257,** 2460–2468 (1982).
205. Wood, F. T., Wu, M. M. & Gerhart, J. C. The radioactive labeling of proteins with an iodinated amidination reagent. *Anal. Biochem.* **69,** 339–349 (1975).
206. Wright, D. E. PEG imidates and protein derivatives thereof. Patent: EP0539167 (1993).
207. Ampon, K. & Means, G. E. Immobilisation of proteins on organic polymer beads. *Biotechnol. Bioeng.* **32,** 689–697 (1988).
208. Gavrilyuk, J. I., Wuellner, U. & Barbas, C. F. Beta-lactam-based approach for the chemical programming of aldolase antibody 38C2. *Bioorg. Med. Chem. Lett.* **19,** 1421–1424 (2009).
209. Gavrilyuk, J. I. *et al.* An efficient chemical approach to bispecific antibodies and antibodies of high valency. *Bioorg. Med. Chem. Lett.* **19,** 3716–3720 (2009).
210. Tietze, L. F. *et al.* Anticancer Agents, 15. Squaric Acid Diethyl Ester: A New Coupling Reagent for the Formation of Drug Biopolymer Conjugates. Synthesis of Squaric Acid Ester Amides and Diamides. *Chem. Ber.* **124,** 1215–1221 (1991).
211. Tietze, L. F. *et al.* Conjugation of p-aminophenyl glycosides with squaric acid diester to a carrier protein and the use of the neoglycoprotein in the histochemical detection of lectins. *Bioconjugate Chem.* **2,** 148–153 (1991).
212. Wurm, F., Dingels, C., Frey, H. & Klok, H.-A. Squaric Acid Mediated Synthesis and Biological Activity of a Library of Linear and Hyperbranched Poly(Glycerol)–Protein Conjugates. *Biomacromolecules* **13,** 1161–1171 (2012).
213. Wurm, F., Steinbach, T. & Klok, H.-A. One-pot squaric acid diester mediated aqueous protein conjugation. *Chem. Commun.* **49,** 7815–7817 (2013).
214. Blakeslee, D. & Baines, M. G. Immunofluorescence using dichlorotriazinylaminofluorescein (DTAF) I. Preparation and fractionation of labelled IgG. *J. Immunol. Methods.* **13,** 305–320 (1976).
215. Blakeslee, D. Immunifluorescence using dichlorotriazinylaminofluorescein (DTAF) II. Preparation, purity and stability of the compound. *J. Immunol. Methods.* **17,** 361–364 (1977).
216. Abuchowski, A., Es, T., Palczuk, N. C. & Davis, F. F. Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol. *J. Biol. Chem.* **252,** 3578–3581 (1977).
217. Abuchowski, A., McCoy, J. R., Palczuk, N. C., Es, T. & Davis, F. F. Effect of covalent attachment of polyethylene glycol on immunogenicity and circulating life of bovine liver catalase. *J. Biol. Chem.* **252,** 3582–3586 (1977).
218. Matsushima, A. Modification of E. coli asparaginase with 2, 4-bis (O-methoxypolyethylene glycol)-6-chloro-s-triazine (activated PEG2); disappearance of binding ability towards anti-serum and retention of enzymic activity. *Chem. Lett.* **7,** 773–776 (1980).
219. Arpicco, S. S. *et al.* Novel poly(ethylene glycol) derivatives for preparation of ribosome-inactivating protein conjugates. *Bioconjugate Chem.* **13,** 757–765 (2002).
220. Ikeda, Y., Katamachi, J., Kawasaki, H. & Nagasaki, Y. Novel Protein PEGylation Chemistry *via* Glutalaldehyde-Functionalized PEG. *Bioconjugate Chem.* 131104112908000 (2013). doi:10.1021/bc400169s
221. Gurd, F. R. [34a] carboxymethylation. *Methods Enzymol.* **25,** 424–438 (1971).
222. Forristal, I. The chemistry of α, β-unsaturated sulfoxides and sulfones: an update. *J. Sulfur Chem.* **26,** 163–185 (2005).
223. Valero, E. *et al.* Magnetic Nanoparticles-Templated Assembly of Protein Subunits: A New Platform for Carbohydrate-Based MRI Nanoprobes. *J. Am. Chem. Soc.* **133,** 4889–4895 (2011).

224. Megia-Fernandez, A., Hernandez-Mateo, F. & Santoyo-Gonzalez, F. Vinyl sulfone-based ferrocenylation reagents: applications in conjugation and bioconjugation. *Org. Biomol. Chem.* **11,** 2586–2596 (2013).
225. Lundblad, R. L. *Chemical Reagents for Protein Modification, Third Edition*. (CRC Press, 2010).
226. Brun, M.-P. & Gauzy-Lazo, L. Protocols for Lysine Conjugation. *Antibody Drug Conjugates* **1045,** 173–187 (2013).
227. Leavell, M. D., Novak, P., Behrens, C. R., Schoeniger, J. S. & Kruppa, G. H. Strategy for selective chemical cross-linking of tyrosine and lysine residues. *J. Am. Soc. Mass Spectrom.* **15,** 1604–1611 (2004).
228. Swaim, C. L., Smith, J. B. & Smith, D. L. Unexpected products from the reaction of the synthetic cross-linker 3,3′-dithiobis(sulfosuccinimidyl propionate), DTSSP with peptides. *Mikrochim. Acta* **15,** 736–749 (2004).
229. Kalkhof, S. & Sinz, A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Anal. Bioanal. Chem.* **392,** 305–312 (2008).
230. Mädler, S., Seitz, M., Robinson, J. & Zenobi, R. Does chemical cross-linking with NHS esters reflect the chemical equilibrium of protein-protein noncovalent interactions in solution? *Mikrochim. Acta* **21,** 1775–1783 (2010).
231. Miller, B. T., Rogers, M. E., Smith, J. S. & Kurosky, A. Identification and Characterisation of O-Biotinylated Hydroxy Amino Acid Residues in Peptides. *Anal. Biochem.* **219,** 240–248 (1994).
232. Mädler, S., Gschwind, S. & Zenobi, R. *Role of arginine in chemical cross-linking with N-hydroxysuccinimide esters. Anal. Biochem.* **398,** 123–125 (2010).
233. Brotzel, F. & Mayr, H. Nucleophilicities of amino acids and peptides. *Org. Biomol. Chem.* **5,** 3814–3820 (2007).
234. Gilis, D., Massar, S., Cerf, N. J. & Rooman, M. Optimality of the genetic code with respect to protein stability and amino-acid frequencies. *Genome Biol.* **2,** 1–12 (2001).
235. Fodje, M. N. & Al-Karadaghi, S. Occurrence, conformational features and amino acid propensities for the π-helix. *Protein Eng.* **15,** 353–358 (2002).
236. Sirk, S. J., Olafsen, T., Barat, B., Bauer, K. B. & Wu, A. M. Site-Specific, Thiol-Mediated Conjugation of Fluorescent Probes to Cysteine-Modified Diabodies Targeting CD20 or HER2. *Bioconjugate Chem.* **19,** 2527–2534 (2008).
237. Gorman, J. J., Corino, G. L. & Mitchell, S. J. Fluorescent labeling of cysteinyl residues. *Eur. J. Biochem.* **168,** 169–179 (1987).
238. Burns, J. A., Butler, J. C., Moran, J. & Whitesides, G. M. Selective reduction of disulfides by tris (2-carboxyethyl) phosphine. *J. Org. Chem.* **56,** 2648–2650 (1991).
239. Getz, E. B., Xiao, M., Chakrabarty, T., Cooke, R. & Selvin, P. R. A Comparison between the Sulfhydryl Reductants Tris(2-carboxyethyl)phosphine and Dithiothreitol for Use in Protein Biochemistry. *Anal. Biochem.* **273,** 73–80 (1999).
240. Hansen, R. E. & Winther, J. R. An introduction to methods for analyzing thiols and disulfides: Reactions, reagents, and practical considerations. *Anal. Biochem.* **394,** 147–158 (2009).
241. Tyagarajan, K., Pretzer, E. & Wiktorowicz, J. E. Thiol-reactive dyes for fluorescence labeling of proteomic samples. *Electrophoresis* **24,** 2348–2358 (2003).
242. Shafer, D. E., Inman, J. K. & Lees, A. Reaction of Tris(2-carboxyethyl)phosphine (TCEP) with maleimide and alpha-haloacyl groups: anomalous elution of TCEP by gel filtration. *Anal. Biochem.* **282,** 161–164 (2000).
243. Goddard, D. R. & Michaelis, L. Derivatives of keratin. *J. Biol. Chem.* **112,** 361–371 (1935).
244. Lundell, N. & Schreitmüller, T. Sample Preparation for Peptide Mapping— A Pharmaceutical Quality-Control Perspective. *Anal. Biochem.* **266,** 31–47 (1999).
245. Gurd, F. R. [34a] carboxymethylation. *Methods Enzymol.* **25,** 424–438 (1972).
246. Nielsen, M. L. *et al.* Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat. Methods* **5,** 459–460 (2008).
247. Gerwin, B. I. Properties of the single sulfhydryl group of streptococcal proteinase. A comparison of the rates of alkylation by chloroacetic acid and chloroacetamide. *J. Biol. Chem.* **242,** 451–456 (1967).
248. Kim, J. R., Yoon, H. W., Kwon, K. S., Lee, S. R. & Rhee, S. G. Identification of proteins containing cysteine residues that are sensitive to oxidation by hydrogen peroxide at neutral pH. *Anal. Biochem.* **283,** 214–221 (2000).
249. Davis, N. J. & Flitsch, S. L. A novel method for the specific glycosylation of proteins. *Tetrahedron Lett.* **32,** 6793–6796 (1991).
250. Macmillan, D., Bill, R., Sage, K., Fern, D. & Flitsch, S. Selective in vitro glycosylation of recombinant proteins: semi-synthesis of novel homogeneous glycoforms of human erythropoietin. *Chem. Biol.* **8,** 133–145 (2001).
251. Swanwick, R. S., Daines, A. M., Tey, L.-H., Flitsch, S. L. & Allemann, R. K. Increased Thermal Stability of Site-Selectively Glycosylated Dihydrofolate Reductase. *Chem. Eur. J. of Chem. Bio.* **6,** 1338–1340 (2005).
252. Mackworth, J. F. The inhibition of thiol enzymes by lachrymators. *Biochem. J.* **42,** 82–90 (1948).
253. Clark, P. I. & Lowe, G. Chemical mutations of papain. The preparation of ser 25- and gly 25-papain. *J. Am. Chem. Soc* **24,** 923–924 (1977).
254. Clark, P. I. & Lowe, G. Conversion of the Active-Site Cysteine Residue of Papain into a Dehydro-serine, a Serine and a Glycine Residue. *Eur. J. Biochem.* **84,** 293–299 (1978).
255. Furlanetto, R. W. & Kaiser, E. T. Reaction of the active site of papain with a 'reporter' group labeled phenacyl halide. *J. Am. Chem. Soc.* **92,** 6980–6982 (1970).
256. Furlanetto, R. W., Mochizuki, M. & Kaiser, E. T. Aryl alpha-haloalkyl ketoximes as enzyme modifying agents. The reaction of alpha-haloacetophenone oximes with the active site of papain. *Biochem. Biophys. Res. Commun.* **58,** 192–196 (1974).
257. Schwartz, K. J., Nakagawa, Y. & Kaiser, E. T. Modification of the reactive sulfhydryl group in phosphofructokinase. *J. Am. Chem. Soc.* **98,** 6369–6378 (1976).
258. Kaiser, E. & Lawrence, D. Chemical mutation of enzyme active sites. *Science* **226,** 505–511 (1984).
259. Friedmann, E. & Marrian, D. H. Antimitotic action of maleimide and related substances. *Br. J. Pharmacol.* **4,** 105–108 (1949).
260. Friedmann, E. Spectrophotometric investigation of the interaction of glutathione with maleimide and N-ethylmaleimide. *Biochim. Biophys. Acta* **9,** 65–75 (1952).
261. Kajfež, T., Kamenar, B., Piližota, V. & Fleš, D. Crystal and Molecular Structures of N-Phenylmaleimide and N-Phenyl-2, 3-dimethylmaleimide. *Croat. Chem. Acta* **76,** 343–346 (2003).
262. Bodige, S. G., Méndez-Rojas, M. A. & Watson, W. H. Structure and properties of N-phenylmaleimide derivatives. *J. Chem. Crystallogr.* **29,** 57–66 (1999).
263. Smyth, D. G., Blumenfeld, O. O. & Konigsberg, W. Reactions of N-ethylmaleimide with peptides and amino acids. *Biochem. J.* **91,** 589–595 (1964).
264. Brown, R. D. & Matthews, K. S. Chemical modification of lactose repressor protein using N-substituted maleimides. *J. Biol. Chem.* **254,** 5128–5134 (1979).
265. Ishii, Y. & Lehrer, S. S. Effects of the state of the succinimido-ring on the fluorescence and structural properties of pyrene maleimide-labeled alpha alpha-tropomyosin. *Biophys. J.* **50,** 75–80 (1986).
266. Wu, C.-W., Yarbrough, L. R. & Wu, F. Y. N-(1-pyrene) maleimide: a fluorescent crosslinking reagent. *Biochemistry* **15,** 2863–2868 (1976).
267. Climent, T., González-Luque, R. & Merchán, M. Theoretical analysis of the excited states in maleimide. *J. Phys. Chem. A* **107,** 6995–7003 (2003).
268. Schelté, P., Boeckler, C., Frisch, B. & Schuber, F. Differential reactivity of maleimide and bromoacetyl functions with thiols: application to the preparation of liposomal diepitope constructs. *Bioconjugate Chem.* **11,** 118–123 (2000).

269. Yoshitake, S., Yamada, Y., Ishikawa, E. & Masseyeff, R. Conjugation of glucose oxidase from Aspergillus niger and rabbit antibodies using N-hydroxysuccinimide ester of N-(4-carboxycyclohexylmethyl)-maleimide. *Eur. J. Biochem.* **101,** 395–399 (1979).
270. Yoshitake, S. *et al.* Mild and efficient conjugation of rabbit Fab'and horseradish peroxidase using a maleimide compound and its use for enzyme immunoassay. *J. Biochem.* **92,** 1413–1424 (1982).
271. Hashida, S., Imagawa, M., Inoue, S., Ruan, K. H. & Ishikawa, E. More useful maleimide compounds for the conjugation of Fab' to horseradish peroxidase through thiol groups in the hinge. *J. Appl. Biochem.* **6,** 56–63 (1984).
272. Ghosh, S. S., Kao, P. M., McCue, A. W. & Chappelle, H. L. Use of maleimide-thiol coupling chemistry for efficient syntheses of oligonucleotide-enzyme conjugate hybridisation probes. *Bioconjugate Chem.* **1,** 71–76 (1990).
273. Dickgreber, N. *et al.* Targeting antigen to MHC class II molecules promotes efficient cross-presentation and enhances immunotherapy. *J. Immunol.* **182,** 1260–1269 (2009).
274. Hu, T. & Su, Z. Bovine serum albumin-bovine hemoglobin conjugate as a candidate blood substitute. *Biotechnol. Lett.* **24,** 275–278 (2002).
275. Kafi, K. *et al.* Maleimide conjugation markedly enhances the immunogenicity of both human and murine idiotype-KLH vaccines. *Mol. Immunol.* **46,** 448–456 (2009).
276. Ehrlich, G. K., Michel, H., Truitt, T. & Riboulet, W. Preparation and Characterisation of Albumin Conjugates of a Truncated Peptide YY Analog for Half-Life Extension. *Bioconjugate Chem.* (2013). doi:10.1021/bc400340z
277. Kratz, F. *et al.* Probing the cysteine-34 position of endogenous serum albumin with thiol-binding doxorubicin derivatives. Improved efficacy of an acid-sensitive doxorubicin derivative with specific albumin-binding properties compared to that of the parent compound. *J. Med. Chem.* **45,** 5523–5533 (2002).
278. Warnecke, A., Fichtner, I., Sass, G. & Kratz, F. Synthesis, cleavage profile, and antitumor efficacy of an albumin-binding prodrug of methotrexate that is cleaved by plasmin and cathepsin B. *Arch. Pharm. (Weinheim, Ger.)* **340,** 389–395 (2007).
279. Kersemans, V., Cornelissen, B., Minden, M. D., Brandwein, J. & Reilly, R. M. Drug-resistant AML cells and primary AML specimens are killed by 111In-Anti-CD33 monoclonal antibodies modified with nuclear localizing peptide sequences. *J. Nucl. Med.* **49,** 1546–1554 (2008).
280. Betting, D. J., Kafi, K., Abdollahi-Fard, A., Hurvitz, S. A. & Timmerman, J. M. Sulfhydryl-based tumor antigen-carrier protein conjugates stimulate superior antitumor immunity against B cell lymphomas. *J. Immunol.* **181,** 4131–4140 (2008).
281. Alisio, A. Relative proximity and orientation of helices 4 and 8 of the GLUT1 glucose transporter. *J. Biol. Chem.* **279,** 26540–26545 (2004).
282. Giron-Monzon, L. Mapping protein-protein interactions between MutL and MutH by cross-linking. *J. Biol. Chem.* **279,** 49338–49345 (2004).
283. Green, N. S., Reisler, E. & Houk, K. N. Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers. *Protein Sci.* **10,** 1293–1304 (2008).
284. Ishmael, F. T., Trakselis, M. A. & Benkovic, S. J. Protein-Protein Interactions in the Bacteriophage T4 Replisome: the leading strand holoenzyme is physically linked to the lagging strand holoenzyme and the primosome. *J. Biol. Chem.* **278,** 3145–3152 (2003).
285. Kovalenko, O. V., Yang, X., Kolesnikova, T. V. & Hemler, M. E. Evidence for specific tetraspanin homodimers: inhibition of palmitoylation makes cysteine residues available for cross-linking. *Biochem. J.* **377,** 407–417 (2004).
286. Kwon, Y. K. Role of the pleckstrin homology domain of PLC 1 in its interaction with the insulin receptor. *J. Cell Biol.* **163,** 375–384 (2003).
287. Stalteri, M. A. & Mather, S. J. A Cross-Linked Monoclonal Antibody Fragment for Improved Tumor Targeting. *Bioconjugate Chem.* **6,** 179–186 (1995).
288. Leary Swan, E. E., Popat, K. C. & Desai, T. A. Peptide-immobilized nanoporous alumina membranes for enhanced osteoblast adhesion. *Biomaterials* **26,** 1969–1976 (2005).
289. Misra, A. & Dwivedi, P. Immobilisation of oligonucleotides on glass surface using an efficient heterobifunctional reagent through maleimide-thiol combination chemistry. *Anal. Biochem.* **369,** 248–255 (2007).
290. Elia, G. Protein biotinylation. *Curr. Protoc. Protein. Sci.* Chapter 3. Unit 3.6 (2010).
291. Oda, Y., Nagasu, T. & Chait, B. T. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome : Abstract : Nature Biotechnology. *Nat. Biotechnol.* **19,** 379–382 (2001).
292. Khurana, S., Kennedy, M., King, L. R. & Golding, H. Identification of a linear peptide recognized by monoclonal antibody 2D7 capable of generating CCR5-specific antibodies with human immunodeficiency virus-neutralizing activity. *J. Virol.* **79,** 6791–6800 (2005).
293. Golding, H. CCR5 N-terminal Region Plays a Critical Role in HIV-1 Inhibition by Toxoplasma gondii-derived Cyclophilin-18. *J. Biol. Chem.* **280,** 29570–29577 (2005).
294. Gast, F.-U., Franke, I., Meiss, G. & Pingoud, A. Immobilisation of sugar-non-specific nucleases by utilizing the streptavidin–biotin interaction. *J. Biotech.* **87,** 131–141 (2001).
295. Cho, I. H., Paek, E. H., Lee, H., Choi, J. W. & Paek, S. H. Site-directed immobilisation of antibody onto solid surfaces for the construction of immunochip. *Biotechnol. Bioprocess Eng.* **9,** 112–117 (2004).
296. Kang, G. Y. *et al.* Label-free protein assay with site-directly immobilized antibody using self-actuating PZT cantilever. *Sens. Actuators, B* **117,** 332–338 (2006).
297. Cho, I.-H. *et al.* Site-directed biotinylation of antibodies for controlled immobilisation on solid surfaces. *Anal. Biochem.* **365,** 14–23 (2007).
298. Pompella, A., Dominici, S., Cambiaggi, C., Frank, J. & Biesalski, H. K. Cytofluorescence techniques for the visualisation of distinct pools of protein thiols at the single cell level. *Methods Mol. Biol.* **196,** 305–312 (2002).
299. Kuiper, J. M. *et al.* A method for site-specific labeling of multiple protein thiols. *Protein Sci.* **18,** 1033–1041 (2009).
300. Chen, X., Zhou, Y., Peng, X. & Yoon, J. Fluorescent and colorimetric probes for detection of thiols. *Chem. Soc. Rev.* **39,** 2120–2135 (2010).
301. Zhang, Y., Bhatt, V. S., Sun, G., Wang, P. G. & Palmer, A. F. Site-Selective Glycosylation of Hemoglobin on Cys β93. *Bioconjugate Chem.* **19,** 2221–2230 (2008).
302. Ahlgren, S. *et al.* Evaluation of Maleimide Derivative of DOTA for Site-Specific Labeling of Recombinant Affibody Molecules. *Bioconjugate Chem.* **19,** 235–243 (2008).
303. Boswell, C. A. *et al.* Enhanced Tumor Retention of a Radiohalogen Label for Site-Specific Modification of Antibodies. *J. Med. Chem.* (2013). doi:10.1021/jm401365h
304. Han, M. K. *et al.* Fluorescence Studies of Pyrene Maleimide-Labeled Translin: Excimer Fluorescence Indicates Subunits Associate in a Tail-to-Tail Configuration to Form Octamer. *Biochemistry* **41,** 3468–3476 (2002).
305. West, J. M. A Fluorescent Probe-labeled Escherichia coli Aspartate Transcarbamoylase That Monitors the Allosteric Conformational State. *J. Biol. Chem.* **279,** 945–951 (2003).
306. Fernandes, F. *et al.* Quantification of Protein-Lipid Selectivity using FRET: Application to the M13 Major Coat Protein. *Biophys. J.* **87,** 344–352 (2004).
307. Nishino, T. G. *et al.* Spatial Orientation of Mitochondrial Processing Peptidase and a Preprotein Revealed by Fluorescence Resonance Energy Transfer. *J. Biochem.* **141,** 889–895 (2007).
308. Wright, S. K. & Viola, R. E. Evaluation of methods for the quantitation of cysteines in proteins. *Anal. Biochem.* **265,** 8–14 (1998).

309. Mastroberardino, P. G., Orr, A. L., Hu, X., Na, H. M. & Greenamyre, J. T. A FRET-based method to study protein thiol oxidation in histological preparations. *Free Radic. Biol. Med.* **45,** 971–981 (2008).
310. Mantovani, G. *et al.* Design and Synthesis of N-Maleimido-Functionalized Hydrophilic Polymers *via* Copper-Mediated Living Radical Polymerisation: A Suitable Alternative to PEGylation Chemistry. *J. Am. Chem. Soc.* **127,** 2966–2973 (2005).
311. Matsui, S. & Aida, H. Hydrolysis of some N-alkylmaleimides. *J. Chem. Soc., Perkin Trans. 2* 1277–1280 (1978). doi:10.1039/P29780001277
312. Kalia, J. & Raines, R. T. Catalysis of imido group hydrolysis in a maleimide conjugate. *Bioorg. Med. Chem. Lett.* **17,** 6286–6289 (2007).
313. Tedaldi, L. M., Smith, M. E. B., Nathani, R. I. & Baker, J. R. Bromomaleimides: new reagents for the selective and reversible modification of cysteine. *Chem. Commun.* 6583–6585 (2009).
314. Roberts, D. D., Lewis, S. D., Ballou, D. P., Olson, S. T. & Shafer, J. A. Reactivity of small thiolate anions and cysteine-25 in papain toward methyl methanethiosulfonate. *Biochemistry* **25,** 5595–5601 (1986).
315. Smith, M. E. B. *et al.* Protein Modification, Bioconjugation, and Disulfide Bridging Using Bromomaleimides. *J. Am. Chem. Soc.* **132,** 1960–1965 (2010).
316. Ryan, C. P. *et al.* Tunable reagents for multi-functional bioconjugation: reversible or permanent chemical modification of proteins and peptides by control of maleimide hydrolysis. *Chem. Commun.* **47,** 5452–5454 (2011).
317. Castañeda, L. *et al.* Acid-cleavable thiomaleamic acid linker for homogeneous antibody–drug conjugation. *Chem. Commun.* **49,** 8187–8189 (2013).
318. Porath, J. & Axén, R. [3] Immobilisation of enzymes to agar, agarose, and sephadex support. **44,** 19–45 (1976).
319. Masri, M. S. & Friedman, M. Protein reactions with methyl and ethyl vinyl sulfones. *J. Protein Chem.* **7,** 49–54 (1988).
320. Houen, G. & Jensen, O. M. Conjugation to preactivated proteins using divinylsulfone and iodoacetic acid. *J. Immunol. Methods.* **181,** 187–200 (1995).
321. Morpurgo, M., Veronese, F. M., Kachensky, D. & Harris, J. M. Preparation and Characterisation of Poly(ethylene glycol) Vinyl Sulfone. *Bioconjugate Chem.* **7,** 363–368 (1996).
322. Meadows, D. C. & Gervay-Hague, J. Vinyl sulfones: Synthetic preparations and medicinal chemistry applications. *Med. Res. Rev.* **26,** 793–814 (2006).
323. Bailon, P. & Won, C.-Y. PEG-modified biopharmaceuticals. *Expert Opin. Drug Deliv.* **6,** 1–16 (2009).
324. Rusmini, F., Zhong, Z. & Feijen, J. Protein immobilisation strategies for protein biochips. *Biomacromolecules* **8,** 1775–1789 (2007).
325. Ortega-Muñoz, M. *et al.* Vinyl sulfone functionalized silica: a 'ready to use' pre-activated material for immobilisation of biomolecules. *J. Mater. Chem.* **20,** 7189–7196 (2010).
326. Morales-Sanfrutos, J. *et al.* Vinyl sulfone: a versatile function for simple bioconjugation and immobilisation. *Org. Biomol. Chem.* **8,** 667–675 (2010).
327. Lopez-Jaramillo, F. J., Ortega-Muñoz, M., Megia-Fernandez, A., Hernandez-Mateo, F. & Santoyo-Gonzalez, F. Vinyl Sulfone Functionalisation: A Feasible Approach for the Study of the Lectin–Carbohydrate Interactions. *Bioconjugate Chem.* **23,** 846–855 (2012).
328. Li, L. *et al.* Vinyl Sulfone Bifunctional Derivatives of DOTA Allow Sulfhydryl- or Amino-Directed Coupling to Antibodies. Conjugates Retain Immunoreactivity and Have Similar Biodistributions. *Bioconjugate Chem.* **13,** 110–115 (2002).
329. Morales-Sanfrutos, J., Lopez-Jaramillo, F. J., Hernandez-Mateo, F. & Santoyo-Gonzalez, F. Vinyl Sulfone Bifunctional Tag Reagents for Single-Point Modification of Proteins. *J. Org. Chem.* **75,** 4039–4047 (2010).
330. Lopez Jaramillo, F., Hernandez Mateo, F., Santoyo Gonzalez, F. & Leung, E. Vinyl sulfone: a multi-purpose function in proteomics. *Integr. Proteom.* **16,** 301–326 (2012).
331. Hemelaar, J. *et al.* Specific and covalent targeting of conjugating and deconjugating enzymes of ubiquitin-like proteins. *Mol. Cell. Biol.* **24,** 84–95 (2004).
332. Posner, T. Beiträge zur Kenntniss der ungesättigten Verbindungen. II. Über die Addition von Mercaptanen an ungesättigte Kohlenwassershoffe. *Chem. Ber.* **38,** 646–657 (1905).
333. Hoyle, C. E. & Bowman, C. N. Thiol–ene click chemistry. *Angew. Chem. Int. Ed. (English)* **49,** 1540–1573 (2010).
334. Dondoni, A. & Marra, A. Recent applications of thiol–ene coupling as a click process for glycoconjugation. *Chem. Soc. Rev.* **41,** 573–586 (2012).
335. Floyd, N., Vijayakrishnan, B., Koeppe, J. R. & Davis, B. G. Thiyl Glycosylation of Olefinic Proteins: S-Linked Glycoconjugate Synthesis. *Angew. Chem. Int. Ed. (English)* **121,** 7938–7942 (2009).
336. Dondoni, A., Massi, A., Nanni, P. & Roda, A. A New Ligation Strategy for Peptide and Protein Glycosylation: Photoinduced Thiol-Ene Coupling. *Chem. Eur. J.* **15,** 11444–11449 (2009).
337. Wu, L.-Z., Sheng, Y.-B., Xie, J.-B. & Wang, W. Photoexcitation of tryptophan groups induced reduction of disulfide bonds in hen egg white lysozyme. *J. Mol. Struct.* **882,** 101–106 (2008).
338. Garber, K. C. A. & Carlson, E. E. Thiol-ene enabled detection of thiophosphorylated kinase substrates. *ACS Chem. Biol.* **8,** 1671–1676 (2013).
339. Stolz, R. M. & Northrop, B. H. Experimental and Theoretical Studies of Selective Thiol–Ene and Thiol–Yne Click Reactions Involving N-Substituted Maleimides. *J. Org. Chem.* **78,** 8105–8116 (2013).
340. Markey, L., Giordani, S. & Scanlan, E. M. Native Chemical Ligation,Thiol–Ene Click: A Methodology for the Synthesis of Functionalized Peptides. *J. Org. Chem.* **78,** 4270–4277 (2013).
341. Massi, A. & Nanni, D. Thiol–yne coupling: revisiting old concepts as a breakthrough for up-to-date applications. *Org. Biomol. Chem.* **10,** 3791–3807 (2012).
342. Bader, H., Cross, L. C., Heilbron, I. & Jones, E. 132. Researches on acetylenic compounds. Part XVIII. The addition of thiolacetic acid to acetylenic hydrocarbons. The conversion of monosubstituted acetylenes into aldehydes and 1: 2-dithiols. *J. Chem. Soc.* 619–623 (1949).
343. Conte, Lo, M., Pacifico, S., Chambery, A., Marra, A. & Dondoni, A. Photoinduced Addition of Glycosyl Thiols to Alkynyl Peptides: Use of Free-Radical Thiol−Yne Coupling for Post-Translational Double-Glycosylation of Peptides. *J. Org. Chem.* **75,** 4644–4647 (2010).
344. Conte, M. L. *et al.* Multi-molecule reaction of serum albumin can occur through thiol-yne coupling. *Chem. Commun.* **47,** 11086–11088 (2011).
345. Crankshaw, M. W. & Grant, G. A. Modification of cysteine. *Curr. Protoc. Protein. Sci.* Chapter 15. Unit 15.1 (2001). doi:10.1002/0471140864.ps1501s03
346. Anson, M. L. The Reactions Of Iodine And Iodoacetamide With Native Egg Albumin. *J. Gen. Physiol.* **23,** 321–331 (1940).
347. Ellman, G. L. Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **82,** 70–77 (1959).
348. Kenyon, G. L. & Bruice, T. W. [40] Novel sulfhydryl reagents. *Methods Enzymol.* **47,** 407–430 (1977).
349. Fontana, A., Scoffone, E. & Benassi, C. A. Sulfenyl halides as modifying reagents for polypeptides and proteins. II. Modification of cysteinyl residues. *Biochemistry* **7,** 980–986 (1968).
350. King, T. P., Li, Y. & Kochoumian, L. Preparation of protein conjugates *via* intermolecular disulfide bond formation. *Biochemistry* **17,** 1499–1506 (1978).
351. Macindoe, W. M., van Oijen, A. H. & Boons, G.-J. A unique and highly facile method for synthesising disulfide linked neoglycoconjugates: a new approach for remodelling of peptides and proteins. *Chem. Commun.* 847–848 (1998). doi:10.1039/A708701B
352. Davis, B. G., Maughan, M. A., Green, M. P., Ullman, A. &

Jones, J. B. Glycomethanethiosulfonates: powerful reagents for protein glycosylation. *Tetrahedron: Asymmetry* **11,** 245–262 (2000).
353. Gamblin, D. P. *et al.* Glycosyl phenylthiosulfonates (Glyco-PTS): novel reagents for glycoprotein synthesis. *Org. Biomol. Chem.* **1,** 3642–3644 (2003).
354. Davis, B. G. Synthesis of Glycoproteins. *Macromolecules* **102,** 579–602 (2002).
355. Zhao, Y. J., Zhai, Y. Q., Su, Z. G. & Ma, G. H. Methoxy poly (ethylene glycol) thiosulfonate: new activated polymer derivatives for thiol-specific modification. *Polym. Adv. Technol.* **21,** 867–873 (2010).
356. Gamblin, D. P. *et al.* Glyco-SeS: Selenenylsulfide-Mediated Protein Glycoconjugation—A New Strategy in Post-Translational Modification. *Angew. Chem. Int. Ed. (English)* **116,** 846–851 (2004).
357. Bernardes, G. J. L., Gamblin, D. P. & Davis, B. G. The Direct Formation of Glycosyl Thiols from Reducing Sugars Allows One-Pot Protein Glycoconjugation. *Angew. Chem. Int. Ed. (English)* **118,** 4111–4115 (2006).
358. Bernardes, G. J. L., Marston, J. P., Batsanov, A. S., Howard, J. A. K. & Davis, B. G. A trisulfide-linked glycoprotein. *Chem. Commun.* 3145–3147 (2007). doi:10.1039/B706682A
359. Steiner, M. *et al.* Spacer length shapes drug release and therapeutic efficacy of traceless disulfide-linked ADCs targeting the tumor neovasculature. *Chem. Sci.* **4,** 297–302 (2013).
360. Kunstelj, M. *et al.* Cysteine-specific PEGylation of rhG-CSF *via* selenylsulfide bond. *Bioconjugate Chem.* **24,** 889–896 (2013).
361. Bernardes, G. J. L. *et al.* From Disulfide- to Thioether-Linked Glycoproteins. *Angew. Chem. Int. Ed. (English)* **47,** 2244–2247 (2008).
362. Rosendahl, M. S. *et al.* A long-acting, highly potent interferon alpha-2 conjugate created using site-specific PEGylation. *Bioconjugate Chem.* **16,** 200–207 (2005).
363. Petersen, M. T., Jonson, P. H. & Petersen, S. B. Amino acid neighbours and detailed conformational analysis of cysteines in proteins. *Protein Eng.* **12,** 535–548 (1999).
364. Leung, H. J., Xu, G., Narayan, M. & Scheraga, H. A. Impact of an easily reducible disulfide bond on the oxidative folding rate of multi-disulfide-containing proteins. *J. Pept. Res.* **65,** 47–54 (2005).
365. Arakawa, T., Prestrelski, S. J., Narhi, L. O., Boone, T. C. & Kenney, W. C. Cysteine 17 of recombinant human granulocyte-colony stimulating factor is partially solvent-exposed. *J. Protein Chem.* **12,** 525–531 (1993).
366. Brocchini, S. *et al.* PEGylation of native disulfide bonds in proteins. *Nat. Protocols* **1,** 2241–2252 (2006).
367. Thornton, J. M. Disulphide bridges in globular proteins. *J. Mol. Biol.* **151,** 261–287 (1981).
368. Bhattacharyya, R., Pal, D. & Chakrabarti, P. Disulfide bonds, their stereospecific environment and conservation in protein structures. *Protein Eng.* **17,** 795–808 (2004).
369. Shaunak, S. *et al.* Site-specific PEGylation of native disulfide bonds in therapeutic proteins. *Nat. Chem. Biol.* **2,** 312–313 (2006).
370. Balan, S. *et al.* Site-Specific PEGylation of Protein Disulfide Bonds Using a Three-Carbon Bridge. *Bioconjugate Chem.* **18,** 61–76 (2007).
371. Brocchini, S. *et al.* Disulfide bridge based PEGylation of proteins. *Adv. Drug Deliv. Rev.* **60,** 3–12 (2008).
372. Khalili, H., Godwin, A., Choi, J.-W., Lever, R. & Brocchini, S. Comparative Binding of Disulfide-Bridged PEG-Fabs. *Bioconjugate Chem.* **23,** 2262–2277 (2012).
373. Lewis, A., Tang, Y., Brocchini, S., Choi, J.-W. & Godwin, A. Poly(2-methacryloyloxyethyl phosphorylcholine) for Protein Conjugation. *Bioconjugate Chem.* **19,** 2144–2155 (2008).
374. Castañeda, L. *et al.* A mild synthesis of N-functionalised bromomaleimides, thiomaleimides and bromopyridazinediones. *Tetrahedron Lett.* **54,** 3493–3495 (2013).
375. Jones, M. W. *et al.* Polymeric dibromomaleimides as extremely efficient disulfide bridging bioconjugation and pegylation agents. *J. Am. Chem. Soc.* **134,** 1847–1852 (2012).
376. Schumacher, F. F. *et al.* Homogeneous antibody fragment conjugation by disulfide bridging introduces 'spinostics'. *Sci. Rep.* **3,** 1–8 (2013).
377. Smith, M. E. B. *et al.* Chemical modification of antibodies. Patent: WO2013132268 (2013).
378. Schumacher, F. F. *et al.* In situ maleimide bridging of disulfides and a new approach to protein PEGylation. *Bioconjugate Chem.* **22,** 132–136 (2011).
379. Jones, M. W. *et al.* Highly efficient disulfide bridging polymers for bioconjugates from radical-compatible dithiophenol maleimides. *Chem. Commun.* **48,** 4064–4066 (2012).
380. Chapman, C. J., Wadsworth, K. J. & Frost, C. G. Enantioselective rhodium-catalysed addition of boronic acids using C2-symmetric aryl diphosphite ligands. *J. Organomet. Chem.* **680,** 206–211 (2003).
381. Chapman, C. J., Matsuno, A., Frost, C. G. & Willis, M. C. Site-selective modification of peptides using rhodium and palladium catalysis: complementary electrophilic and nucleophilic arylation. *Chem. Commun.* 3903–3905 (2007). doi:10.1039/b711533d
382. Chapman, C. J., Hargrave, J. D., Bish, G. & Frost, C. G. Peptide modification through site-selective residue interconversion: application of the rhodium-catalysed 1,4-addition of aryl siloxanes and boronates. *Tetrahedron* **64,** 9528–9539 (2008).
383. Navarre, L., Martinez, R., Genet, J.-P. & Darses, S. Access to enantioenriched α-amino esters *via* Rhodium-catalyzed 1, 4-addition/enantioselective protonation. *J. Am. Chem. Soc.* **130,** 6159–6169 (2008).
384. Bernardes, G. J. L., Chalker, J. M., Errey, J. C. & Davis, B. G. Facile Conversion of Cysteine and Alkyl Cysteines to Dehydroalanine on Protein Surfaces: Versatile and Switchable Access to Functionalized Proteins. *J. Am. Chem. Soc.* **130,** 5052–5053 (2008).
385. Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J. & Davis, B. G. Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post-Translational Modifications. *Angew. Chem. Int. Ed. (English)* **124,** 1871–1875 (2012).
386. Collier, P. N., Campbell, A. D., Patel, I. & Taylor, R. J. K. The direct synthesis of novel enantiomerically pure α-amino acids in protected form *via* Suzuki cross-coupling. *Tetrahedron Lett.* **41,** 7115–7119 (2000).
387. Chalker, J. M. *et al.* Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2,** 1666–1676 (2011).
388. Photaki, I. Transformation of Serine to Cysteine. β-Elimination Reactions in Serine Derivatives. *J. Am. Chem. Soc.* **85,** 1123–1126 (1963).
389. Strumeyer, D. H., White, W. N. & Koshland, D. E. Role of serine in chymotrypsin action. Conversion of the active serine to dehydroalanine. *Proc. Natl. Acad. Sci. U. S. A.* **50,** 931–935 (1963).
390. Weiner, H., White, W. N., Hoare, D. G. & Koshland, D. E. *J. Am. Chem. Soc.* **88,** 3851–3859 (1966).
391. Rich, D. H., Tam, J., Mathiaparanam, P., Grant, J. A. & Mabuni, C. General synthesis of didehydroamino-acids and peptides. 897–898 (1974).
392. Chalker, J., Bernardes, G. ., Lin, Y. & Davis, B. Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. **4,** 630–640 (2009).
393. Holmes, T. J. & Lawton, R. G. Cysteine modification and cleavage of proteins with 2-methyl-N1-benzenesulfonyl-N4-bromoacetylquinonediimide. *J. Am. Chem. Soc.* **99,** 1984–1986 (1977).
394. Rowan, F. C. *et al.* Insights into Aurora-A kinase activation using unnatural amino acids incorporated by chemical modification. *ACS Chem. Biol.* **8,** 2184–2191 (2013).
395. Nathani, R. I. *et al.* A novel approach to the site-selective dual labelling of a protein *via* chemoselective cysteine modification.

Chem. Sci. **4,** 3455–3458 (2013).
396. Hashimoto, K., Sakai, M., Okuno, T. & Shirahama, H. β-Phenylselenoalanine as a dehydroalanine precursor–efficient synthesis of alternariolide (AM-toxin I). *Chem. Commun.* 1139–1140 (1996).
397. Okeley, N. M., Zhu, Y. & van der Donk, W. A. Facile Chemoselective Synthesis of Dehydroalanine-Containing Peptides. *Org. Lett* **2,** 3603–3606 (2000).
398. Seebeck, F. P. & Szostak, J. W. Ribosomal Synthesis of Dehydroalanine-Containing Peptides. *J. Am. Chem. Soc.* **128,** 7150–7151 (2006).
399. Wang, J., Schiller, S. M. & Schultz, P. G. A Biosynthetic Route to Dehydroalanine-Containing Proteins. *Angew. Chem. Int. Ed. (English)* **119,** 6973–6975 (2007).
400. Guo, J., Wang, J., Lee, J. S. & Schultz, P. G. Site-Specific Incorporation of Methyl- and Acetyl-Lysine Analogues into Recombinant Proteins. *Angew. Chem. Int. Ed. (English)* **120,** 6499–6501 (2008).
401. You, Y. O., Levengood, M. R., Ihnken, L. A. F., Knowlton, A. K. & van der Donk, W. A. Lacticin 481 Synthetase as a General Serine/Threonine Kinase. *ACS Chem. Biol.* **4,** 379–385 (2009).
402. Lin, Y. A., Boutureira, O. & Lercher, L. Rapid Cross-Metathesis for Reversible Protein Modifications *via* Chemical Access to Se-Allyl-selenocysteine in Proteins. *J. Am. Chem. Soc.* **135,** 12156–12159 (2013).
403. Lindley, H. A new synthetic substrate for trypsin and its application to the determination of the amino-acid sequence of proteins. **178,** 647–648 (1956).
404. Raftery, M. A. & Cole, R. D. On the aminoethylation of proteins. *J. Biol. Chem.* **241,** 3457–3461 (1966).
405. Simon, M. D. *et al.* The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell* **128,** 1003–1012 (2007).
406. Spokoyny, A. M. *et al.* A Perfluoroaryl-Cysteine SNAr Chemistry Approach to Unprotected Peptide Stapling. *J. Am. Chem. Soc.* **135,** 5946–5949 (2013).
407. Zhang, C., Spokoyny, A. M., Zou, Y., Simon, M. D. & Pentelute, B. L. Enzymatic 'Click' Ligation: Selective Cysteine Modification in Polypeptides Enabled by Promiscuous Glutathione S-Transferase. *Angew. Chem. Int. Ed. Engl.* **52,** 1–6 (2013).
408. Toda, N., Asano, S. & Barbas, C. F., III. Rapid, Stable, Chemoselective Labeling of Thiols with Julia-Kocieński-like Reagents: A Serum-Stable Alternative to Maleimide-Based Protein Conjugation. *Angew. Chem. Int. Ed. (English)* **52,** 12592–12596 (2013).
409. Shiu, H.-Y. *et al.* Electron-Deficient Alkynes as Cleavable Reagents for the Modification of Cysteine-Containing Peptides in Aqueous Medium. *Chem. Eur. J.* **15,** 3839–3850 (2009).
410. Chan, A. O.-Y. *et al.* Gold-mediated selective cysteine modification of peptides using allenes. *Chem. Commun.* **49,** 1428–1430 (2013).
411. Halphen, P. D. & Owen, T. C. Carboxylthioacrylates. *J. Org. Chem.* **38,** 3507–3510 (1973).
412. Truce, W. E. & Tichenor, G. J. W. Effect of activating group on trans-stereoselectivity of thiolate additions to activated acetylenes. *J. Org. Chem.* **37,** 2391–2396 (1972).
413. De Lucchi, O., Lucchini, V., Marchioro, C., Valle, G. & Modena, G. Self-induced diastereoselective oxidation of vinyl sulfoxides bearing a chiral hydroxy group as a way of preparation of optically active sulfinyl dienophiles and their use in the asymmetric Diels-Alder reactions to cyclopentadiene. *J. Org. Chem.* **51,** 1457–1466 (1986).
414. Journet, M., Rouillard, A., Cai, D. & Larsen, R. D. Double Radical Cyclisation/β-Fragmentation of Acyclic ω-Yne Vinyl Sulfides. Synthesis of 3-Vinyldihydrothiophene and Dihydrothiopyran Derivatives. A New Example of a 5-endo-trig Radical Cyclisation. *J. Org. Chem.* **62,** 8630–8631 (1997).
415. Arjona, O., Medel, R., Rojas, J., Costa, A. M. & Vilarrasa, J. Chemoselective protection of thiols versus alcohols and phenols. The Tosvinyl group. *Tetrahedron Lett.* **44,** 6369–6373 (2003).
416. Trofimov, B. A. *et al.* Synthesis of Functionalized L-Cysteine and L-Methionine by Reaction with Electron-Deficient Acetylenes. *Synthesis* 3136–3142 (2009).
417. Shiu, H.-Y., Chong, H.-C., Leung, Y.-C., Wong, M.-K. & Che, C.-M. A Highly Selective FRET-Based Fluorescent Probe for Detection of Cysteine and Homocysteine. *Chem. Eur. J.* **16,** 3308–3313 (2010).
418. Hong, V., Kislukhin, A. A. & Finn, M. G. Thiol-Selective Fluorogenic Probes for Labeling and Release. *J. Am. Chem. Soc.* **131,** 9986–9994 (2009).
419. Dziadek, S., Jacques, S. & Bundle, D. R. A novel linker methodology for the synthesis of tailored conjugate vaccines composed of complex carbohydrate antigens and specific TH-cell peptide epitopes. *Chemistry* **14,** 5908–5917 (2008).
420. Jones, M. W. *et al.* Phosphine-mediated one-pot thiol-ene 'click' approach to polymer-protein conjugates. *Chem. Commun.* 5272–5274 (2009).
421. Haase, C. & Seitz, O. Internal Cysteine Accelerates Thioester-Based Peptide Ligation. *Justus Liebigs Ann. Chem.* 2096–2101 (2009).
422. Hansen, F. K. *et al.* Microwave-assisted chemical ligation of S-acyl peptides containing non-terminal cysteine residues. *Org. Biomol. Chem.* **9,** 7162–7167 (2011).
423. Ha, K. *et al.* Long-Range Intramolecular S → N Acyl Migration: A Study of the Formation of Native Peptide Analogues *via* 13-, 15-, and 16-Membered Cyclic Transition States. *J. Org. Chem.* **77,** 2637–2648 (2012).
424. Bol'shakov, O. *et al.* S- to N-Acyl transfer in S-acylcysteine isopeptides *via* 9-, 10-, 12-, and 13-membered cyclic transition states. *J. Pept. Sci.* **18,** 704–709 (2012).
425. Panda, S. S. *et al.* Study of chemical ligation *via* 17-, 18- and 19-membered cyclic transition states. *Chem Biol Drug Design* **80,** 821–827 (2012).
426. Monbaliu, J.-C. M., Dive, G., Stevens, C. V. & Katritzky, A. R. Governing Parameters of Long-Range Intramolecular S-to-NAcyl Transfers within (S)-Acyl Isopeptides. *J. Chem. Theory Comput.* **9,** 927–934 (2013).
427. Gevaert, K., Van Damme, P., Martens, L. & Vandekerckhove, J. Diagonal reverse-phase chromatography applications in peptide-centric proteomics: ahead of catalogue-omics? *Anal. Biochem.* **345,** 18–29 (2005).
428. Koshland, D. E., Karkhanis, Y. D. & Latham, H. G. An Environmentally-Sensitive Reagent with Selectivity for the Tryptophan Residue in Proteins. *J. Am. Chem. Soc.* **86,** 1448–1450 (1964).
429. Savige, W. E. & Fontana, A. [42] Modification of tryptophan to oxindolylalanine by dimethyl sulfoxide-hydrochloric acid. **47,** 442–453 (1977).
430. Bell, J. E., Castellino, F. J., Trayer, I. P. & Hill, R. L. Modification of bovine alpha-lactalbumin with N-bromosuccinimide and 2-hydroxy-5-nitrobenzylbromide. *J. Biol. Chem.* **250,** 7579–7585 (1975).
431. Horton, H. R. & Koshland, D. E. A Highly Reactive Colored Reagent with Selectivity for the Tryptophan Residue in Proteins. 2-Hydroxy-5-nitrobenzyl Bromide1. *J. Am. Chem. Soc.* **87,** 1126–1132 (1965).
432. Strohalm, M., Šantrůček, J., Hynek, R. & Kodíček, M. Analysis of tryptophan surface accessibility in proteins by MALDI-TOF mass spectrometry. *Biochem. Biophys. Res. Commun.* **323,** 1134–1138 (2004).
433. Teuber, H. J., Glosauer, O. & Hochmuth, U. Ein neues Kondensationsprodukt aus Tryptamin und Acetessigaldehyd. *Eur. J. Inorg. Chem.* **97,** 557–562 (1964).
434. Foettinger, A., Melmer, M., Leitner, A. & Lindner, W. Reaction of the Indole Group with Malondialdehyde: Application for the Derivatisation of Tryptophan Residues in Peptides. *Bioconjugate Chem.* **18,** 1678–1683 (2007).
435. Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat. Protocols* **7,** 1511–1522 (2012).

436. Foettinger, A., Leitner, A. & Lindner, W. Derivatisation of arginine residues with malondialdehyde for the analysis of peptides and protein digests by LC-ESI-MS/MS. *J. Mass Spectrom.* **41,** 623–632 (2006).
437. Foettinger, A., Leitner, A. & Lindner, W. Selective enrichment of tryptophan-containing peptides from protein digests employing a reversible derivatisation with malondialdehyde and solid-phase capture on hydrazide beads. *J. Proteome Res.* **6,** 3827–3834 (2007).
438. Antos, J. M. & Francis, M. B. Selective Tryptophan Modification with Rhodium Carbenoids in Aqueous Solution. *J. Am. Chem. Soc.* **126,** 10256–10257 (2004).
439. Ye, T. & McKervey, M. A. Organic Synthesis with .alpha.-Diazo Carbonyl Compounds. *Macromolecules* **94,** 1091–1160 (1994).
440. Antos, J. M., McFarland, J. M., Iavarone, A. T. & Francis, M. B. Chemoselective Tryptophan Labeling with Rhodium Carbenoids at Mild pH. *J. Am. Chem. Soc.* **131,** 6301–6308 (2009).
441. Popp, B. V. & Ball, Z. T. Structure-Selective Modification of Aromatic Side Chains with Dirhodium Metallopeptide Catalysts. *J. Am. Chem. Soc.* **132,** 6660–6662 (2010).
442. Woolfson, D. N. The design of coiled-coil structures and assemblies. *Adv. Protein Chem.* **70,** 79–112 (2005).
443. Chen, Z., Popp, B. V., Bovet, C. L. & Ball, Z. T. Site-specific protein modification with a dirhodium metallopeptide catalyst. *ACS Chem. Biol.* **6,** 920–925 (2011).
444. Howland, J. L. *The Encyclopedia of Molecular Biology*. (Biochemical Education, 1995).
445. Li, S. & Hong, M. Protonation, Tautomerisation, and Rotameric Structure of Histidine: A Comprehensive Study by Magic-Angle-Spinning Solid-State NMR. *J. Am. Chem. Soc.* **133,** 1534–1544 (2011).
446. Li, X., Ma, H., Dong, S., Duan, X. & Liang, S. Selective labeling of histidine by a designed fluorescein-based probe. *Talanta* **62,** 367–371 (2004).
447. Li, X., Ma, H., Nie, L., Sun, M. & Xiong, S. A novel fluorescent probe for selective labeling of histidine. *Anal. Chim. Acta* **515,** 255–260 (2004).
448. Takaoka, Y., Tsutsumi, H., Kasagi, N., Nakata, E. & Hamachi, I. One-Pot and Sequential Organic Chemistry on an Enzyme Surface to Tether a Fluorescent Probe at the Proximity of the Active Site with Restoring Enzyme Activity. *J. Am. Chem. Soc.* **128,** 3273–3280 (2006).
449. Wakabayashi, H. *et al.* Affinity-Labeling-Based Introduction of a Reactive Handle for Natural Protein Modification. *Chem. Asian J.* **3,** 1134–1139 (2008).
450. Miyaura, N. & Suzuki, A. Palladium-Catalyzed Cross-Coupling Reactions of Organoboron Compounds. *Macromolecules* **95,** 2457–2483 (1995).
451. Meldal, M. & Tornøe, C. W. Cu-Catalyzed Azide−Alkyne Cycloaddition. *Macromolecules* **108,** 2952–3015 (2008).
452. Hearon, J. Z. The configuration of cobaltodihistidine and oxy-bis (cobaltodihistidine). *J. Natl. Cancer Inst.* **9,** 1–11 (1948).
453. Wang, X., Jia, J., Huang, Z., Zhou, M. & Fei, H. Luminescent Peptide Labeling Based on a Histidine-Binding Iridium(III) Complex for Cell Penetration and Intracellular Targeting Studies. *Chem. Eur. J.* **17,** 8028–8032 (2011).
454. Ma, D.-L. *et al.* A Highly Selective Luminescent Switch-On Probe for Histidine/Histidine-Rich Proteins and Its Application in Protein Staining. *Angew. Chem. Int. Ed. (English)* **47,** 3735–3739 (2008).
455. K Uchida, E. R. S. Modification of histidine residues in proteins by reaction with 4-hydroxynonenal. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 4544–4548 (1992).
456. Zamora, R., Alaiz, M. & Hidalgo, F. J. Modification of Histidine Residues by 4,5-Epoxy-2-alkenals. *Chem. Res. Toxicol.* **12,** 654–660 (1999).
457. Hidalgo, F. J. & Zamora, R. Fluorescent pyrrole products from carbonyl-amine reactions. *J. Biol. Chem.* **268,** 16190–16197 (1993).
458. Uchida, K. & Stadtman, E. R. Modification of histidine residues in proteins by reaction with 4-hydroxynonenal. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 4544–4548 (1992).
459. Chand Bollineni, R., Fedorova, M. & Hoffmann, R. Qualitative and quantitative evaluation of derivatisation reagents for different types of protein-bound carbonyl groups. *Analyst* **138,** 5081–5088 (2013).
460. Wylie, D. *et al.* Carboxyalkylated Histidine Is a pH-Dependent Product of Pegylation with SC-PEG. *Pharm. Res.* **18,** 1354–1360–1360 (2001).
461. Nishino, T., Massey, V. & Williams, C. H. Chemical modifications of D-amino acid oxidase. Evidence for active site histidine, tyrosine, and arginine residues. *J. Biol. Chem.* **255,** 3610–3616 (1980).
462. Williams, E. B., Krishnaswamy, S. & Mann, K. G. Zymogen/enzyme discrimination using peptide chloromethyl ketones. *J. Biol. Chem.* **264,** 7536–7545 (1989).
463. Fruchter, R. G. & Crestfield, A. M. The specific alkylation by iodoacetamide of histidine-12 in the active site of ribonuclease. *J. Biol. Chem.* **242,** 5807–5812 (1967).
464. Lin, M. C., Stein, W. H. & Moore, S. Further studies on the alkylation of the histidine residues at the active site of pancreatic ribonuclease. *J. Biol. Chem.* **243,** 6167–6170 (1968).
465. Inagami, T. T. & Hatano, H. H. Effect of alkylguanidines on the inactivation of trypsin by alkylation and phosphorylation. *J. Biol. Chem.* **244,** 1176–1182 (1969).
466. Liu, T. Y. Demonstration of the presence of a histidine residue at the active site of streptococcal proteinase. *J. Biol. Chem.* **242,** 4029–4032 (1967).
467. Pauly, H. Über die Einwirkung von Diazoniumverbindungen auf Imidazole. *H.-S. Z. Physiol. Chem.* 159–160 (1905).
468. Pauly, H. Zur Kenntnis der Diazoreaktion des Eiweißes. *H.-S. Z. Physiol. Chem.* **94,** 284–290 (1915).
469. Dominici, P., Tancini, B. & Borri Voltattorni, C. Chemical modification of pig kidney 3,4-dihydroxyphenylalanine decarboxylase with diethyl pyrocarbonate. Evidence for an essential histidyl residue. *J. Biol. Chem.* **260,** 10583–10589 (1985).
470. Marks, F., Klingmüller, U. & Müller-Decker, K. *Cellular signal processing : an introduction to the molecular mechanisms of signal transduction*. (lavoisier.eu, 2009).
471. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141,** 1117–1134 (2010).
472. Koshi, Y. *et al.* Target-Specific Chemical Acylation of Lectins by Ligand-Tethered DMAP Catalysts. *J. Am. Chem. Soc.* **130,** 245–251 (2008).
473. Håkansson, S., Viljanen, J. & Broo, K. S. Programmed Delivery of Novel Functional Groups to the Alpha Class Glutathione Transferases. *Biochemistry* **42,** 10260–10268 (2003).
474. Riordan, J. F., Sokolovsky, M. & Vallee, B. L. *J. Am. Chem. Soc.* **88,** 4104–4105 (1966).
475. Jiao, K., Mandapati, S., Skipper, P. L., Tannenbaum, S. R. & Wishnok, J. S. Site-selective nitration of tyrosine in human serum albumin by peroxynitrite. *Anal. Biochem.* **293,** 43–52 (2001).
476. Shirai, T. *et al.* Crystal structure of a conger eel galectin (congerin II) at 1.45A resolution: implication for the accelerated evolution of a new ligand-binding site following gene duplication. *J. Mol. Biol.* **321,** 879–889 (2002).
477. Neurath, H. & Kenner, R. A. Activity and fluorescent derivatives of aminotyrosyl trypsin and trypsinogen. *Biochemistry* **10,** 551–557 (1971).
478. Nikov, G., Bhat, V., Wishnok, J. S. & Tannenbaum, S. R. Analysis of nitrated proteins by nitrotyrosine-specific affinity probes and mass spectrometry. *Anal. Biochem.* **320,** 214–222 (2003).
479. Kai, M., Ishida, J. & Ohkura, Y. High-performance liquid chromatographic determination of leucine-enkephalin-like peptide in rat brain by pre-column fluorescence derivatisation involving formylation followed by reaction with 1,2-diamino-4,5-dimethoxybenzene. *J. Chromatogr. B* **430,** 271–278 (1988).
480. Ishida, J., Kai, M. & Ohkura, Y. High-performance liquid chromatography of tyrosine-containing peptides by pre-column deri-

vatisation involving formylation followed by fluorescence reaction with 1,2-diamino-4,5-dimethoxybenzene. *J. Chromatogr. A* **356,** 171–177 (1986).
481. Trost, B. M. & Toste, F. D. A Catalytic Enantioselective Approach to Chromans and Chromanols. A Total Synthesis of (−)-Calanolides A and B and the Vitamin E Nucleus. *J. Am. Chem. Soc.* **120,** 9074–9075 (1998).
482. Trost, B. M. & Toste, F. D. Asymmetric O- and C-Alkylation of Phenols. *J. Am. Chem. Soc.* **120,** 815–816 (1998).
483. Tilley, S. & Francis, M. Tyrosine-selective protein alkylation using π-allylpalladium complexes. *J. Am. Chem. Soc.* **128,** 1080–1081 (2006).
484. Brown, K. C., Yang, S.-H. & Kodadek, T. Highly Specific Oxidative Crosslinking of Proteins Mediated by a Nickel-Peptide Complex. *Biochemistry* **34,** 4733–4739 (1995).
485. Brown, K. C., Yu, Z., Burlingame, A. L. & Craik, C. S. Determining Protein−Protein Interactions by Oxidative Cross-Linking of a Glycine-Glycine-Histidine Fusion Protein. *Biochemistry* **37,** 4397–4406 (1998).
486. Fancy, D. A., Melcher, K., Johnston, S. A. & Kodadek, T. New chemistry for the study of multiprotein complexes: the six-histidine tag as a receptor for a protein crosslinking reagent. *Chem. Biol.* **3,** 551–559 (1996).
487. Fancy, D. A. & Kodadek, T. Site-directed oxidative protein crosslinking. *Tetrahedron* **53,** 11953–11960 (1997).
488. Gill, G., Richter-Rusli, A. A., Ghosh, M., Burrows, C. J. & Rokita, S. E. Nickel-Dependent Oxidative Cross-Linking of a Protein †. *Chem. Res. Toxicol.* **10,** 302–309 (1997).
489. Kim, K., Fancy, D. A., Carney, D. & Kodadek, T. Photoinduced Protein Cross-Linking Mediated by Palladium Porphyrins. *J. Am. Chem. Soc.* **121,** 11896–11897 (1999).
490. Fancy, D. A. & Kodadek, T. Chemistry for the analysis of protein-protein interactions: Rapid and efficient cross-linking triggered by long wavelength light. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 6020–6024 (1999).
491. Fancy, D. A. *et al.* Scope, limitations and mechanistic aspects of the photo-induced cross-linking of proteins by water-soluble metal complexes. *Chem. Biol.* **7,** 697–708 (2000).
492. Person, M. D., Brown, K. C., Mahrus, S., Craik, C. S. & Burlingame, A. L. Novel inter-protein cross-link identified in the GGH-ecotin D137Y dimer. *Protein Sci.* **10,** 1549–1562 (2009).
493. Kodadek, T., Duroux-Richard, I. & Bonnafous, J. Techniques: Oxidative cross-linking as an emergent tool for the analysis of receptor-mediated signalling events. *Trends Pharmacol. Sci.* **26,** 210–217 (2005).
494. Antos, J. & Francis, M. Transition metal catalyzed methods for site-selective protein modification. *Curr. Opin. Chem. Biol.* **10,** 253–262 (2006).
495. Seim, K. L., Obermeyer, A. C. & Francis, M. B. Oxidative Modification of Native Protein Residues Using Cerium(IV) Ammonium Nitrate. *J. Am. Chem. Soc.* **133,** 16970–16976 (2011).
496. Duroux-Richard, I. *et al.* Crosslinking photosensitized by a ruthenium chelate as a tool for labeling and topographical studies of G-protein-coupled receptors. *Chem. Biol.* **12,** 15–24 (2005).
497. Denison, C. Toward a General Chemical Method for Rapidly Mapping Multi-Protein Complexes. *J. Proteome Res.* **3,** 417–425 (2004).
498. Meunier, S., Strable, E. & Finn, M. G. Crosslinking of and coupling to viral capsid proteins by tyrosine oxidation. *Chem. Biol.* **11,** 319–326 (2004).
499. Higgins, H. G. & Harrington, K. J. Reaction of amino acids and proteins with diazonium compounds. II. Spectra of protein derivatives. *Arch. Biochem. Biophys.* **85,** 409–425 (1959).
500. Roglans, A., Pla-Quintana, A. & Moreno-Mañas, M. Diazonium Salts as Substrates in Palladium-Catalyzed Cross-Coupling Reactions. *Chem. Rev.* **106,** 4622–4643 (2006).
501. Hooker, J. M., Kovacs, E. W. & Francis, M. B. Interior Surface Modification of Bacteriophage MS2. *J. Am. Chem. Soc.* **126,** 3718–3719 (2004).
502. Hooker, J. M., Datta, A., Botta, M., Raymond, K. N. & Francis, M. B. Magnetic Resonance Contrast Agents from Viral Capsid Shells: A Comparison of Exterior and Interior Cargo Strategies. *Nano Lett.* **7,** 2207–2210 (2007).
503. Schlick, T. L., Ding, Z., Kovacs, E. W. & Francis, M. B. Dual-Surface Modification of the Tobacco Mosaic Virus. *J. Am. Chem. Soc.* **127,** 3718–3723 (2005).
504. Jones, M. W. *et al.* Direct Peptide Bioconjugation/PEGylation at Tyrosine with Linear and Branched Polymeric Diazonium Salts. *J. Am. Chem. Soc.* **134,** 7406–7413 (2012).
505. Gavrilyuk, J., Ban, H., Nagano, M., Hakamata, W. & Barbas, C. F. Formylbenzene Diazonium Hexafluorophosphate Reagent for Tyrosine-Selective Modification of Proteins and the Introduction of a Bioorthogonal Aldehyde. *Bioconjugate Chem.* **23,** 2321–2328 (2012).
506. Fraenkel-Conrat, H. & Olcott, H. S. Reaction of formaldehyde with proteins VI. cross-linking of amino groups with phenol, imidazole, or indole groups. *J. Biol. Chem.* **174,** 827–843 (1948).
507. Joshi, N. S., Whitaker, L. R. & Francis, M. B. A Three-Component Mannich-Type Reaction for Selective Tyrosine Bioconjugation. *J. Am. Chem. Soc.* **126,** 15942–15943 (2004).
508. Romanini, D. & Francis, M. Attachment of peptide building blocks to proteins through tyrosine bioconjugation. *Bioconjugate Chem.* **19,** 153–157 (2008).
509. McFarland, J. M., Joshi, N. S. & Francis, M. B. Characterisation of a Three-Component Coupling Reaction on Proteins by Isotopic Labeling and Nuclear Magnetic Resonance Spectroscopy. *J. Am. Chem. Soc.* **130,** 7639–7644 (2008).
510. Guo, H.-M., Minakawa, M. & Tanaka, F. Fluorogenic Imines for Fluorescent Detection of Mannich-Type Reactions of Phenols in Water. *Chem. Rev.* **73,** 3964–3966 (2008).
511. Minakawa, M., Guo, H.-M. & Tanaka, F. Imines that React with Phenols in Water over a Wide pH Range. *J. Org. Chem.* **73,** 8669–8672 (2008).
512. Schroeter, S. H. Reaction of phenols with ethyl azodicarboxylate. *J. Org. Chem.* **34,** 4012–4015 (1969).
513. Mitchell, H. & Leblanc, Y. Amination of Arenes with Electron-Deficient Azodicarboxylates. *J. Org. Chem.* **59,** 682–687 (1994).
514. Leblanc, Y. & Boudreault, N. Para-Directed Amination of Electron-Rich Arenes with Bis(2,2,2-Trichloroethyl) Azodicarboxylate. *J. Org. Chem.* **60,** 4268–4271 (1995).
515. Yadav, J. S., Reddy, B. V. S., Kumar, G. M. & Madan, C. InCl3-SiO2 Catalyzed Electrophilic Amination of Arenes: A Facile and Rapid Synthesis of Aryl Hydrazides. *Synlett* **2001,** 1781–1783 (2001).
516. Yadav, J. S., Reddy, BVS, Veerendhar, G., Rao, R. S. & Nagaiah, K. Sc(OTf)3 catalyzed electrophilic amination of arenes: an expeditious synthesis of aryl hydrazides. *Chem. Lett.* **31,** 318–319 (2002).
517. Bombek, S., Lenaršič, R. & Kočevar, M. ZrCl4-promoted halogen migration during an electrophilic amination of halogenated phenols. *Chem. Commun.* 1494–1495 (2002).
518. Kinart, W. J. & Kinart, C. M. Studies on the catalysis of the reaction of organotin phenoxides with diethyl azodicarboxylate by lithium perchlorate. *J. Organomet. Chem.* **665,** 233–236 (2003).
519. Chee, G. L. Efficient Synthesis of Bifenazate. *Synth. Comm.* **36,** 2151–2156 (2006).
520. Brandes, S., Bella, M., Kjærsgaard, A. & Jørgensen, K. A. Chirally Aminated 2-Naphthols—Organocatalytic Synthesis of Non-Biaryl Atropisomers by Asymmetric Friedel–Crafts Amination. *Angew. Chem. Int. Ed. (English)* **45,** 1147–1151 (2006).
521. Desimoni, G., Faita, G., Righetti, P. P., Sfulcini, A. & Tsyganov, D. Solvent effect in pericyclic reactions. IX. The ene reaction. *Tetrahedron* **50,** 1821–1832 (1994).
522. Ban, H., Gavrilyuk, J. & Barbas, C. F. Tyrosine Bioconjugation through Aqueous Ene-Type Reactions: A Click-Like Reaction for Tyrosine. *J. Am. Chem. Soc.* **132,** 1523–1525 (2010).
523. Ban, H. *et al.* Facile and stabile linkages through tyrosine: bioconjugation strategies with the tyrosine-click reaction. *Bioconjugate Chem.* **24,** 520–532 (2013).

524. Wong, O. Y., Pradeepkumar, P. I. & Silverman, S. K. DNA-catalyzed covalent modification of amino acid side chains in tethered and free peptide substrates. *Biochemistry* **50,** 4741–4749 (2011).
525. Borders, C. L. *et al.* A structural role for arginine in proteins: Multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Sci.* **3,** 541–548 (1994).
526. Shimoni, L. & Glusker, J. P. Hydrogen bonding motifs of protein side chains: Descriptions of binding of arginine and amide groups. *Protein Sci.* **4,** 65–74 (1995).
527. Dawson, R. M. C. *Data for Biochemical Research*. (Oxford University Press, USA, 1989).
528. McMahon, B. H. *et al.* FTIR studies of internal proton transfer reactions linked to inter-heme electron transfer in bovine cytochrome c oxidase. *Biochim. Biophys. Acta* **1655,** 321–331 (2004).
529. Xiao, Y., Hutson, M. S., Belenky, M., Herzfeld, J. & Braiman, M. S. Role of Arginine-82 in Fast Proton Release during the Bacteriorhodopsin Photocycle: A Time-Resolved FT-IR Study of Purple Membranes Containing 15N-Labeled Arginine. *Biochemistry* **43,** 12809–12818 (2004).
530. Leitner, A. & Lindner, W. Functional Probing of Arginine Residues in Proteins Using Mass Spectrometry and an Arginine-Specific Covalent Tagging Concept. *Anal. Chem.* **77,** 4481–4488 (2005).
531. Takahashi, K. The reaction of phenylglyoxal with arginine residues in proteins. *J. Biol. Chem.* **243,** 6171–6179 (1968).
532. R Manohar, N. A. R. Identification of active-site residues of sheep liver serine hydroxymethyltransferase. *Biochem. J.* **224,** 703–707 (1984).
533. Borders, C. L., Saunders, J. E., Blech, D. M. & Fridovich, I. Essentiality of the active-site arginine residue for the normal catalytic activity of Cu,Zn superoxide dismutase. *Biochem. J.* **230,** 771–776 (1985).
534. Bailin, G. Reaction of phenylglyoxal with chicken gizzard myosin. *Biochem. Int.* **11,** 161–170 (1985).
535. Varalakshmi, K., Savithri, H. S. & Rao, N. A. Identification of amino acid residues essential for enzyme activity of sheep liver 5,10-methylenetetrahydrofolate reductase. *Biochem. J.* **236,** 295–298 (1986).
536. Eun, H. M. Arginine modification by phenylglyoxal and (p-hydroxyphenyl) glyoxal: reaction rates and intermediates. *Biochem. Int.* 6171–6179 (1988).
537. Yamasaki, R. B., Vega, A. & Feeney, R. E. Modification of available arginine residues in proteins by p-hydroxyphenylglyoxal. *Anal. Biochem.* **109,** 32–40 (1980).
538. Yamasaki, R. B., Shimer, D. A. & Feeney, R. E. Colorimetric determination of arginine residues in proteins by p-nitrophenylglyoxal. *Anal. Biochem.* **111,** 220–226 (1981).
539. Strévey, J., Vachon, V., Beaumier, B., Giroux, S. & Béliveau, R. Characterisation of essential arginine residues implicated in the renal transport of phosphate and glucose. *Biochim. Biophys. Acta* **1106,** 110–116 (1992).
540. Mukouyama, E. B., Hirose, T. & Suzuki, H. Chemical modification of L-phenylalanine oxidase from Pseudomonas sp. P-501 by phenylglyoxal. Identification of one essential arginyl residue. *J. Biochem.* **123,** 1097–1103 (1998).
541. Cheung, S.-T. & Fonda, M. L. Reaction of phenylglyoxal with arginine. The effect of buffers and pH. *Biochem. Biophys. Res. Commun.* **90,** 940–947 (1979).
542. Yankeelov, J. A., Jr. Modification of arginine in proteins by oligomers of 2, 3-butanedione. *Biochemistry* **9,** 2433–2439 (1970).
543. Yankeelov, J. A., Jr. [52] Modification of arginine by diketones. *Methods Enzymol.* **25,** 566–579 (1972).
544. Toi, K., Bynum, E., Norris, E. & Itano, H. A. Studies on the chemical modification of arginine. I. The reaction of 1,2-cyclohexanedione with arginine and arginyl residues of proteins. *J. Biol. Chem.* **242,** 1036–1043 (1967).
545. Riordan, J. F. Functional arginyl residues in carboxypeptidase A. Modification with butanedione. *Biochemistry* **12,** 3915–3923 (1973).
546. Patthy, L. & Smith, E. L. Reversible modification of arginine residues. Application to sequence studies by restriction of tryptic hydrolysis to lysine residues. *J. Biol. Chem.* **250,** 557–564 (1975).
547. Foettinger, A., Leitner, A. & Lindner, W. Solid-phase capture and release of arginine peptides by selective tagging and boronate affinity chromatography. *J. Chromatogr. A* **1079,** 187–196 (2005).
548. Ngo, T. T., Yam, C. F., Lenhoff, H. M. & Ivy, J. p-Azidophenylglyoxal. A heterobifunctional photoactivable cross-linking reagent selective for arginyl residues. *J. Biol. Chem.* **256,** 11313–11318 (1981).
549. Politz, S. M., Noller, H. F. & McWhirter, P. D. Ribonucleic acid-protein cross-linking in Escherichia coli ribosomes: (4-azidophenyl)glyoxal, a novel heterobifunctional reagent. *Biochemistry* **20,** 372–378 (1981).
550. Gauthier, M. A. & Klok, H.-A. Arginine-Specific Modification of Proteins with Polyethylene Glycol. *Biomacromolecules* **12,** 482–493 (2011).
551. Takahashi, K. Specific modification of arginine residues in proteins with ninhydrin. *J. Biochem.* **80,** 1173–1176 (1976).
552. Pischetsrieder, M. Reaction of l-Ascorbic Acid with l-Arginine Derivatives. *J. Agric. Food Chem.* **44,** 2081–2085 (1996).
553. Hoare, D. G. & Koshland, D. E., Jr. A Procedure for the Selective Modification of Carboxyl Groups in Proteins. *J. Am. Chem. Soc.* **88,** 2057–2058 (1966).
554. Hoare, D. G. & Koshland, D. E. A method for the quantitative modification and estimation of carboxylic acid groups in proteins. *J. Biol. Chem.* **242,** 2447–2453 (1967).
555. Khorana, H. G. The Chemistry of Carbodiimides. *Chem. Rev.* **53,** 145–166 (1953).
556. Renthal, R., Cothran, M., Dawson, N. & Harris, G. J. Fluorescent labeling of bacteriorhodopsin: implications for helix connections. *Biochim. Biophys. Acta* **897,** 384–394 (1987).
557. Das, A. & Ljungdahl, L. G. Composition and primary structure of the F1F0 ATP synthase from the obligately anaerobic bacterium Clostridium thermoaceticum. *J. Bacteriol.* **179,** 3746–3755 (1997).
558. Cook, G. M. *et al.* Purification and biochemical characterisation of the F1Fo-ATP synthase from thermoalkaliphilic Bacillus sp. strain TA2.A1. *J. Bacteriol.* **185,** 4442–4449 (2003).
559. Sheehan, J. C. & Hlavka, J. J. The Use of Water-Soluble and Basic Carbodiimides in Peptide Synthesis. *J. Org. Chem.* **21,** 439–441 (1956).
560. Sheehan, J. C. & Hlavka, J. J. The cross-linking of gelatin using a water-soluble carbodiimide. *J. Am. Chem. Soc.* **79,** 4528–4529 (1957).
561. Carraway, K. L. & Koshland, D. E., Jr. [56] Carbodiimide modification of proteins. **25,** 616–623 (1972).
562. Goodfriend, T. L., Levine, L. & Fasman, G. D. Antibodies to bradykinin and angiotensin: a use of carbodiimides in immunology. *Science* **144,** 1344–1346 (1964).
563. Xia, N. *et al.* Probing of EDC/NHSS-Mediated Covalent Coupling Reaction by the Immobilisation of Electrochemically Active Biomolecules. *Int. J. Electrochem. Sci* **8,** 2459–2467 (2013).
564. Gilles, M. A., Hudson, A. Q. & Borders, C. L. Stability of water-soluble carbodiimides in aqueous solution. *Anal. Biochem.* **184,** 244–248 (1990).
565. Valeur, E. & Bradley, M. Amide bond formation: beyond the myth of coupling reagents. *Chem. Soc. Rev.* **38,** 606–631 (2009).
566. Woodward, R. B., Olofson, R. A. & Mayer, H. *J. Am. Chem. Soc.* **83,** 1010–1012 (1961).
567. Bodlaender, P., Feinstein, G. & Shaw, E. Use of isoxazolium salts for carboxyl group modification in proteins. Trypsin. *Biochemistry* **8,** 4941–4949 (1969).
568. Erlanger, B. F., Vratsanos, S. M., Wassermann, N. & Cooper, A. G. Specific and reversible inactivation of pepsin. *J. Biol. Chem.* **240,** PC3447–PC3448 (1965).

569. Gross, E. & Morell, J. L. Evidence for an active carboxyl group in pepsin. *J. Biol. Chem.* **241,** 3638–3639 (1966).
570. Takahashi, K., Stein, W. H. & Moore, S. The identification of a glutamic acid residue as part of the active site of ribonuclease T1. *J. Biol. Chem.* **242,** 4682–4690 (1967).
571. Vogt, W. Oxidation of methionyl residues in proteins: tools, targets, and reversal. *Free Radic. Biol. Med.* **18,** 93–105 (1995).
572. Glazer, A. N. Specific chemical modification of proteins. *Annu. Rev. Biochem.* **39,** 101–130 (1970).
573. Lindorff-Larsen, K. & Winther, J. R. Thiol Alkylation below Neutral pH. *Anal. Biochem.* **286,** 308–310 (2000).
574. Kramer, J. R. & Deming, T. J. Reversible chemoselective tagging and functionalisation of methionine containing peptides. *Chem. Commun.* **49,** 5144–5146 (2013).
575. Toennies, G. Sulfonium reactions of methionine and their possible metabolic significance. *J. Biol. Chem.* **132,** 455–456 (1940).
576. Toennies, G. & Kolb, J. J. Methionine Studies. VII. Sulfonium Derivatives1. *J. Am. Chem. Soc.* **67,** 849–851 (1945).
577. Ho, T.-L. Hard soft acids bases (HSAB) principle and organic chemistry. *Macromolecules* **75,** 1–20 (1975).
578. Reid, E. Relative Importance of Free α-and **ε**-Amino Groups for the Biological Activity of the Growth Hormone. *Nature* **1,** 955 (1951).
579. Kinstler, O., Molineux, G., Treuheit, M., Ladd, D. & Gegg, C. Mono-N-terminal poly(ethylene glycol)–protein conjugates. *Adv. Drug Deliv. Rev.* **54,** 477–485 (2002).
580. Gaudriault, G. & Vincent, J. P. Selective labeling of alpha- or epsilon-amino groups in peptides by the Bolton-Hunter reagent. *Peptides* **13,** 1187–1192 (1992).
581. Sélo, I., Négroni, L., Créminon, C., Grassi, J. & Wal, J. M. Preferential labeling of α-amino N-terminal groups in peptides by biotin: application to the detection of specific anti-peptide antibodies by enzyme immunoassays. *J. Immunol. Methods.* **199,** 127–138 (1996).
582. Yamaguchi, M. *et al.* High-Throughput Method for N-Terminal Sequencing of Proteins by MALDI Mass Spectrometry. *Anal. Chem.* **77,** 645–651 (2005).
583. Xu, G. & Jaffrey, S. R. N-CLAP: global profiling of N-termini by chemoselective labeling of the alpha-amine of proteins. *Cold. Spring. Harb. Protoc.* **2010,** 1–15 (pdb.prot5528) (2010).
584. Bertaccini, D. *et al.* An improved stable isotope N-terminal labeling approach with light/heavy TMPP to automate proteogenomics data validation: dN-TOP. *J. Proteome Res.* **12,** 3063−3070 (2013).
585. Chan, W.-K., Ho, C.-M., Wong, M.-K. & Che, C.-M. Oxidative Amide Synthesis and N-Terminal α-Amino Group Ligation of Peptides in Aqueous Medium. *J. Am. Chem. Soc.* **128,** 14796–14797 (2006).
586. Chan, A. O.-Y. *et al.* Modification of N-Terminal α-Amino Groups of Peptides and Proteins Using Ketenes. *J. Am. Chem. Soc.* **134,** 2589–2598 (2012).
587. Dixon, H. & Weitkamp, L. R. Conversion of the N-terminal serine residue of corticotrophin into glycine. *Biochem. J.* **34,** 462–468 (1962).
588. Jencks, W. P. Studies on the mechanism of oxime and semicarbazone formation1. *J. Am. Chem. Soc.* **81,** 475–481 (1959).
589. Kalia, J. & Raines, R. T. Hydrolytic Stability of Hydrazones and Oximes. *Angew. Chem. Int. Ed. (English)* **47,** 7523–7526 (2008).
590. Metzler, D. E. & Snell, E. E. Some Transamination Reactions Involving Vitamin B61. *J. Am. Chem. Soc.* **74,** 979–983 (1952).
591. Cennamo, C. Non-enzymatic transamination between peptides and pyridoxal. *Naturwissenschaften* **41,** 39–39 (1954).
592. Cennamo, C., Carafoli, B. & Bonetti, E. P. Non-enzymatic Transamination between Peptides and Pyridoxal. Isolation of the 2,4-Dinitrophenylhydrazones of Some Ketopeptides. *J. Am. Chem. Soc.* **78,** 3523–3527 (1956).
593. Dixon, H. Transamination of peptides. *Biochem. J.* **92,** 661–666 (1964).
594. Dixon, H. & Moret, V. Removal of the N-terminal residue of a protein after transamination. *Biochem. J.* **94,** 463–469 (1965).
595. Dixon, H. B. F. N-terminal modification of proteins - A review. *J. Protein Chem.* **3,** 99–108 (1984).
596. Mix, H. & Wilcke, F. W. Nichtenzymatische Reaktionen zwischen α-Amino-und α-Ketosäuren, II. Durch Kupfer (II)-Ionen und Pyridin katalysierte Umsetzungen zwischen α-Aminosäuren und Pyruvat. *H.-S. Z. Physiol. Chem.* **318,** 148–158 (1960).
597. Van Heyningen, S. & Dixon, H. B. Catalysis of transamination by acetate. *Biochem. J.* **114,** 70–71 (1969).
598. Dixon, H. B. & Fields, R. [33] Specific modification of NH(2)-terminal residues by transamination. *Methods Enzymol.* **25,** 409–419 (1972).
599. Gilmore, J. M., Scheck, R. A., Esser-Kahn, A. P., Joshi, N. S. & Francis, M. B. N-Terminal Protein Modification through a Biomimetic Transamination Reaction. *Angew. Chem. Int. Ed. (English)* **45,** 5307–5311 (2006).
600. Witus, L. S. & Francis, M. Site-Specific Protein Bioconjugation *via* a Pyridoxal 5'-Phosphate-Mediated N-Terminal Transamination Reaction. *Curr. Protoc. Chem. Biol.* **2,** 125–134 (2010).
601. Scheck, R. A., Dedeo, M. T., Iavarone, A. T. & Francis, M. B. Optimisation of a biomimetic transamination reaction. *J. Am. Chem. Soc.* **130,** 11762–11770 (2008).
602. Witus, L. S. *et al.* Identification of Highly Reactive Sequences For PLP-Mediated Bioconjugation Using a Combinatorial Peptide Library. *J. Am. Chem. Soc.* **132,** 16812–16817 (2010).
603. Scheck, R. A. & Francis, M. B. Regioselective labeling of antibodies through N-terminal transamination. *ACS Chem. Biol.* **2,** 247–251 (2007).
604. Witus, L. S. *et al.* Site-Specific Protein Transamination Using N-Methylpyridinium-4-carboxaldehyde. *J. Am. Chem. Soc.* **135,** 17223–17229 (2013).
605. Zhang, M., Zhang, X., Li, J., Guo, Q. & Xiao, Q. A New Pyridoxal Derivative for Transamination of N-Terminus of Proteins. *Chin. J. Chem.* **29,** 1715–1720 (2011).
606. Dettin, M. *et al.* Chemoselective Surface Immobilisation of Proteins through a Cleavable Peptide. *Bioconjugate Chem.* **22,** 1753–1757 (2011).
607. Crochet, A. P., Kabir, M. M., Francis, M. B. & Paavola, C. D. Site-selective dual modification of periplasmic binding proteins for sensing applications. *Biosens. Bioelectron.* **26,** 55–61 (2010).
608. Gao, W. *et al.* In situ growth of a stoichiometric PEG-like conjugate at a protein's N-terminus with significantly improved pharmacokinetics. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 15231–15236 (2009).
609. Carrico, Z. M. *et al.* N-Terminal labeling of filamentous phage to create cancer marker imaging agents. *ACS Nano* **6,** 6675–6680 (2012).
610. Palaniappan, K. K. *et al.* Molecular Imaging of Cancer Cells Using a Bacteriophage-Based 129Xe NMR Biosensor. *Angew. Chem. Int. Ed. (English)* **125,** 4949–4953 (2013).
611. Sonomura, K., Kuyama, H., Matsuo, E.-I., Tsunasawa, S. & Nishimura, O. A method for terminus proteomics: selective isolation and labeling of N-terminal peptide from protein through transamination reaction. *Bioorg. Med. Chem. Lett.* **19,** 6544–6547 (2009).
612. Sonomura, K. *et al.* Selective isolation of N-blocked peptide by combining AspN digestion, transamination, and tosylhydrazine glass treatment. *Anal. Biochem.* **410,** 214–223 (2011).
613. Han, M.-J., Xiong, D.-C. & Ye, X.-S. Enabling Wittig reaction on site-specific protein modification. *Chem. Commun.* **48,** 11079–11081 (2012).
614. Sasaki, T., Kodama, K., Suzuki, H., Fukuzawa, S. & Tachibana, K. N-terminal labeling of proteins by the Pictet–Spengler reaction. *Bioorg. Med. Chem. Lett.* **18,** 4550–4553 (2008).
615. Agarwal, P., van der Weijden, J., Sletten, E. M., Rabuka, D. & Bertozzi, C. R. A Pictet-Spengler ligation for protein chemical modification. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 46–51 (2013).
616. Waller, J. P. & Dixon, H. B. Selective acetylation of the terminal amino group of corticotrophin. *Biochem. J.* **75,** 320–328 (1960).
617. Brenner, M., Zimmermann, J. P., Wehrmüller, J., Quitt, P. &

Photaki, I. Eine neue Umlagerungsreaktion und ein neues Prinzip zum Aufbau von Peptidketten. *Experientia* **11,** 397–399 (1955).
618. Liu, C. F. C. & Tam, J. P. J. Peptide segment ligation strategy without use of protecting groups. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 6584–6588 (1994).
619. Liu, C.-F. & Tam, J. P. Chemical ligation approach to form a peptide bond between unprotected peptide segments. Concept and model study. *J. Am. Chem. Soc.* **116,** 4149–4153 (1994).
620. Tam, J. P., Rao, C., Liu, C. F. & Shao, J. Specificity and formation of unusual amino acids of an amide ligation strategy for unprotected peptides. *Int. J. Pept. Protein Res.* **45,** 209–216 (1995).
621. Miao, Z. & Tam, J. P. Bidirectional Tandem Pseudoproline Ligations of Proline-Rich Helical Peptides. *J. Am. Chem. Soc.* **122,** 4253–4260 (2000).
622. Tam, J. P., Yu, Q. & Yang, J. L. Tandem ligation of unprotected peptides through thiaprolyl and cysteinyl bonds in water. *J. Am. Chem. Soc.* **123,** 2487–2494 (2001).
623. Fields, R. & Dixon, H. A spectrophotometric method for the microdetermination of periodate. *Biochem. J.* **106,** 883–887 (1968).
624. Geoghegan, K. F. & Stroh, J. G. Site-directed conjugation of nonpeptide groups to peptides and proteins *via* periodate oxidation of a 2-amino alcohol. Application to modification at N-terminal serine. *Bioconjugate Chem.* **3,** 138–146 (1992).
625. Nicolet, B. H. & Shinn, L. A. *J. Am. Chem. Soc.* **61,** 1615–1615 (1939).
626. Zhang, L. & Tam, J. P. Thiazolidine Formation as a General and Site-Specific Conjugation Method for Synthetic Peptides and Proteins. *Anal. Biochem.* **233,** 87–93 (1996).
627. Gaertner, H. F. *et al.* Construction of protein analogs by site-specific condensation of unprotected fragments. *Bioconjugate Chem.* **3,** 262–268 (1992).
628. Chen, J., Zeng, W., Offord, R. & Rose, K. A Novel Method for the Rational Construction of Well-Defined Immunogens: The Use of Oximation To Conjugate Cholera Toxin B Subunit to a Peptide−Polyoxime Complex. *Bioconjugate Chem.* **14,** 614–618 (2003).
629. Geoghegan, K. F., Emery, M. J., Martin, W. H., McColl, A. S. & Daumy, G. O. Site-directed double fluorescent tagging of human renin and collagenase (MMP-1) substrate peptides using the periodate oxidation of N-terminal serine. An apparently general strategy for provision of energy-transfer substrates for proteases. *Bioconjugate Chem.* **4,** 537–544 (1993).
630. Mikolajczyk, S. D. *et al.* High Yield, Site-Specific Coupling of N-Terminally Modified .beta.-Lactamase to a Proteolytically Derived Single-Sulfhydryl Murine Fab'. *Bioconjugate Chem.* **5,** 636–646 (1994).
631. Rose, K. *et al.* Natural Peptides as Building Blocks for the Synthesis of Large Protein-like Molecules with Hydrazone and Oxime Linkages. *Bioconjugate Chem.* **7,** 552–556 (1996).
632. Gaertner, H. F. & Offord, R. E. Site-Specific Attachment of Functionalized Poly(ethylene glycol) to the Amino Terminus of Proteins. *Bioconjugate Chem.* **7,** 38–44 (1996).
633. Nardin, E. H. *et al.* Plasmodium falciparum polyoxymes: highly immunogenic synthetic vaccines constructed by chemoselective ligation of repeat B-cell epitopes and a universal T-cell epitope of CS protein. *Vaccine* **16,** 590–600 (1998).
634. Chelius, D. & Shaler, T. A. Capture of Peptides with N-Terminal Serine and Threonine: A Sequence-Specific Chemical Method for Peptide Mixture Simplification. *Bioconjugate Chem.* **14,** 205–211 (2003).
635. Zeng, Y., Ramya, T. N. C., Dirksen, A., Dawson, P. E. & Paulson, J. C. High-efficiency labeling of sialylated glycoproteins on living cells. *Nat. Methods* **6,** 207–209 (2009).
636. Dixon, H. B. The conversion of glyoxyloyl groups into glycyl groups and their formation from maleyl residues. *Biochem. J.* **107,** 124 (1968).
637. Mahdi, El, O. & Melnyk, O. α-Oxo Aldehyde or Glyoxylyl Group Chemistry in Peptide Bioconjugation. *Bioconjugate Chem.* **24,** 735–765 (2013).
638. Thomas, G. L. & Payne, R. J. Phosphate -assisted peptide ligation. *Chem. Commun.* 4260–4262 (2009).
639. Chen, J., Wang, P., Zhu, J., Wan, Q. & Danishefsky, S. J. A program for ligation at threonine sites: application to the controlled total synthesis of glycopeptides. *Tetrahedron* **66,** 2277–2283 (2010).
640. Okamoto, R. & Kajihara, Y. Uncovering a Latent Ligation Site for Glycopeptide Synthesis. *Angew. Chem. Int. Ed. (English)* **47,** 5402–5406 (2008).
641. Kaiser, R. & Metzka, L. Enhancement of Cyanogen Bromide Cleavage Yields for Methionyl-Serine and Methionyl-Threonine Peptide Bonds. *Anal. Biochem.* **266,** 1–8 (1999).
642. Merrifield, R. B. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **85,** 2149–2154 (1963).
643. Chattopadhaya, S., Srinivasan, R., Yeo, D. S. Y., Chen, G. Y. J. & Yao, S. Q. Site-specific covalent labeling of proteins inside live cells using small molecule probes. *Bioorg. Med. Chem.* **17,** 981–989 (2009).
644. Xiao, J., Burn, A. & Tolbert, T. J. Increasing Solubility of Proteins and Peptides by Site-Specific Modification with Betaine. *Bioconjugate Chem.* **19,** 1113–1118 (2008).
645. Nguyen, D. P., Elliott, T., Holt, M., Muir, T. W. & Chin, J. W. Genetically Encoded 1,2-Aminothiols Facilitate Rapid and Site-Specific Protein Labeling *via* a Bio-orthogonal Cyanobenzothiazole Condensation. *J. Am. Chem. Soc.* **133,** 11418–11421 (2011).
646. Schnolzer, M. & Kent, S. Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease. *Science* **256,** 221 (1992).
647. Dawson, P., Muir, T., Clark-Lewis, I. & Kent, S. Synthesis of proteins by native chemical ligation. *Science* **266,** 776–776 (1994).
648. Wieland, T., Bokelmann, E., Bauer, L., Lang, H. U. & Lau, H. Über Peptidsynthesen. 8. Mitteilung Bildung von S-haltigen Peptiden durch intramolekulare Wanderung von Aminoacylresten. *Justus Liebigs Ann. Chem.* **583,** 129–149 (1953).
649. Hackeng, T. M., Griffin, J. H. & Dawson, P. E. Protein synthesis by native chemical ligation: Expanded scope by using straightforward methodology. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 10068–10073 (1999).
650. Johnson, E. C. B. & Kent, S. B. H. Insights into the Mechanism and Catalysis of the Native Chemical Ligation Reaction. *J. Am. Chem. Soc.* **128,** 6640–6646 (2006).
651. Durek, T. & Alewood, P. F. Preformed Selenoesters Enable Rapid Native Chemical Ligation at Intractable Sites. *Angew. Chem. Int. Ed. (English)* **50,** 12042–12045 (2011).
652. Dang, B., Kubota, T., Mandal, K., Bezanilla, F. & Kent, S. B. H. Native Chemical Ligation at Asx-Cys, Glx-Cys: Chemical Synthesis and High-Resolution X-ray Structure of ShK Toxin by Racemic Protein Crystallography. *J. Am. Chem. Soc.* **135,** 11911–11919 (2013).
653. Lobban, P. E. & Kaiser, A. D. Enzymatic end-to-end joining of DNA molecules. *J. Mol. Biol.* **78,** 453–471 (1973).
654. Jackson, D. A., Symons, R. H. & Berg, P. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **69,** 2904–2909 (1972).
655. Sohma, Y. *et al.* Comparative Properties of Insulin-like Growth Factor 1 (IGF-1) and [Gly7D-Ala]IGF-1 Prepared by Total Chemical Synthesis. *Angew. Chem. Int. Ed. (English)* **47,** 1102–1106 (2008).
656. Raibaut, L., Ollivier, N. & Melnyk, O. Sequential native peptide ligation strategies for total chemical protein synthesis. *Chem. Soc. Rev.* **41,** 7001–7015 (2012).
657. Harpaz, Z., Siman, P., Kumar, K. S. A. & Brik, A. Protein Synthesis Assisted by Native Chemical Ligation at Leucine. *Chem. Eur. J. of Chem. Bio.* **11,** 1232–1235 (2010).
658. Bang, D., Pentelute, B. L. & Kent, S. B. H. Kinetically controlled ligation for the convergent chemical synthesis of proteins.

Angew. Chem. Int. Ed. (English) **45,** 3985–3988 (2006).
659. Macmillan, D. Evolving strategies for protein synthesis converge on native chemical ligation. *Angew. Chem. Int. Ed. (English)* **45,** 7668–7672 (2006).
660. Yuan, Y. *et al.* Toward Homogeneous Erythropoietin: Fine Tuning of the C-Terminal Acyl Donor in the Chemical Synthesis of the Cys 29−Gly 77Glycopeptide Domain. *J. Am. Chem. Soc.* **131,** 5432–5437 (2009).
661. Gordon, W. R., Bang, D., Hoff, W. D. & Kent, S. B. Total chemical synthesis of fully functional Photoactive Yellow Protein. *Bioorg. Med. Chem.* **21,** 3436–3442 (2013).
662. Tesser, G. I. & Balvert-Geers, I. C. The methylsulfonylethyloxycarbonyl group, a new and versatile amino protective function. *Int. J. Pept. Protein Res.* **7,** 295–305 (1975).
663. Gorse, A. D. & Gready, J. E. Molecular dynamics simulations of the docking of substituted N5- deazapterins to dihydrofolate reductase. *Protein Eng.* **10,** 23–30 (1997).
664. Camarero, J. A., Cotton, G. J., Adeva, A. & Muir, T. W. Chemical ligation of unprotected peptides directly from a solid support. *J. Pept. Res.* **51,** 303–316 (2009).
665. Veber, D., Milkowski, J., Varga, S., Denkewalter, R. & Hirschmann, R. Acetamidomethyl. A Novel Thiol Protecting Group for Cysteine. *J. Am. Chem. Soc.* **94,** 5456–5461 (1972).
666. Bang, D. & Kent, S. A one-pot total synthesis of crambin. *Angew. Chem. Int. Ed. (English)* **43,** 2534–2538 (2004).
667. Lahiri, S., Brehs, M., Olschewski, D. & Becker, C. F. W. Total Chemical Synthesis of an Integral Membrane Enzyme: Diacylglycerol Kinase from Escherichia coli. *Angew. Chem. Int. Ed. (English)* **50,** 3988–3992 (2011).
668. Ueda, S., Fujita, M., Tamamura, H., Fujii, N. & Otaka, A. Photolabile Protection for One-Pot Sequential Native Chemical Ligation. *Chem. Eur. J. of Chem. Bio.* **6,** 1983–1986 (2005).
669. Botti, P., Villain, M., Manganiello, S. & Gaertner, H. Native chemical ligation through in situ O to S acyl shift. *Org. Lett* **6,** 4861–4864 (2004).
670. Zheng, J.-S., Cui, H.-K., Fang, G.-M., Xi, W.-X. & Liu, L. Chemical Protein Synthesis by Kinetically Controlled Ligation of Peptide O-Esters. *Chem. Eur. J. of Chem. Bio.* **11,** 511–515 (2010).
671. Ingenito, R., Bianchi, E., Fattori, D. & Pessi, A. Solid Phase Synthesis of Peptide C-Terminal Thioesters by Fmoc/*t*-Bu Chemistry. *J. Am. Chem. Soc.* **121,** 11369–11374 (1999).
672. Aimoto, S. Polypeptide synthesis by the thioester method. *Biopolymers* **51,** 247–265 (1999).
673. Schnölzer, M., Alewood, P., Jones, A. & Alewood, D. In situ neutralisation in Boc-chemistry solid phase peptide synthesis. *Int. J. Pept. Protein Res.* **13,** 31–44 (2007).
674. Eggelkraut-Gottanka, von, R., Klose, A., Beck-Sickinger, A. G. & Beyermann, M. Peptide αthioester formation using standard Fmoc-chemistry. *Tetrahedron Lett.* **44,** 3551–3554 (2003).
675. Blanco-Canosa, J. B. & Dawson, P. E. An Efficient Fmoc-SPPS Approach for the Generation of Thioester Peptide Precursors for Use in Native Chemical Ligation. *Angew. Chem. Int. Ed. (English)* **47,** 6851–6855 (2008).
676. Thomas, F. Fmoc-based peptide thioester synthesis with self-purifying effect: heading to native chemical ligation in parallel formats. *J. Pept. Sci.* **19,** 141–147 (2013).
677. Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: a general method for protein engineering. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 6705–6710 (1998).
678. David, R., Richter, M. P. O. & Beck-Sickinger, A. G. Expressed protein ligation. *Eur. J. Biochem.* **271,** 663–677 (2004).
679. Flavell, R. R. & Muir, T. W. Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. *Acc. Chem. Res.* **42,** 107–116 (2009).
680. Warren, J. D., Miller, J. S., Keding, S. J. & Danishefsky, S. J. Toward fully synthetic glycoproteins by ultimately convergent routes: a solution to a long-standing problem. *J. Am. Chem. Soc.* **126,** 6576–6578 (2004).
681. Sato, K. *et al.* N-Sulfanylethylanilide Peptide as a Crypto-Thioester Peptide. *Chem. Eur. J. of Chem. Bio.* **12,** 1840–1844 (2011).
682. Ohta, Y. *et al.* Cysteine-Derived S-Protected Oxazolidinones: Potential Chemical Devices for the Preparation of Peptide Thioesters. *Org. Lett* **8,** 467–470 (2006).
683. Romanelli, A., Shekhtman, A., Cowburn, D. & Muir, T. W. Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 6397–6402 (2004).
684. Nagaike, F. *et al.* Efficient Microwave-Assisted Tandem N-to S-Acyl Transfer and Thioester Exchange for the Preparation of a Glycosylated Peptide Thioester. *Org. Lett* **8,** 4465–4468 (2006).
685. Hojo, H., Onuma, Y., Akimoto, Y., Nakahara, Y. & Nakahara, Y. N-Alkyl cysteine-assisted thioesterification of peptides. *Tetrahedron Lett.* **48,** 25–28 (2007).
686. Otaka, A., Sato, K., Ding, H. & Shigenaga, A. One-Pot/Sequential Native Chemical Ligation Using N-Sulfanylethylanilide Peptide. *Chem. Record.* **12,** 479–490 (2012).
687. Kang, J. & Macmillan, D. Peptide and protein thioester synthesis *via* N→S acyl transfer. *Org. Biomol. Chem.* **8,** 1993–2002 (2010).
688. Erlich, L. A., Kumar, K. S. A., Haj-Yahya, M., Dawson, P. E. & Brik, A. N-Methylcysteine-mediated total chemical synthesis of ubiquitin thioester. *Org. Biomol. Chem.* **8,** 2392–2396 (2010).
689. Sato, K. *et al.* Chemical Synthesis of Biologically Active Monoglycosylated GM2-Activator Protein Analogue Using N-Sulfanylethylanilide Peptide. *Angew. Chem. Int. Ed. (English)* **52,** 7855–7859 (2013).
690. Ollivier, N., Dheur, J., Mhidia, R., Blanpain, A. & Melnyk, O. Bis(2-sulfanylethyl)amino Native Peptide Ligation. *Org. Lett* **12,** 5238–5241 (2010).
691. Dheur, J., Ollivier, N., Vallin, A. & Melnyk, O. Synthesis of peptide alkylthioesters using the intramolecular N, S-acyl shift properties of bis (2-sulfanylethyl) amido peptides. *J. Org. Chem.* **76,** 3194–3202 (2011).
692. Hou, W., Zhang, X., Li, F. & Liu, C.-F. Peptidyl N, N-Bis(2-mercaptoethyl)-amides as Thioester Precursors for Native Chemical Ligation. *Org. Lett* **13,** 386–389 (2011).
693. Ollivier, N. *et al.* Tidbits for the synthesis of bis(2-sulfanylethyl)amido (SEA) polystyrene resin, SEA peptides and peptide thioesters. *J. Pept. Sci.* (2013). doi:10.1002/psc.2580
694. Ollivier, N. *et al.* A One-Pot Three-Segment Ligation Strategy for Protein Chemical Synthesis. *Angew. Chem. Int. Ed. (English)* **51,** 209–213 (2011).
695. Yan, L. & Dawson, P. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurisation. *J. Am. Chem. Soc.* **123,** 526–533 (2001).
696. Crich, D. & Banerjee, A. Native Chemical Ligation at Phenylalanine. *J. Am. Chem. Soc.* **129,** 10064–10065 (2007).
697. Botti, P. & Tchertchian, S. Side-Chain Extended Ligation. Patent: WO2006133962 (2006).
698. Canne, L. E., Bark, S. J. & Kent, S. B. Extending the applicability of native chemical ligation. *J. Am. Chem. Soc.* **118,** 5891–5896 (1996).
699. Haase, C., Rohde, H. & Seitz, O. Native Chemical Ligation at Valine. *Angew. Chem. Int. Ed. (English)* **47,** 6807–6810 (2008).
700. Chen, J., Wan, Q., Yuan, Y., Zhu, J. & Danishefsky, S. J. Native Chemical Ligation at Valine: A Contribution to Peptide and Glycopeptide Synthesis. *Angew. Chem. Int. Ed. (English)* **47,** 8521–8524 (2008).
701. Tan, Z., Shang, S. & Danishefsky, S. J. Insights into the Finer Issues of Native Chemical Ligation: An Approach to Cascade Ligations. *Angew. Chem. Int. Ed. (English)* **49,** 9500–9503 (2010).
702. Shang, S., Tan, Z., Dong, S. & Danishefsky, S. J. An advance in proline ligation. *J. Am. Chem. Soc.* **133,** 10784–10786 (2011).
703. Townsend, S. D. *et al.* Advances in Proline Ligation. *J. Am. Chem. Soc.* **134,** 3912–3916 (2012).

704. Ding, H., Shigenaga, A., Sato, K., Morishita, K. & Otaka, A. Dual Kinetically Controlled Native Chemical Ligation Using a Combination of Sulfanylproline and Sulfanylethylanilide Peptide. *Org. Lett* **13,** 5588–5591 (2011).
705. Thompson, R. E., Chan, B., Radom, L., Jolliffe, K. A. & Payne, R. J. Chemoselective Peptide Ligation-Desulfurisation at Aspartate. *Angew. Chem. Int. Ed. (English)* **52,** 9723–9727 (2013).
706. Siman, P., Karthikeyan, S. V. & Brik, A. Native chemical ligation at glutamine. *Org. Lett* **14,** 1520–1523 (2012).
707. Tam, J. P. & Yu, Q. Methionine ligation strategy in the biomimetic synthesis of parathyroid hormones. *Biopolymers* **46,** 319–327 (1998).
708. Dose, C. & Seitz, O. New isocysteine building blocks and chemoselective peptide ligation. *Org. Biomol. Chem.* **2,** 59–65 (2004).
709. Malins, L. R., Cergol, K. M. & Payne, R. J. Peptide Ligation–Desulfurisation Chemistry at Arginine. *Chem. Eur. J. of Chem. Bio.* **14,** 559–563 (2013).
710. Pasunooti, K. K. *et al.* Synthesis of 4-mercapto-L-lysine derivatives: potential building blocks for sequential native chemical ligation. *Bioorg. Med. Chem. Lett.* **19,** 6268–6271 (2009).
711. Yang, R., Pasunooti, K. K., Li, F., Liu, X.-W. & Liu, C.-F. Dual Native Chemical Ligation at Lysine. *J. Am. Chem. Soc.* **131,** 13592–13593 (2009).
712. Yang, Y.-Y., Ficht, S., Brik, A. & Wong, C.-H. Sugar-Assisted Ligation in Glycoprotein Synthesis. *J. Am. Chem. Soc.* **129,** 7690–7701 (2007).
713. Wan, Q. & Danishefsky, S. J. Free-Radical-Based, Specific Desulfurisation of Cysteine: A Powerful Advance in the Synthesis of Polypeptides and Glycopolypeptides. *Angew. Chem. Int. Ed. (English)* **119,** 9408–9412 (2007).
714. Haase, C. & Seitz, O. Extending the scope of native chemical peptide coupling. *Angew. Chem. Int. Ed. (English)* **47,** 1553–1556 (2008).
715. Siman, P. *et al.* Chemical Synthesis and Expression of the HIV-1 Rev Protein. *Chem. Eur. J. of Chem. Bio.* **12,** 1097–1104 (2011).
716. Moyal, T., Hemantha, H. P., Siman, P., Refua, M. & Brik, A. Highly efficient one-pot ligation and desulfurisation. *Chem. Sci.* **4,** 2496–2501 (2013).
717. Pentelute, B. L. & Kent, S. B. Selective desulfurisation of cysteine in the presence of Cys (Acm) in polypeptides obtained by native chemical ligation. *Org. Lett* **9,** 687–690 (2007).
718. Metanis, N., Keinan, E. & Dawson, P. E. Traceless Ligation of Cysteine Peptides Using Selective Deselenisation. *Angew. Chem. Int. Ed. (English)* **49,** 7049–7053 (2010).
719. Hondal, R. J., Nilsson, B. L. & Raines, R. T. Selenocysteine in native chemical ligation and expressed protein ligation. *J. Am. Chem. Soc.* **123,** 5140–5141 (2001).
720. Gieselman, M. D., Xie, L. & van der Donk, W. A. Synthesis of a Selenocysteine-Containing Peptide by Native Chemical Ligation. *Org. Lett* **3,** 1331–1334 (2001).
721. Quaderer, R., Sewing, A. & Hilvert, D. Selenocysteine-Mediated Native Chemical Ligation. **84,** 1197–1206 (2001).
722. McGrath, N. A. & Raines, R. T. Chemoselectivity in chemical biology: acyl transfer reactions with sulfur and selenium. *Acc. Chem. Res.* **44,** 752–761 (2011).
723. Gorlatov, S. N. & Stadtman, T. C. Human thioredoxin reductase from HeLa cells: selective alkylation of selenocysteine in the protein inhibits enzyme activity and reduction with NADPH influences affinity to heparin. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 8520–8525 (1998).
724. Metanis, N., Keinan, E. & Dawson, P. E. Synthetic seleno-glutaredoxin 3 analogues are highly reducing oxidoreductases with enhanced catalytic efficiency. *J. Am. Chem. Soc.* **128,** 16684–16691 (2006).
725. Dawson, P. E. Native Chemical Ligation Combined with Desulfurisation and Deselenisation: A General Strategy for Chemical Protein Synthesis. *Isr. J. Chem.* **51,** 862–867 (2011).
726. Wang, Z., Rejtar, T., Zhou, Z. S. & Karger, B. L. Desulfurisation of cysteine-containing peptides resulting from sample preparation for protein characterisation by mass spectrometry. *Rapid Commun. Mass Spectrom.* **24,** 267–275 (2010).
727. Malins, L. R. & Payne, R. J. Synthesis and Utility of β-Selenol-Phenylalanine for Native Chemical Ligation–Deselenisation Chemistry. *Org. Lett* **14,** 3142–3145 (2012).
728. Adams, A. L. & Macmillan, D. Investigation of peptide thioester formation *via* N→ Se acyl transfer. *J. Pept. Sci.* **19,** 65–73 (2013).
729. Ghassemian, A., Vila-Farrés, X., Alewood, P. F. & Durek, T. Solid phase synthesis of peptide-selenoesters. *Bioorg. Med. Chem.* **21,** 3473–3478 (2013).
730. Coltart, D. M. Peptide segment coupling by prior ligation and proximity-induced intramolecular acyl transfer. *Tetrahedron* **56,** 3449–3491 (2000).
731. Fang, G.-M., Cui, H.-K., Zheng, J.-S. & Liu, L. Chemoselective Ligation of Peptide Phenyl Esters with N-Terminal Cysteines. *Chem. Eur. J. of Chem. Bio.* **11,** 1061–1065 (2010).
732. Monbaliu, J.-C. M. & Katritzky, A. R. Recent trends in Cys- and Ser/Thr-based synthetic strategies for the elaboration of peptide constructs. *Chem. Commun.* **48,** 11601–11622 (2012).
733. Kemp, D. S., Leung, S.-L. & Kerkman, D. J. Models that demonstrate peptide bond formation by prior thiol capture I. Capture by disulfide formation. *Tetrahedron Lett.* **22,** 181–184 (1981).
734. Kemp, D. S. & Kerkman, D. J. Models that demonstrate peptide bond formation by prior thiol capture--II capture by organomercury derivatives. *Tetrahedron Lett.* **22,** 185–186 (1981).
735. Kemp, D. S., Kerkman, D. J., Leung, S.-L. & Hanson, G. Intramolecular O,N-acyl transfer *via* cyclic intermediates of nine and twelve members. Models for extensions of the amine capture strategy for peptide synthesis. *J. Org. Chem.* **46,** 490–498 (1981).
736. Kemp, D. S. & Galakatos, N. G. Peptide synthesis by prior thiol capture. 1. A convenient synthesis of 4-hydroxy-6-mercaptodibenzofuran and novel solid-phase synthesis of peptide-derived 4-(acyloxy)-6-mercaptodibenzofurans. *J. Org. Chem.* **51,** 1821–1829 (1986).
737. Kemp, D. S., Galakatos, N. G., Bowen, B. & Tan, K. Peptide synthesis by prior thiol capture. 2. Design of templates for intramolecular O,N-acyl transfer. 4,6-Disubstituted dibenzofurans as optimal spacing elements. *J. Org. Chem.* **51,** 1829–1838 (1986).
738. Kemp, D. S. *et al.* Peptide synthesis by prior thiol capture. 4. Amide bond formation. The effect of a side-chain substituent on the rates of intramolecular O,N-acyl transfer. *J. Org. Chem.* **51,** 3320–3324 (1986).
739. Kemp, D. S. & Carey, R. I. Boc-L-Dmt-OH as a fully N,S-blocked cysteine derivative for peptide synthesis by prior thiol capture. Facile conversion of N-terminal Boc-L-Dmt-peptides to H-Cys(Scm)-peptides. *J. Org. Chem.* **54,** 3640–3646 (1989).
740. Kemp, D. S. & Buckler, D. R. New templates for prior thiol capture from xanthene, dibenzo[c,h]xanthen-7-one and 2-methylenedihydrobenzofuran. *Tetrahedron Lett.* **32,** 3009–3012 (1991).
741. Kemp, D. S. & Buckler, D. R. Highly enantioselective amide ligation by prior thiol capture. *Tetrahedron Lett.* **32,** 3013–3016 (1991).
742. Liu, C.-F., Rao, C. & Tam, J. P. Acyl disulfide-mediated intramolecular acylation for orthogonal coupling between unprotected peptide segments. Mechanism and application. *Tetrahedron Lett.* **37,** 933–936 (1996).
743. Tan, X.-H., Zhang, X., Yang, R. & Liu, C.-F. A Simple Method for Preparing Peptide C-Terminal Thioacids and Their Application in Sequential Chemoenzymatic Ligation. *Chem. Eur. J. of Chem. Bio.* **9,** 1052–1056 (2008).
744. Zhang, X., Li, F. & Liu, C.-F. Synthesis of histone H3 proteins by a thioacid capture ligation strategy. *Chem. Commun.* **47,** 1746–1748 (2011).
745. Dirksen, A. & Dawson, P. E. Expanding the scope of chemoselective peptide ligations in chemical biology. *Curr. Opin. Chem. Biol.* **12,** 760–766 (2008).

746. Zou, G. *et al.* Chemoenzymatic synthesis and Fcγ receptor binding of homogeneous glycoforms of antibody Fc domain. Presence of a bisecting sugar moiety enhances the affinity of Fc to FcγIIIa receptor. *J. Am. Chem. Soc.* **133,** 18975–18991 (2011).
747. Ratner, S. & Clarke, H. T. The Action of Formaldehyde upon Cysteine. *J. Am. Chem. Soc.* **59,** 200–206 (1937).
748. Rao, C. & Tam, J. P. Synthesis of Peptide Dendrimer. *J. Am. Chem. Soc.* **116,** 6975–6976 (1994).
749. Botti, P., Pallin, T. D. & Tam, J. P. Cyclic Peptides from Linear Unprotected Peptide Precursors through Thiazolidine Formation †. *J. Am. Chem. Soc.* **118,** 10018–10024 (1996).
750. Shao, J. & Tam, J. P. Unprotected peptides as building blocks for the synthesis of peptide dendrimers with oxime, hydrazone, and thiazolidine linkages. *J. Am. Chem. Soc.* **117,** 3893–3899 (1995).
751. Wöhr, T. *et al.* Pseudo-Prolines as a Solubilizing, Structure-Disrupting Protection Technique in Peptide Synthesis. *J. Am. Chem. Soc.* **118,** 9218–9227 (1996).
752. Villain, M., Vizzavona, J. & Rose, K. Covalent capture: a new tool for the purification of synthetic and recombinant polypeptides. *Chem. Biol.* **8,** 673–679 (2001).
753. Muir, T. W. Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* **72,** 249–289 (2003).
754. Casi, G. Convergent protein synthesis. *Curr. Opin. Struct. Biol.* **13,** 589–594 (2003).
755. Casi, G., Huguenin-Dezot, N., Zuberbühler, K., Scheuermann, J. & Neri, D. Site-Specific Traceless Coupling of Potent Cytotoxic Drugs to Recombinant Antibodies for Pharmacodelivery. *J. Am. Chem. Soc.* **134,** 5887–5892 (2012).
756. Bernardes, G. J. L., Steiner, M., Hartmann, I., Neri, D. & Casi, G. Site-specific chemical modification of antibody fragments using traceless cleavable linkers. *Nat. Protocols* **8,** 2079–2089 (2013).
757. White, E. H., McCapra, F. & Field, G. F. The structure and synthesis of firefly luciferin. *J. Am. Chem. Soc.* **85,** 337–343 (1963).
758. Bitler, B. & McElroy, W. D. The preparation and properties of crystalline firefly luciferin. *Arch. Biochem. Biophys.* **72,** 358–368 (1957).
759. Li, J., Chen, L., Du, L. & Li, M. Cage the firefly luciferin! – A strategy for developing bioluminescent probes. *Chem. Soc. Rev.* **42,** 662–676 (2012).
760. Ren, H. *et al.* A Biocompatible Condensation Reaction for the Labeling of Terminal Cysteine Residues on Proteins. *Angew. Chem. Int. Ed. (English)* **48,** 9658–9662 (2009).
761. Ramil, C. P. & Lin, Q. Bioorthogonal chemistry: strategies and recent developments. *Chem. Commun.* **49,** 11007–11022 (2013).
762. Liang, G., Ren, H. & Rao, J. A biocompatible condensation reaction for controlled assembly of nanostructures in living cells. *Nature Chem.* **2,** 54–60 (2010).
763. Liang, G. *et al.* Controlled Self-Assembling of Gadolinium Nanoparticles as Smart Molecular Magnetic Resonance Imaging Contrast Agents. *Angew. Chem. Int. Ed. (English)* **50,** 6283–6286 (2011).
764. Yuan, Y. *et al.* Detection of Glutathione in Vitroand in Cells by the Controlled Self-Assembly of Nanorings. *Anal. Chem.* **85,** 1280–1284 (2013).
765. Jeon, J. *et al.* Efficient Method for Site-Specific 18F-Labeling of Biomolecules Using the Rapid Condensation Reaction between 2-Cyanobenzothiazole and Cysteine. *Bioconjugate Chem.* **23,** 1902–1908 (2012).
766. Yuan, Y. & Liang, G. A biocompatible, highly efficient click reaction and its applications. *Org. Biomol. Chem.* (2013). doi:10.1039/C3OB41241E
767. Scoffone, E., Fontana, A. & Rocchi, R. Selective modification of the tryptophan residue in peptides and proteins using sulfenyl halides. *Biochem. Biophys. Res. Commun.* **25,** 170–174 (1966).
768. Scoffone, E., Fontana, A. & Rocchi, R. Sulfenyl halides as modifying reagents for polypeptides and proteins. I. Modification of tryptophan residues. *Biochemistry* **7,** 971–979 (1968).
769. Malins, L. R., Cergol, K. M. & Payne, R. J. Chemoselective sulfenylation and peptide ligation at tryptophan. *Chem. Sci.* (2014). doi:10.1039/c3sc51497h
770. Offer, J. Native chemical ligation with Nα acyl transfer auxiliaries. *Biopolymers* **94,** 530–541 (2010).
771. Li, X., Zhang, L., Hall, S. E. & Tam, J. P. A new ligation method for N-terminal tryptophan-containing peptides using the Pictet–Spengler reaction. *Tetrahedron Lett.* **41,** 4069–4073 (2000).
772. Pictet, A. & Spengler, T. Über die Bildung von Isochinolinderivaten durch Einwirkung von Methylal auf Phenyl-äthylamin, Phenyl-alanin und Tyrosin. *Eur. J. Inorg. Chem.* **44,** 2030–2036 (1911).
773. Popov, V., Panda, S. S. & Katritzky, A. R. Ligations of N-acyl tryptophan units to give native peptides *via* 7-, 10-, 11- and 12-membered cyclic transition states. *Org. Biomol. Chem.* **11,** 1594 (2013).
774. Zhang, L. & Tam, J. Orthogonal coupling of unprotected peptide segments through histidyl amino terminus. *Tetrahedron Lett.* **1,** 3–6 (1997).
775. Metanis, N., Beld, J. & Hilvert, D. in (2011).
776. Craik, D. J. & Conibear, A. C. The Chemistry of Cyclotides. *J. Org. Chem.* **76,** 4805–4817 (2011).
777. Malins, L. R., Mitchell, N. J. & Payne, R. J. Peptide ligation chemistry at selenol amino acids. *J. Pept. Sci.* (2013). doi:10.1002/psc.2581
778. Wessjohann, L. A., Schneider, A., Abbas, M. & Brandt, W. Selenium in chemistry and biochemistry in comparison to sulfur. *Biol. Chem.* **388,** 997–1006 (2007).
779. Casi, G., Roelfes, G. & Hilvert, D. Selenoglutaredoxin as a Glutathione Peroxidase Mimic. *Chem. Eur. J. of Chem. Bio.* **9,** 1623–1631 (2008).
780. Berry, S. M., Gieselman, M. D., Nilges, M. J., van der Donk, W. A. & Lu, Y. An Engineered Azurin Variant Containing a Selenocysteine Copper Ligand. *J. Am. Chem. Soc.* **124,** 2084–2085 (2002).
781. Ralle, M. *et al.* The Selenocysteine-Substituted Blue Copper Center: Spectroscopic Investigations of Cys112SeCys Pseudomonas aeruginosaAzurin. *J. Am. Chem. Soc.* **126,** 7244–7256 (2004).
782. Quaderer, R. & Hilvert, D. Selenocysteine-mediated backbone cyclisation of unprotected peptides followed by alkylation, oxidative elimination or reduction of the selenol. *Chem. Commun.* 2620–2621 (2002).
783. Forchhammer, K., Heider, J. & Baron, C. Selenoprotein synthesis: an expansion of the genetic code. *Trends Biochem. Sci.* **16,** 463–467 (1991).
784. Muttenthaler, M. & Alewood, P. F. Selenopeptide chemistry. *J. Pept. Sci.* **14,** 1223–1239 (2008).
785. Yi, L. *et al.* A highly efficient strategy for modification of proteins at the C terminus. *Angew. Chem. Int. Ed. (English)* **49,** 9417–9421 (2010).
786. Schwarz, A., Wandrey, C., Bayer, E. A. & Wilchek, M. [15] Enzymatic C-terminal biotinylation of proteins. *Methods Enzymol.* **184,** 160–162 (1990).
787. Keppler, A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **21,** 86–89 (2002).
788. Chen, I., Howarth, M., Lin, W. & Ting, A. Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* **2,** 99–104 (2005).
789. Flavell, R. R. *et al.* Site-Specific 18F-Labeling of the Protein Hormone Leptin Using a General Two-Step Ligation Procedure. *J. Am. Chem. Soc.* **130,** 9106–9112 (2008).
790. Antos, J. M. *et al.* Site-Specific N- and C-Terminal Labeling of a Single Polypeptide Using Sortases of Different Specificity. *J. Am. Chem. Soc.* **131,** 10800–10801 (2009).
791. Tsukiji, S. & Nagamune, T. Sortase-Mediated Ligation: A Gift from Gram-Positive Bacteria to Protein Engineering. *Chem. Eur. J. of Chem. Bio.* **10,** 787–798 (2009).
792. Soh, N. Selective Chemical Labeling of Proteins with Small Fluorescent Molecules Based on Metal-Chelation Methodology. **8,** 1–21 (2008).

793. Ojida, A. *et al.* Oligo-Asp Tag/Zn(II) Complex Probe as a New Pair for Labeling and Fluorescence Imaging of Proteins. *J. Am. Chem. Soc.* **128,** 10452–10459 (2006).
794. Hauser, C. T. & Tsien, R. Y. A hexahistidine-Zn2+-dye label reveals STIM1 surface exposure. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 3693–3697 (2007).
795. Adams, S. R. *et al.* New Biarsenical Ligands and Tetracysteine Motifs for Protein Labeling in Vitro and in Vivo: Synthesis and Biological Applications. *J. Am. Chem. Soc.* **124,** 6063–6076 (2002).
796. Yamaguchi, T. *et al.* Turn-ON fluorescent affinity labeling using a small bifunctional O-nitrobenzoxadiazole unit. *Chem. Sci.* (2014). doi:10.1039/C3SC52704B
797. Burkhard, P., Stetefeld, J. & Strelkov, S. V. Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* **11,** 82–88 (2001).
798. Grigoryan, G. & Keating, A. E. Structural specificity in coiled-coil interactions. *Curr. Opin. Struct. Biol.* **18,** 477–483 (2008).
799. Parry, D. A. D., Fraser, R. D. B. & Squire, J. M. Fifty years of coiled-coils and α-helical bundles: A close relationship between sequence and structure. *J. Struct. Biol.* **163,** 258–269 (2008).
800. Popp, B. V. & Ball, Z. T. Proximity-driven metallopeptide catalysis: Remarkable side-chain scope enables modification of the Fos bZip domain. *Chem. Sci.* **2,** 690 (2011).
801. Zaykov, A. N., MacKenzie, K. R. & Ball, Z. T. Controlling Peptide Structure with Coordination Chemistry: Robust and Reversible Peptide-Dirhodium Ligation. *Chem. Eur. J.* **15,** 8961–8965 (2009).
802. Keil, B. *Specificity of Proteolysis*. (Springer Verlag, 1992).
803. Zhi-Heng Huang *et al.* A Picomole-Scale Method for Charge Derivatisation of Peptides for Sequence Analysis by Mass Spectrometry. *Anal. Chem.* **69,** 137–144 (1997).
804. Huang, Z.-H., Shen, T., Wu, J., Gage, D. A. & Watson, J. T. Protein Sequencing by Matrix-Assisted Laser Desorption Ionisation–Postsource Decay–Mass Spectrometry Analysis of theN-Tris(2,4,6-trimethoxyphenyl)phosphine-Acetylated Tryptic Digests. *Anal. Biochem.* **268,** 305–317 (1999).
805. Bonnett, R. in *Carbon-Nitrogen Double Bonds (1970)* (Patai, S.) 597–662 (John Wiley & Sons, Ltd., 2010). doi:10.1002/9780470771204.ch13
806. Hameršak, Z. *et al.* Preparation and Mechanism of Solvolysis of N-Hydroxy-α-oxobenzeneethanimidoyl Chloride, a 2-(Hydroxyimino)-1-phenylethan-1-one Derivative: Molecular Structure of α-Oxo-oximes (=α-(Hydroxyimino) Ketones). *HCA* **82,** 1289–1301 (1999).
807. Kornblum, N. & Weaver, W. M. The Reaction of Sodium Nitrite with Ethyl Bromoacetate and with Benzyl Bromide 1,2. *J. Am. Chem. Soc.* **80,** 4333–4337 (1958).
808. Wolfbeis, O. S. Synthesen von Fluoreszenzfarbstoffen, 11 über die Umlagerung von Alkoxymethylen- und Aminomethylenhomophthalsäureanhydriden zu Isocumarinen bzw. Isochinolinonen. *Eur. J. Org. Chem.* **1981,** 819–827 (1981).
809. Matoba, K. & Yamazaki, T. Reduction of O-alkylated imides. *Chem Pharm Bull* **22,** 2999–3001 (1974).
810. Lundblad, R. L. *Application of Solution Protein Chemistry to Biotechnology*. (CRC Press, 2009).
811. Valente, C. *et al.* The 1,4-naphthoquinone scaffold in the design of cysteine protease inhibitors. *Bioorg. Med. Chem.* **15,** 5340–5350 (2007).
812. Cheng, Y. *et al.* Synthesis, cytotoxic activities and structure–activity relationships of topoisomerase I inhibitors: Indolizinoquinoline-5,12-dione derivatives. *Bioorg. Med. Chem.* **16,** 4617–4625 (2008).
813. Korobitsyna, I. K., Khalikova, A. V., Rodina, L. L. & Shusherina, N. P. 4-Phenyl-1,2,4-triazoline-3,5-dione in organic synthesis (review). *Chem. Heterocycl. Compd.(N.Y.)* **19,** 117–136 (1983).
814. Wisastra, R. *et al.* Isothiazolones; thiol-reactive inhibitors of cysteine protease cathepsin B and histone acetyltransferase PCAF. *Org. Biomol. Chem.* **9,** 1817–1822 (2011).
815. Kozhushkov, S. I. & Brandl, M. Surprisingly facile addition of thiols to the double bonds of bicyclopropylidene and other methylenecyclopropanes. *Eur. J. Org. Chem.* 1535–1542 (1998).
816. De Meijere, A., Kozhushkov, S. I. & Spaeth, T. A new general approach to bicyclopropylidenes. *J. Org. Chem.* **58,** 502–505 (1993).
817. Bloxham, J. & Dell, C. P. A convenient route to polyfunctionalised indeno[1,2 - b]pyran derivatives. *Tetrahedron Lett.* **32,** 4051–4054 (1991).
818. Heard, N. E. & Turner, J. Synthesis of a Novel N-Hydroxypyrrole *via* Lithium Perchlorate Accelerated Diels-Alder Methodology. *J. Org. Chem.* **60,** 4302–4304 (1995).
819. Sobenina, L. *et al.* The Addition Reactions of Pyrrolecarbodithioates to Activated Alkenes and Alkynes. *Synthesis* 293–299 (2001).
820. Heukeshoven, J. Side reaction of methionine with 4-vinylpyridine during acid hydrolysis of modified proteins. *Anal. Biochem.* **109,** 421–425 (1980).
821. Boja, E. S. & Fales, H. M. Overalkylation of a Protein Digest with Iodoacetamide. *Anal. Chem.* **73,** 3576–3582 (2001).
822. Roberts, M. J., Bentley, M. D. & Harris, J. M. Chemistry for peptide and protein PEGylation. *Adv. Drug Deliv. Rev.* **54,** 459–476 (2002).
823. Kim, Y. *et al.* Efficient Site-Specific Labeling of Proteins *via* Cysteines. *Bioconjugate Chem.* **19,** 786–791 (2008).
824. McAllister, G. D., Wilfred, C. D. & Taylor, R. J. K. Tandem oxidation processes: The direct conversion of activated alcohols into nitriles. *Synlett* 1291–1292 (2002).
825. Fen-Tair, L., Fen-Tair, L., Ren-Tzong, W. & Ren-Tzong, W. A novel synthesis of cyanoalkynes *via* iodide-catalyzed cyanation of terminal acetylenes with cuprous cyanide. *The International Journal for the Rapid Publication of Preliminary Communications in Organic Chemistry* **34,** 5911–5914 (1993).
826. Murray, R. E. & Zweifel, G. Preparation of phenyl cyanate and its utilisation for the synthesis of α, β-unsaturated nitriles. *Synthesis* **1980,** 150–151 (1980).
827. Kim, J.-G., Lee, E. H. & Jang, D. O. One-pot synthesis of conjugated alkynenitriles from aldehydes. *The International Journal for the Rapid Publication of Preliminary Communications in Organic Chemistry* **48,** 2299–2301 (2007).
828. Swindell, C. S. in *e-EROS Encyclopedia of Reagents for Organic Synthesis* 1–2 (John Wiley & Sons, Ltd, 2001). at <http://onlinelibrary.wiley.com/doi/10.1002/047084289X.rp276/full>
829. Alley, S. C. *et al.* Contribution of Linker Stability to the Activities of Anticancer Immunoconjugates. *Bioconjugate Chem.* **19,** 759–765 (2008).
830. Shen, B.-Q. *et al.* Conjugation site modulates the in vivo stability and therapeutic activity of antibody-drug conjugates. *Nat. Biotechnol.* **30,** 184–189 (2012).
831. King, M., Baati, R. & Wagner, A. New tetramethylthiepinium (TMTI) for copper-free click chemistry. *Chem. Commun.* **48,** 9308 (2012).
832. Leriche, G., Budin, G., Brino, L. & Wagner, A. Optimisation of the Azobenzene Scaffold for Reductive Cleavage by Dithionite; Development of an Azobenzene Cleavable Linker for Proteomic Applications. *Eur. J. Org. Chem.* **2010,** 4360–4364 (2010).
833. Lee, S. R. *et al.* Mammalian thioredoxin reductase: oxidation of the C-terminal cysteine/selenocysteine active site forms a thioselenide, and replacement of selenium with sulfur markedly reduces catalytic activity. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 2521–2526 (2000).
834. Mugesh, G. & Singh, H. B. Synthetic organoselenium compounds as antioxidants: glutathione peroxidase activity. *Chem. Soc. Rev.* **29,** 347–357 (2000).
835. Urig, S. & Becker, K. On the potential of thioredoxin reductase inhibitors for cancer therapy. *Semin. Cancer Biol.* **16,** 452–465 (2006).
836. Cai, W. *et al.* Free Radical Biology & Medicine. *Free Radic. Biol. Med.* **52,** 257–265 (2012).

837. Kofoed, J. & Reymond, J.-L. A general method for designing combinatorial peptide libraries decodable by amino acid analysis. *J. Comb. Chem.* **9,** 1046–1052 (2007).
838. Kofoed, J. & Reymond, J.-L. Identification of protease substrates by combinatorial profiling on TentaGel beads. *Chem. Commun.* 4453–4455 (2007). doi:10.1039/b713595e
839. Huang, X., Rickman, B. H. & Borhan, B. Zinc porphyrin tweezer in host-guest complexation: determination of absolute configurations of diamines, amino acids, and amino alcohols by circular dichroism. *J. Am. Chem. Soc.* 6185–6186 (1998). doi:10.1021/ja973539e
840. Castellano, S. *et al.* Synthesis and Biochemical Evaluation of Δ2-Isoxazoline Derivatives as DNA Methyltransferase 1 Inhibitors. *J. Med. Chem.* **54,** 7663–7677 (2011).
841. Matos, M. R. P. N., Gois, P. M. P., Mata, M. L. E. N., Cabrita, E. J. & Afonso, C. A. M. Studies on the Preparation of 4-Ethoxyalkyliden and 4-Aminoalkyliden-5(4H)-oxazolones. *Synth. Comm.* **33,** 1285–1299 (2011).
842. de Meijere, A., Kozhushkov, S. & Spath, T. Bicyclopropylidene. *Org. Synth.* 142–149 (2001).
843. Bio, M., Nkepang, G. & You, Y. Click and photo-unclick chemistry of aminoacrylate for visible light-triggered drug release. *Chem. Commun.* **48,** 6517–6519 (2012).
844. Dombrowski, G. W. *et al.* Efficient unimolecular deprotonation of aniline radical cations. *J. Org. Chem.* **70,** 3791–3800 (2005).
845. Rulíšek, L. *et al.* An Experimental and Theoretical Study of Stereoselectivity of Furan−Maleic Anhydride and Furan−Maleimide Diels−Alder Reactions. *The Journal of Organic Chemistry* **70,** 6295–6302 (2005).
846. Wang, Y. & Lin, Q. Synthesis and Evaluation of Photoreactive Tetrazole Amino Acids. *Org. Lett* **11,** 3570–3573 (2009).
847. Budin, G. *et al.* Nondenaturing Chemical Proteomics for Protein Complex Isolation and Identification. *Chem. Eur. J. of Chem. Bio.* **11,** 2359–2361 (2010).
848. Kvach, M. V. *et al.* Practical Synthesis of Isomerically Pure 5- and 6-Carboxytetramethylrhodamines, Useful Dyes for DNA Probes. *Bioconjugate Chem.* **20,** 1673–1682 (2009).

F. ANNEXES

ANNEX 1: *IN SILICO* DIGESTION OF LYSOZYME BY TRYPSIN.

The full list of the peptides (400 Da < M_w < 6000 Da) generated by *in silico* digestion of hen white egg lysozyme (pdb: 1LYZ) using "Protein Digestion Simulator" software (http://omics.pnl.gov/software/ProteinDigestionSimulator.php).

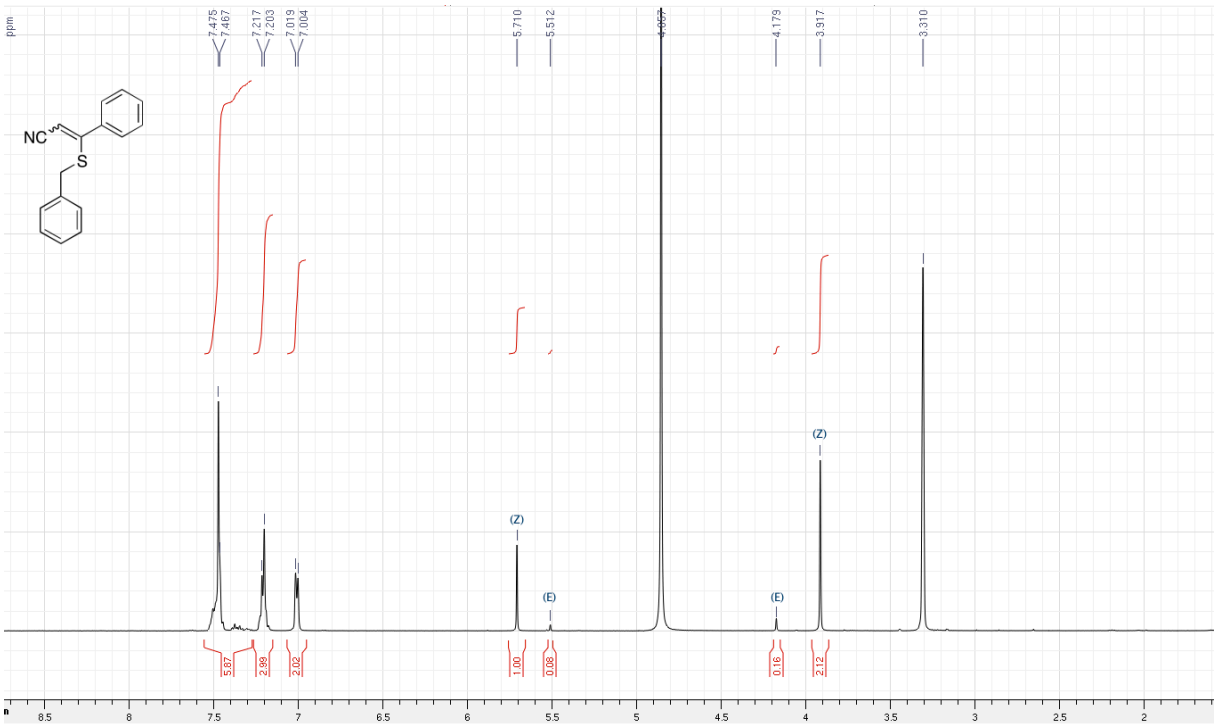
FASTA SEQUENCE: >1LYZ:A|PDBID|CHAIN|SEQUENCE
KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGS
TDYGILQINSRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVS
DGNGMNAWVAWRNRCKGTDVQAWIRGCRL

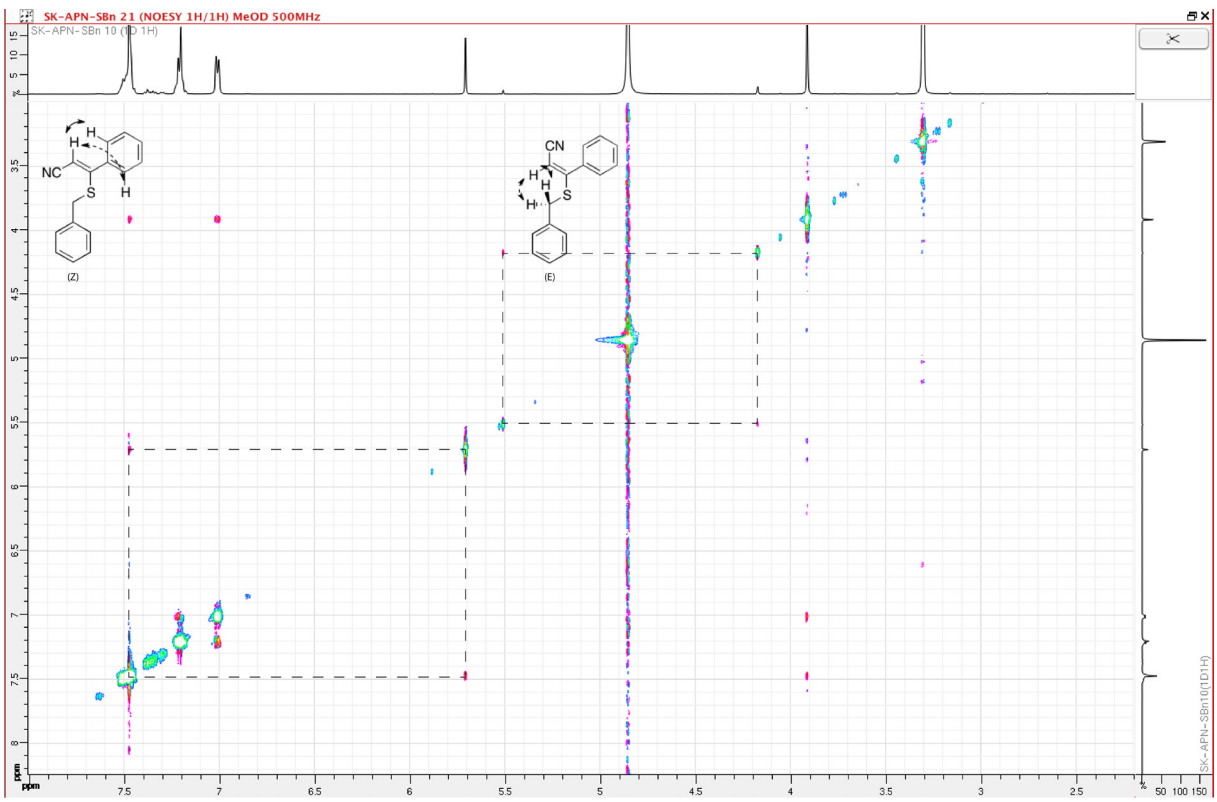
TRYPTIC DIGEST CONSTITUTION (CELLS CONTAINING CYS-CONTAINING PEPTIDES ARE HIGHLIGHTED IN GREEN):

#	Sequence	Monoisotopic Mass
1	KVFGR	605,36
2	KVFGRCELAAAMK	1422,75
3	KVFGRCELAAAMKR	1578,85
4	VFGR	477,27
5	VFGRCELAAAMK	1294,65
6	VFGRCELAAAMKR	1450,75
7	VFGRCELAAAMKRHGLDNYR	2306,15
8	CELAAAMK	835,39
9	CELAAAMKR	991,49
10	CELAAAMKRHGLDNYR	1846,89
11	CELAAAMKRHGLDNYRGYSLGNWVCAAK	3096,48
12	RHGLDNYR	1029,51
13	RHGLDNYRGYSLGNWVCAAK	2279,10
14	RHGLDNYRGYSLGNWVCAAKFESNFNTQATNR	3688,73
15	HGLDNYR	873,41
16	HGLDNYRGYSLGNWVCAAK	2123,00
17	HGLDNYRGYSLGNWVCAAKFESNFNTQATNR	3532,63
18	HGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSR	5267,45
19	GYSLGNWVCAAK	1267,60
20	GYSLGNWVCAAKFESNFNTQATNR	2677,23
21	GYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSR	4412,05
22	GYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSRWWCNDGR	5329,41
23	FESNFNTQATNR	1427,64
24	FESNFNTQATNRNTDGSTDYGILQINSR	3162,46
25	FESNFNTQATNRNTDGSTDYGILQINSRWWCNDGR	4079,82
26	FESNFNTQATNRNTDGSTDYGILQINSRWWCNDGRTPGSR	4578,08
27	NTDGSTDYGILQINSR	1752,83

28	NTDGSTDYGILQINSRWWCNDGR	2670,19
29	NTDGSTDYGILQINSRWWCNDGRTPGSR	3168,44
30	NTDGSTDYGILQINSRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAK	5486,55
31	WWCNDGR	935,37
32	WWCNDGRTPGSR	1433,63
33	WWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAK	3751,73
34	WWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKK	3879,83
35	TPGSR	516,27
36	TPGSRNLCNIPCSALLSSDITASVNCAK	2834,37
37	TPGSRNLCNIPCSALLSSDITASVNCAKK	2962,47
38	TPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWR	4619,25
39	NLCNIPCSALLSSDITASVNCAK	2336,12
40	NLCNIPCSALLSSDITASVNCAKK	2464,21
41	NLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWR	4121,00
42	NLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNR	4391,14
43	KIVSDGNGMNAWVAWR	1802,89
44	KIVSDGNGMNAWVAWRNR	2073,03
45	KIVSDGNGMNAWVAWRNRCK	2304,14
46	IVSDGNGMNAWVAWR	1674,79
47	IVSDGNGMNAWVAWRNR	1944,94
48	IVSDGNGMNAWVAWRNRCK	2176,04
49	IVSDGNGMNAWVAWRNRCKGTDVQAWIR	3202,57
50	NRCK	519,26
51	NRCKGTDVQAWIR	1545,78
52	NRCKGTDVQAWIRGCR	1861,92
53	CKGTDVQAWIR	1275,64
54	CKGTDVQAWIRGCR	1591,77
55	CKGTDVQAWIRGCRL	1704,86
56	GTDVQAWIR	1044,54
57	GTDVQAWIRGCR	1360,67
58	GTDVQAWIRGCRL	1473,75
59	GCRL	447,23

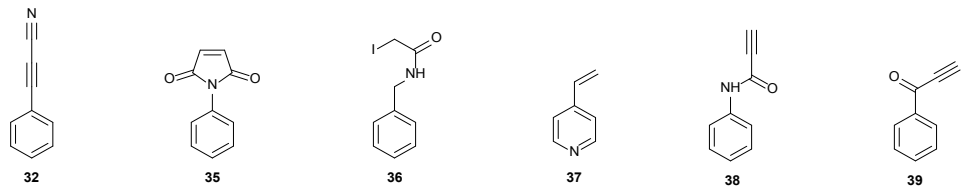
ANNEX 2: EXAMPLE OF APN-THIOL ADDITION PRODUCT.





ANNEX 3: STUDY OF HYDROLYTIC STABILITY OF 32, 35-39.

STRUCTURES OF THE STUDIED ELECTROPHILES:



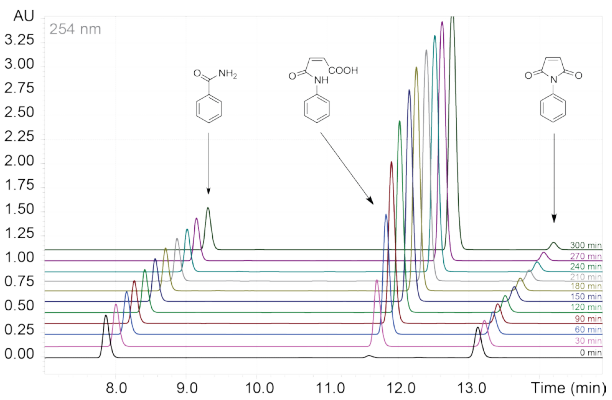
32 35 36 37 38 39

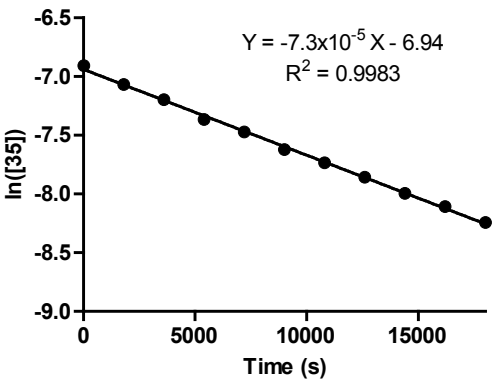
GENERAL CONSIDERATIONS:

Only probes 33 and 37 were found to degrade in PBS (1x, pH 7.6) at 25 ˚C. Conversion was monitored by disappearance of the starting material. Benzamide was used as an internal standard. Pseudo first order rate constant for the reaction was determined by plotting the ln([**C**]) versus time and analysing by linear regression. The constant corresponds to the absolute value of determined slope.

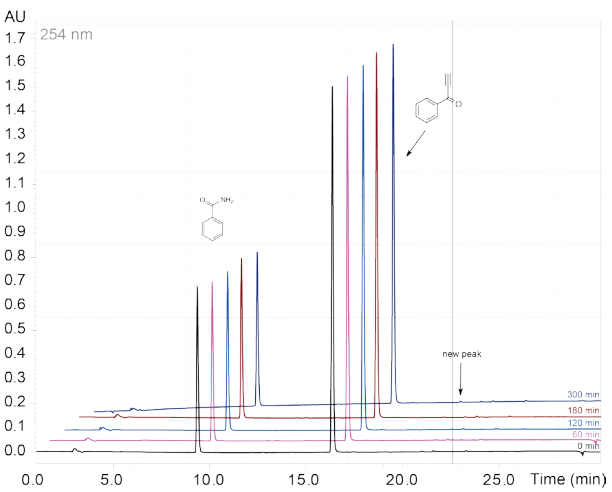
HPLC TRACES AND KINETICS PLOTS FOR HYDROLYTICALLY UNSTABLE SUBSTRATES:

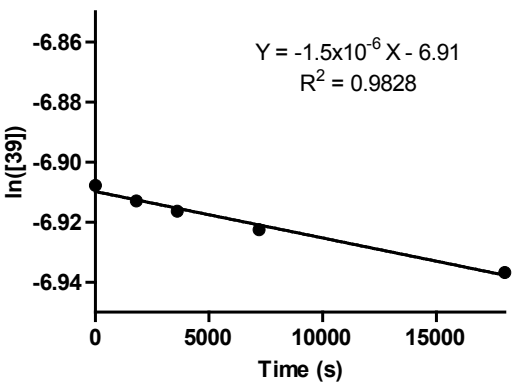
35



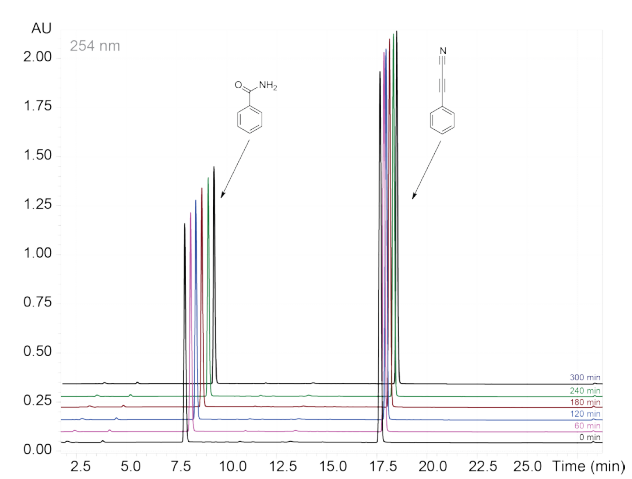


39

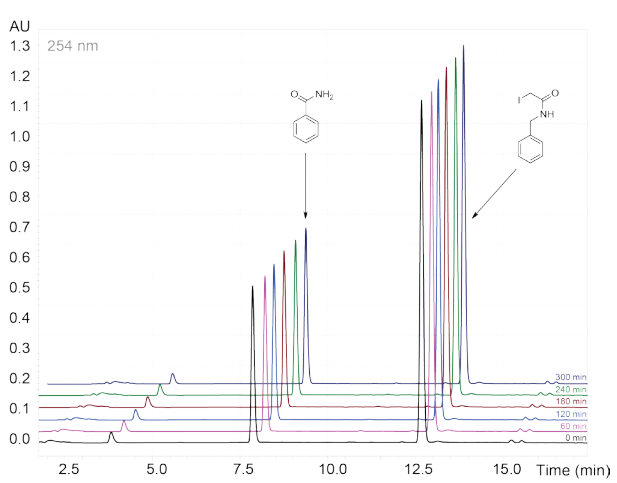




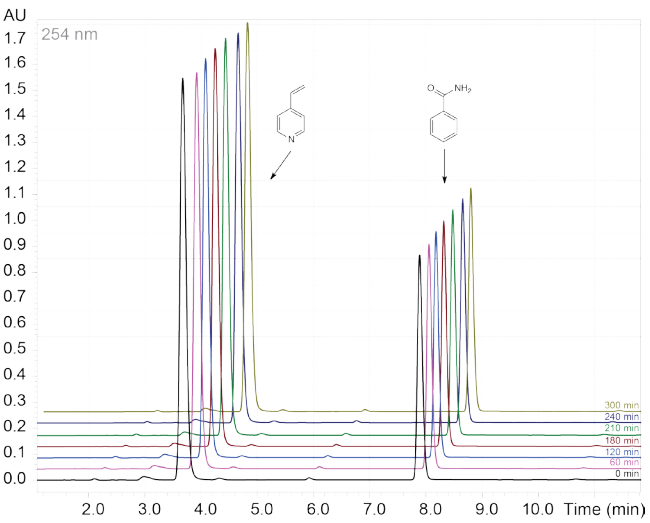
32


AU
254 nm
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0 22.5 25.0 Time (min)
300 min
240 min
180 min
120 min
60 min
0 min

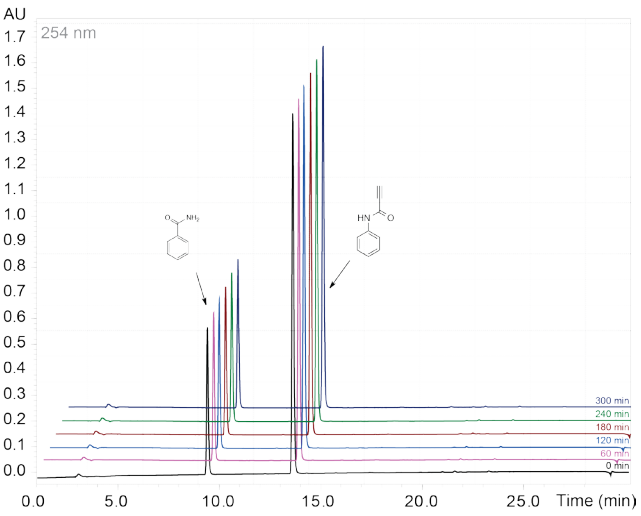
36


AU
254 nm
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
2.5 5.0 7.5 10.0 12.5 15.0 Time (min)
300 min
240 min
180 min
120 min
60 min
0 min

37

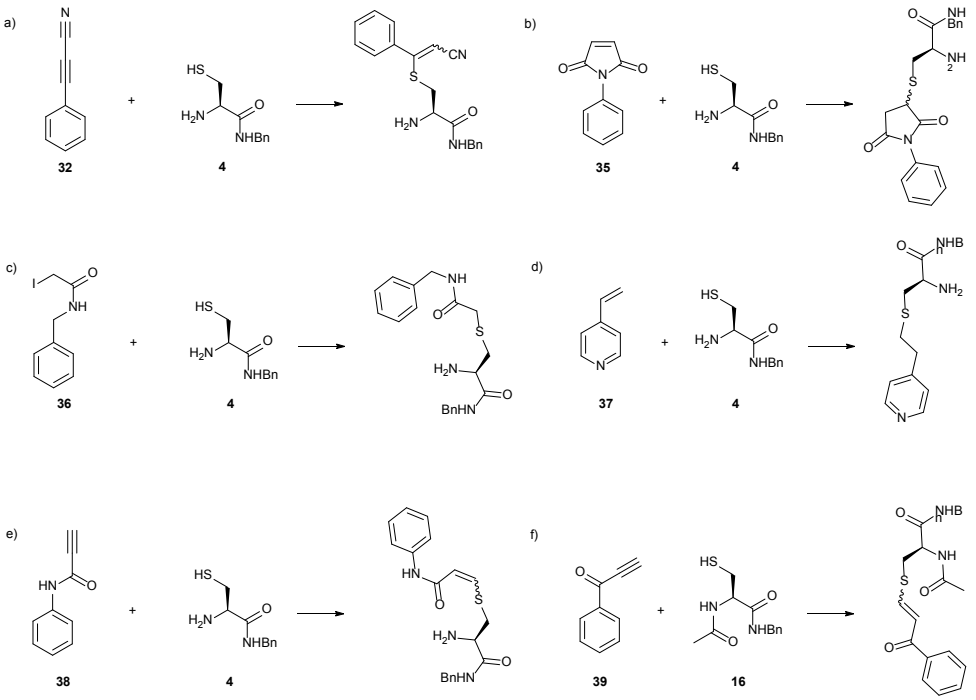

AU
254 nm
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 Time (min)
300 min
240 min
210 min
180 min
120 min
60 min
0 min

38


AU
254 nm
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 5.0 10.0 15.0 20.0 25.0 Time (min)
300 min
240 min
180 min
120 min
60 min
0 min

ANNEX 4: STUDY OF THE REACTIVITY OF CYSTEINE-SELECTIVE PROBES 32, 35-39.

STUDIED REACTIONS:



GENERAL CONSIDERATIONS:

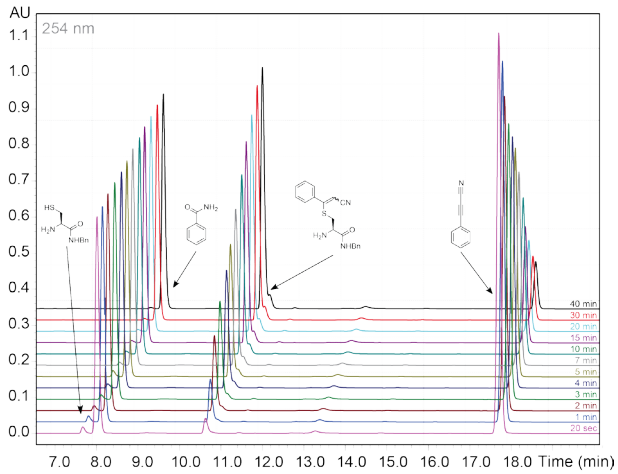
All reactions were monitored by HPLC at 100 μM concentration of reagents in PBS:DMSO (4:1) mixture (pH = 7.4) at 25 ˚C. Benzamide was used as an internal standard. The reaction of alkynone **37** with cysteine benzamide **4** gave a complex mixture, perhaps, because of formation of thiazoline derivative together with the product of thiol addition to the acetylene moiety, which is why acetylated cysteine model 16 was used instead.

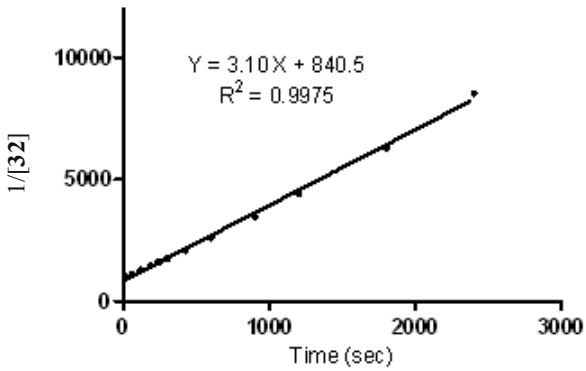
In cases of maleimide **33** and alkynone **37**, starting material and reaction product were hydrolysed during the quenching of the reaction with 5% TFA. As a consequence, only an estimated value of the second order constant was determined, based on measurements at 0 and 30 min of reaction (99% conversion was observed for electrophile). Constant value was estimated basing of following equation:

$$k_2 \geq \frac{\frac{1}{C_1} - \frac{1}{C_0}}{\tau} = \frac{\frac{1}{(1-0.99)*0.001} - \frac{1}{0.001}}{30*60} \geq 55\ (M^{-1}s^{-1}) > 50\ (M^{-1}s^{-1})$$

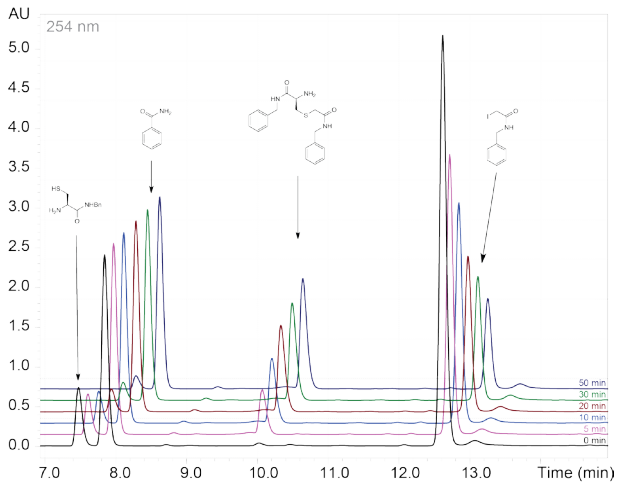
HPLC TRACES AND KINETICS PLOTS FOR HYDROLYTICALLY UNSTABLE SUBSTRATES:

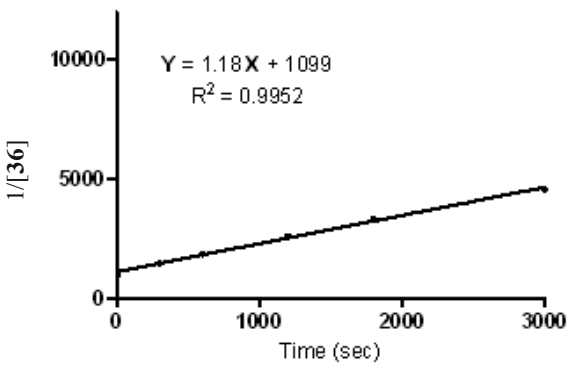
32



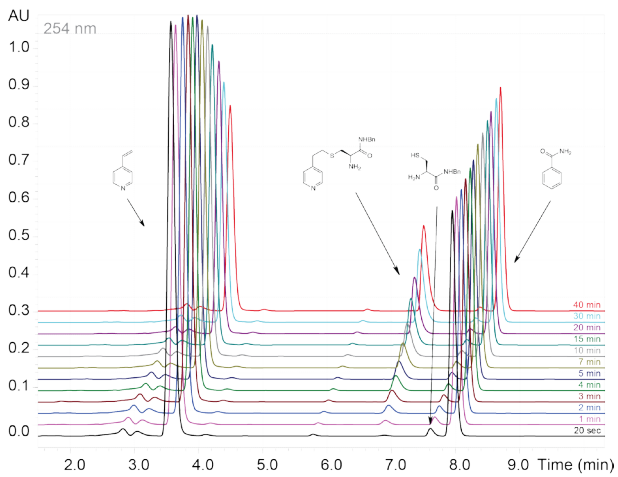


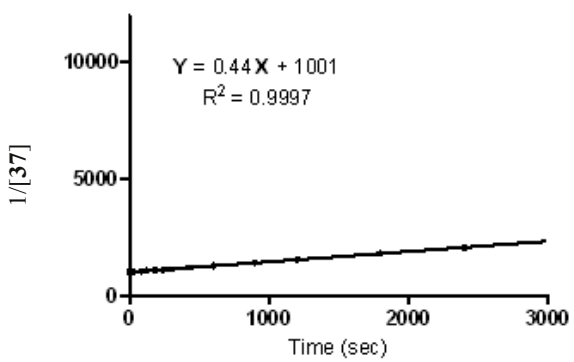
36



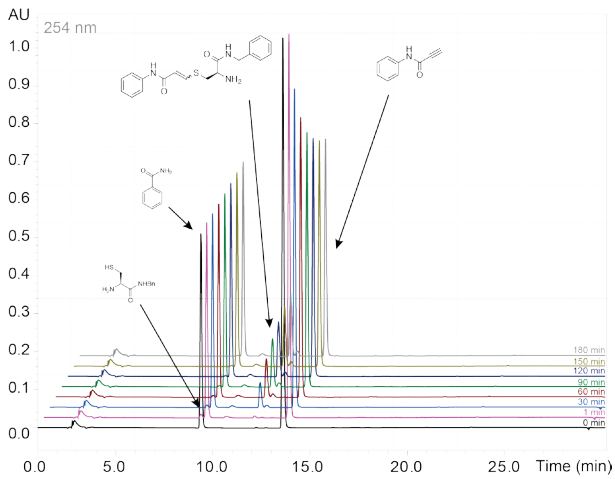


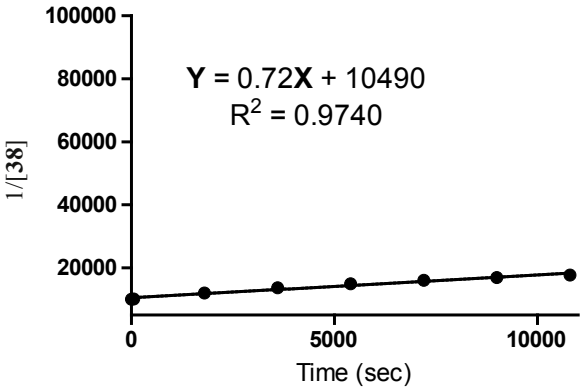
37



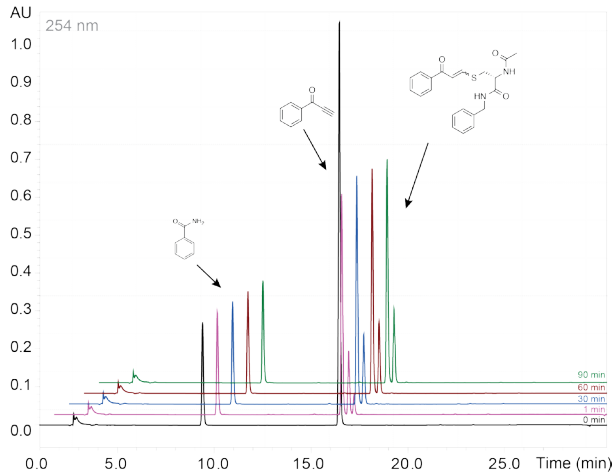


38


AU
254 nm
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
180 min
150 min
120 min
90 min
60 min
30 min
1 min
0 min
0.0
5.0
10.0
15.0
20.0
25.0
Time (min)


100000
80000
60000
40000
20000
1/[38]
Y = 0.72X + 10490
R2 = 0.9740
0
5000
10000
Time (sec)

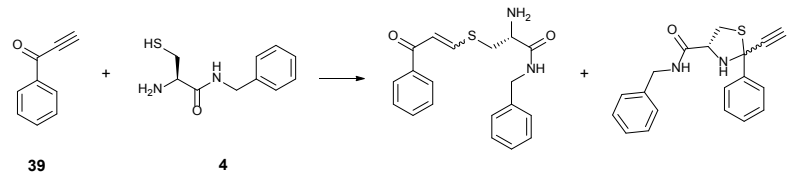
39


AU
254 nm
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
90 min
60 min
30 min
1 min
0 min
0.0
5.0
10.0
15.0
20.0
25.0
Time (min)

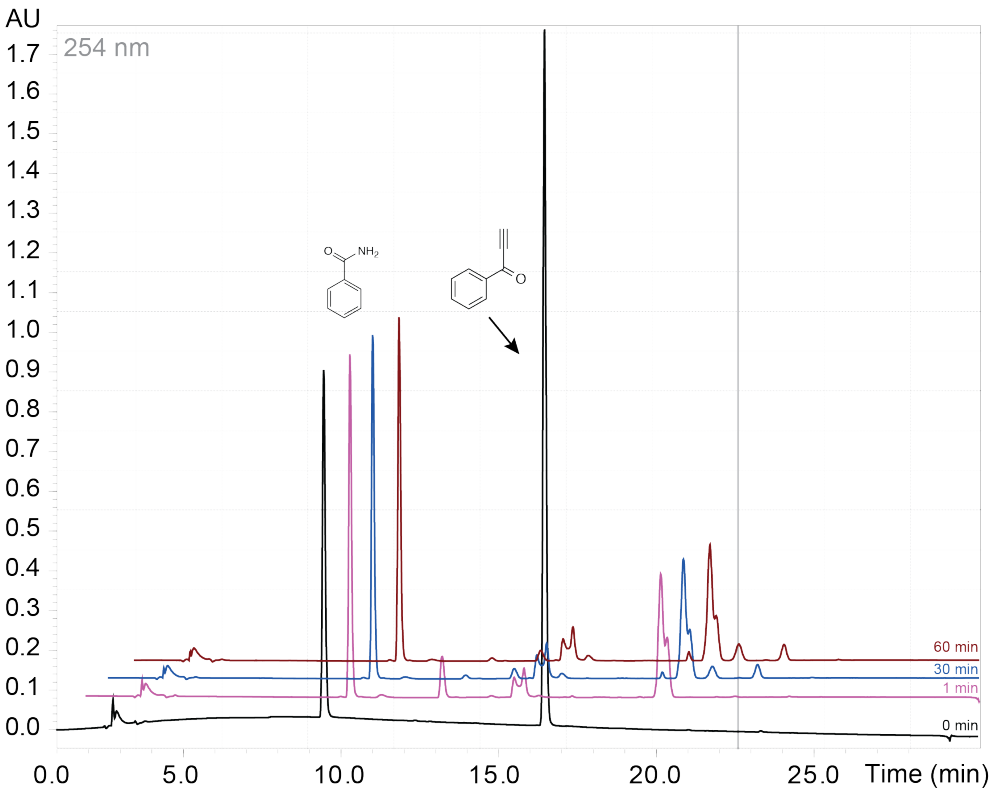
ANNEX 5: REACTION BETWEEN ALKYNONE 39 AND CYSTEINE DERIVATIVE 4.

The reaction of alkynone **39** with cysteine benzamide **4** gave a complex mixture, because of formation of thiazoline derivative together with the product of thiol addition to the acetylene moiety 8a.

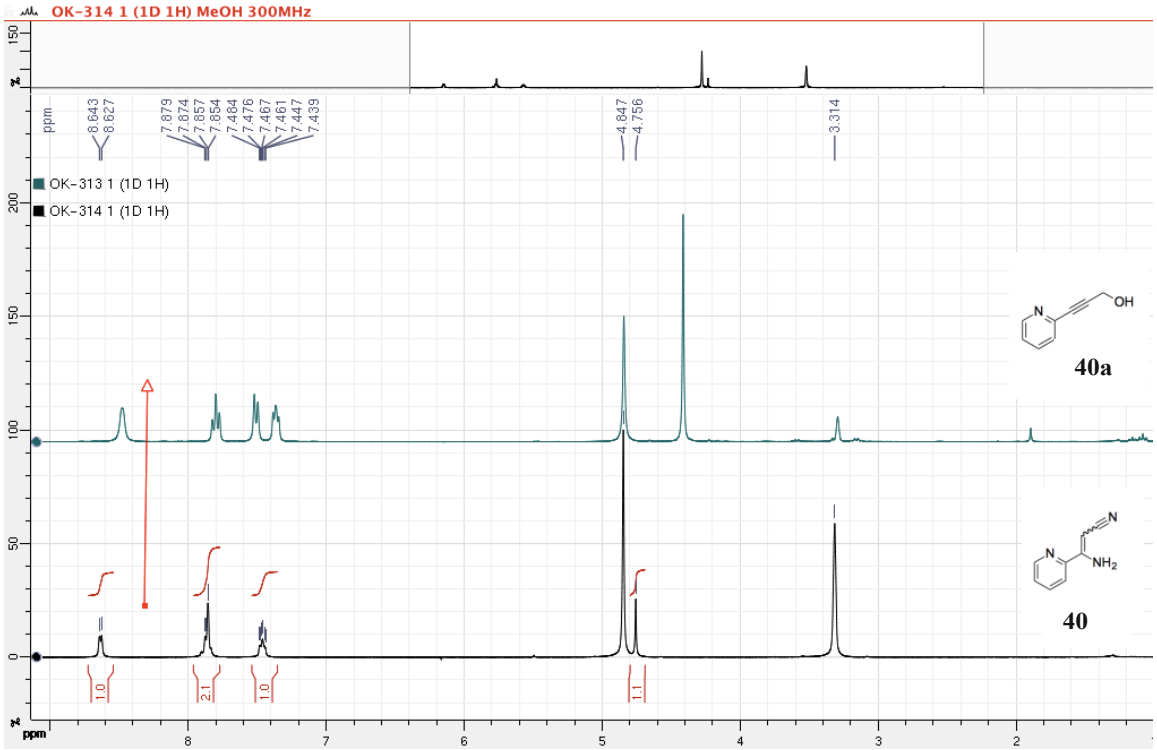
REACTION SCHEME:

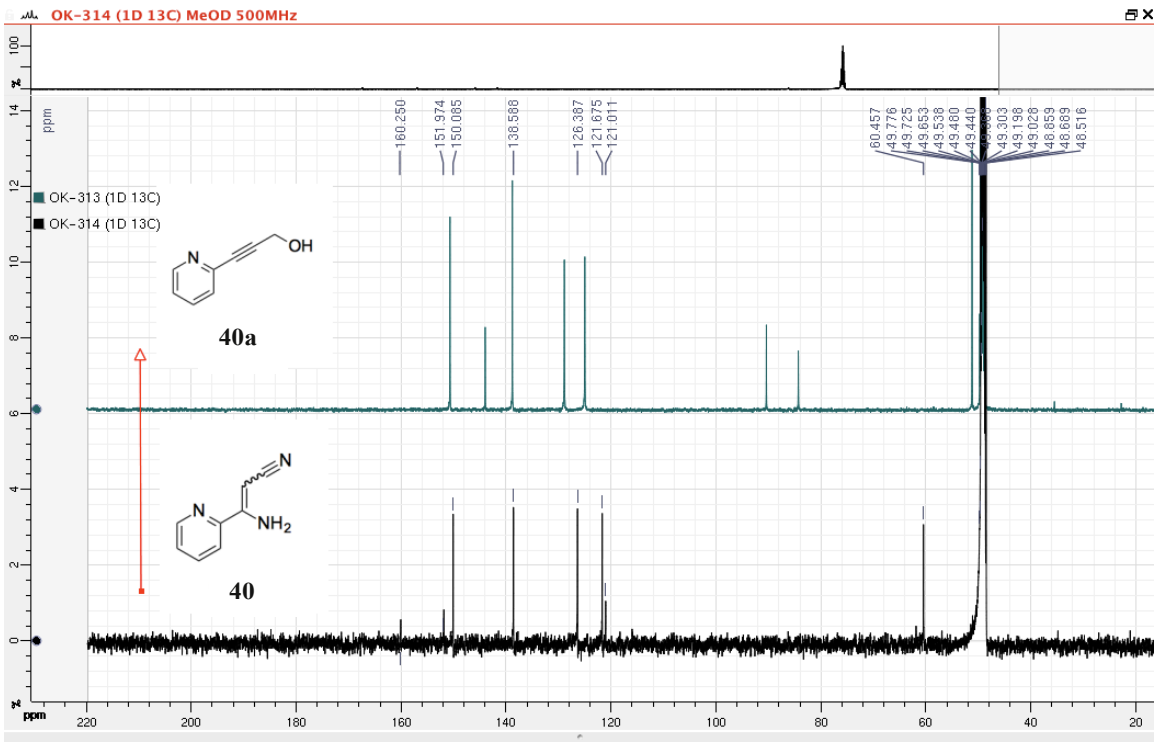


HPLC TRACES:

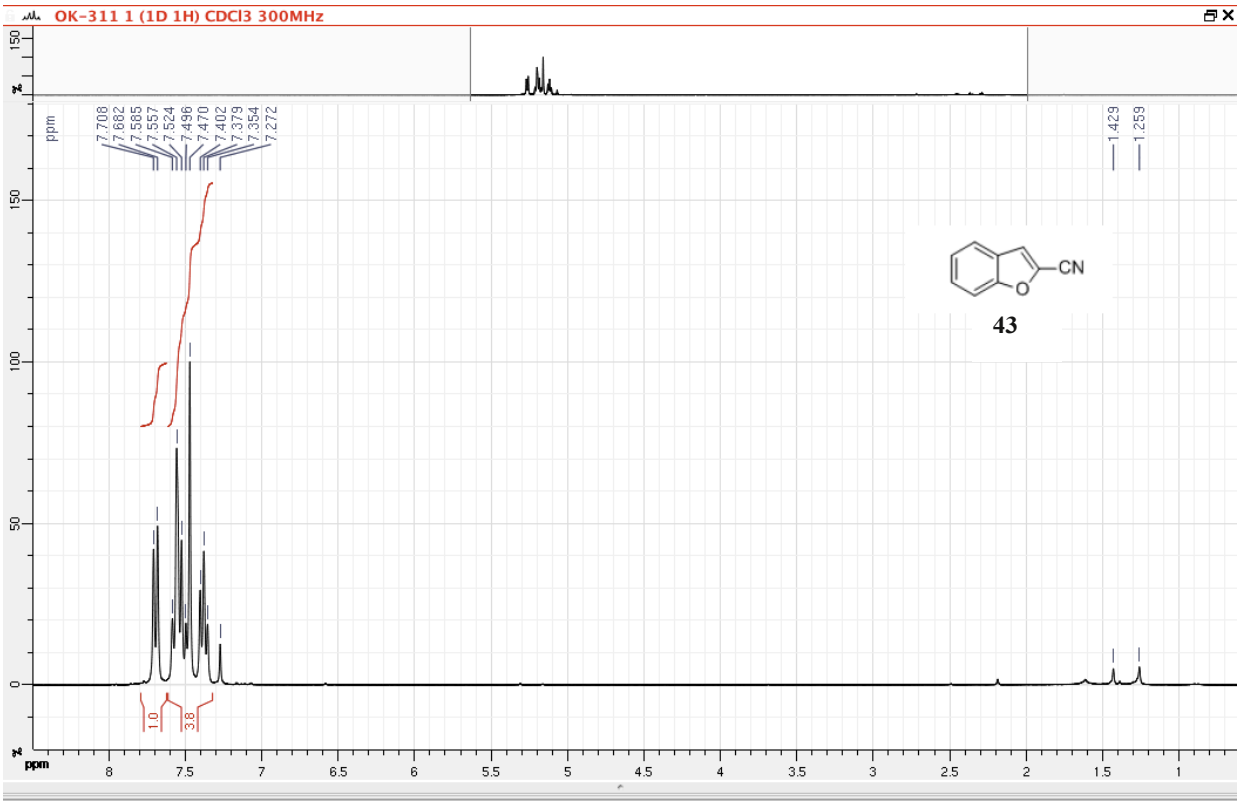


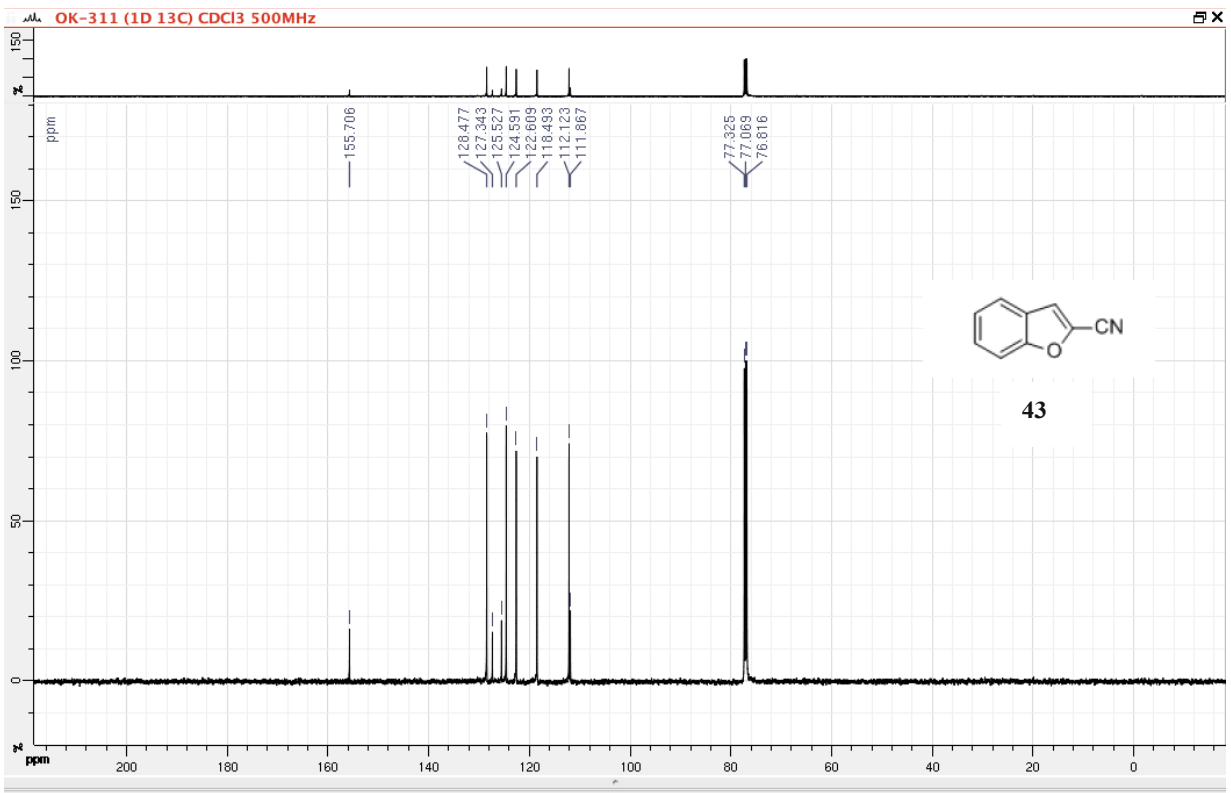
ANNEX 6: NMR SPECTRA OF 40A AND 40.





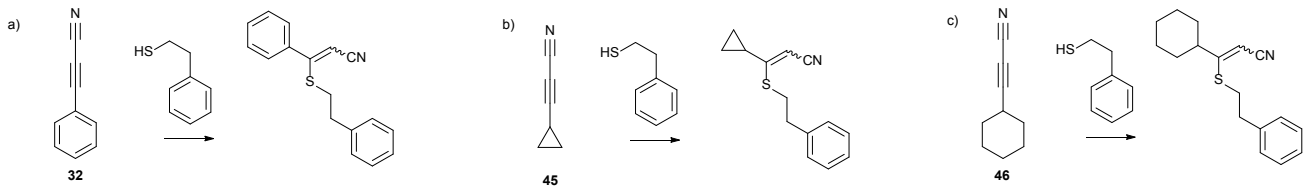
ANNEX 7: NMR SPECTRA OF 43.





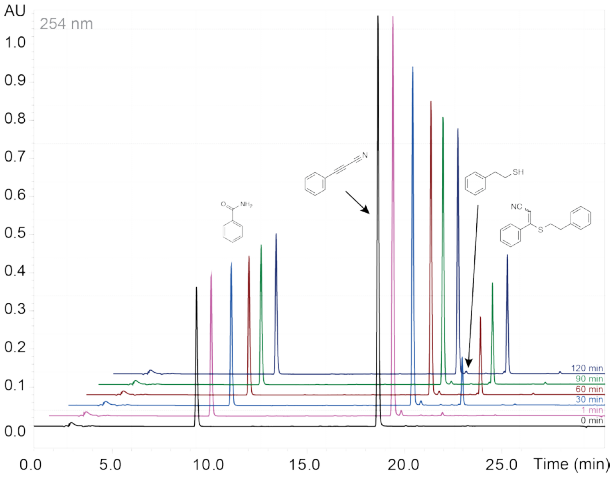
ANNEX 8: REACTIVITY OF 4-PHENYL-, 4-CYPROPYL-, AND 4-CYCLOHEXYL- SUBSTITUTED PROPIOLONITRILES TOWARDS PHENYLETHYLENEMERCAPTANE.

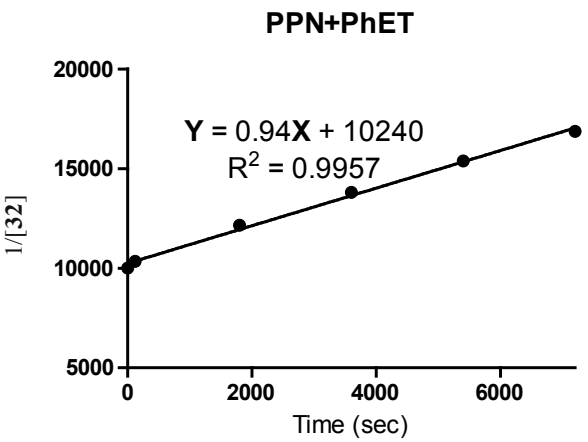
STUDIED REACTIONS:



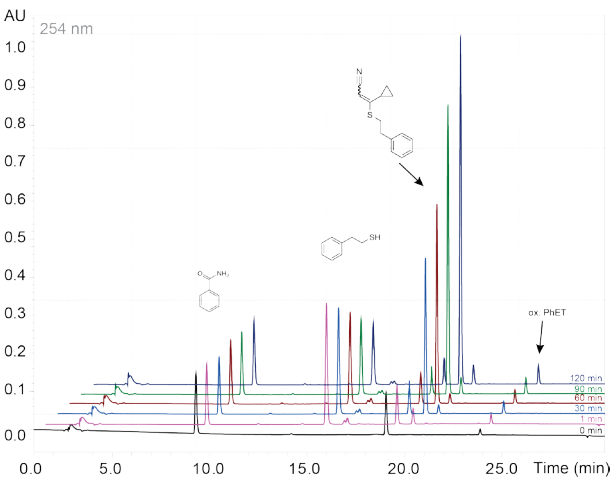
HPLC TRACES AND KINETICS PLOTS:

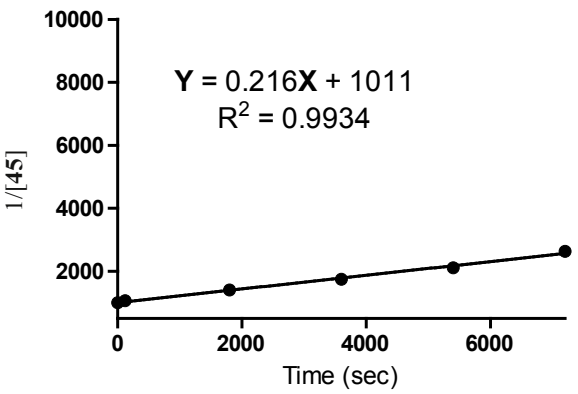
a)



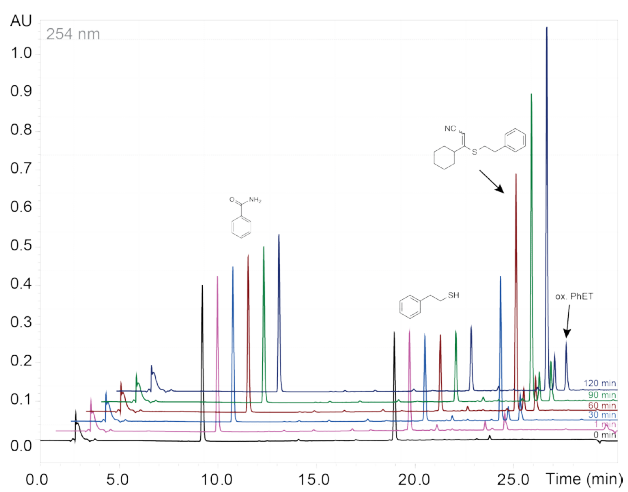


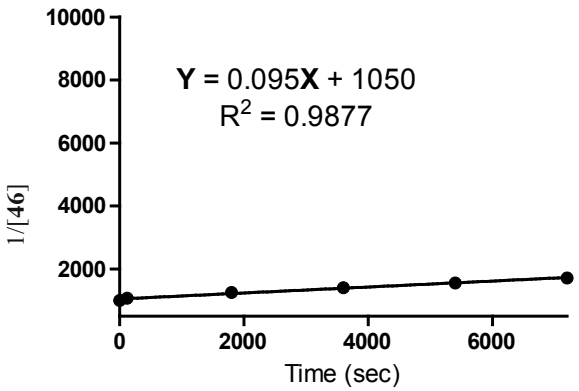
b)



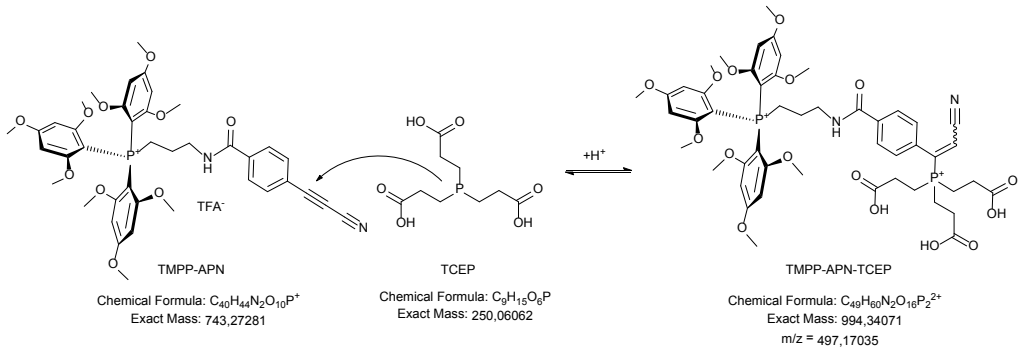


c)


AU
254 nm
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
5.0
10.0
15.0
20.0
25.0
Time (min)
ox. PhET
120 min
90 min
60 min
30 min
1 h
0 min

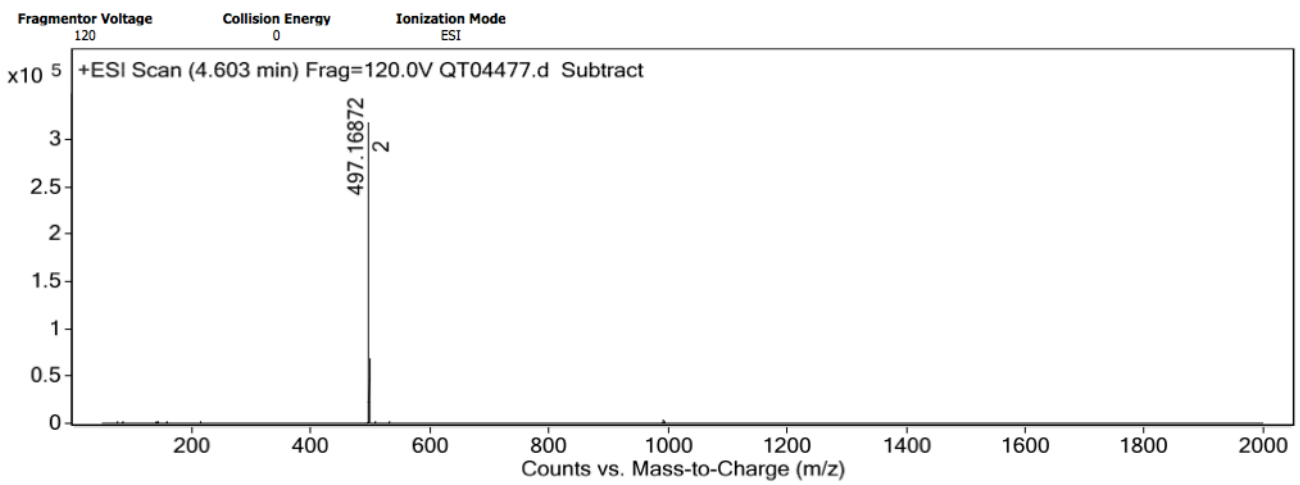

10000
8000
6000
4000
2000
Y = 0.095X + 1050
R2 = 0.9877
1/[46]
0
2000
4000
6000
Time (sec)

ANNEX 9: MS OF APN-TCEP ADDUCT.



The addition product TMPP-APN-TCEP was only detectable by MS.

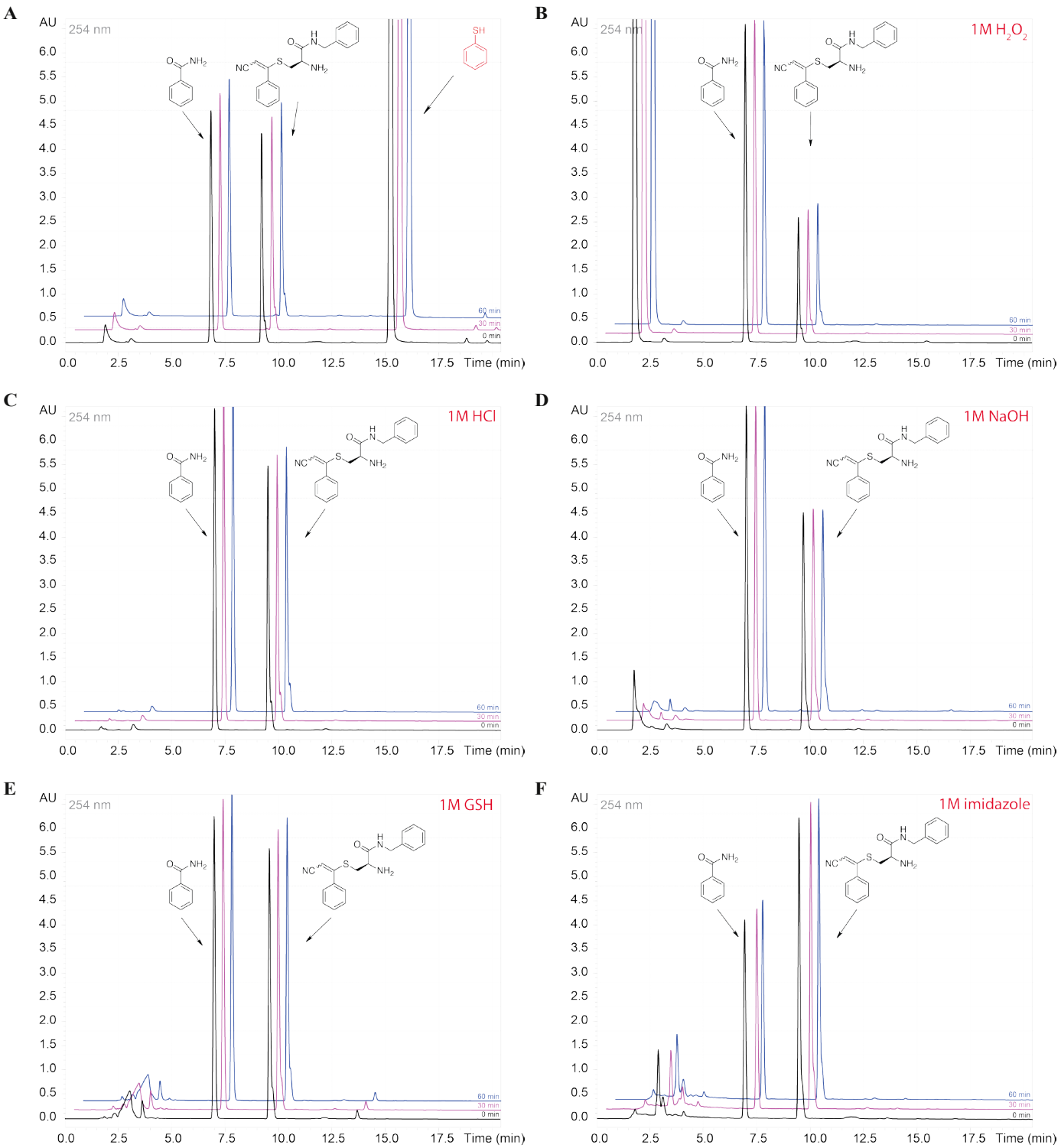
Qualitative Analysis Report

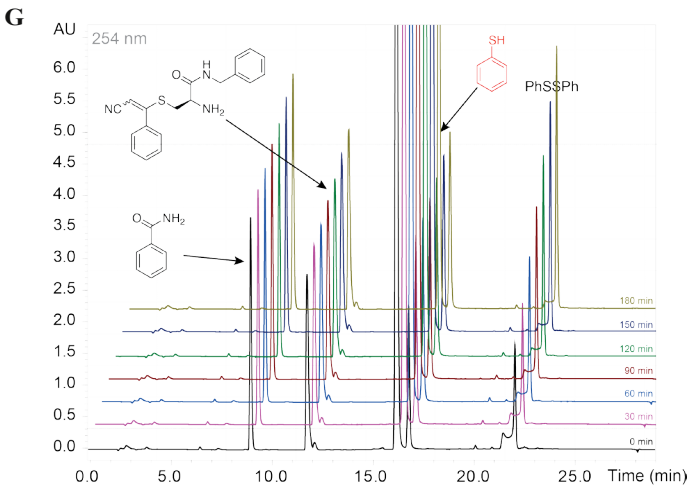


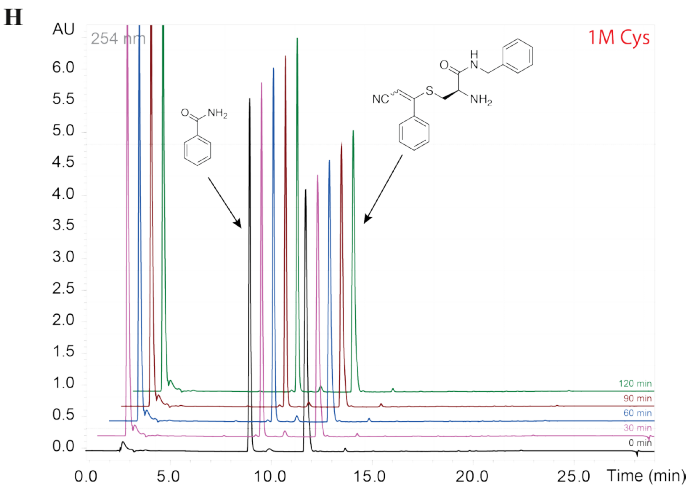
Peak List

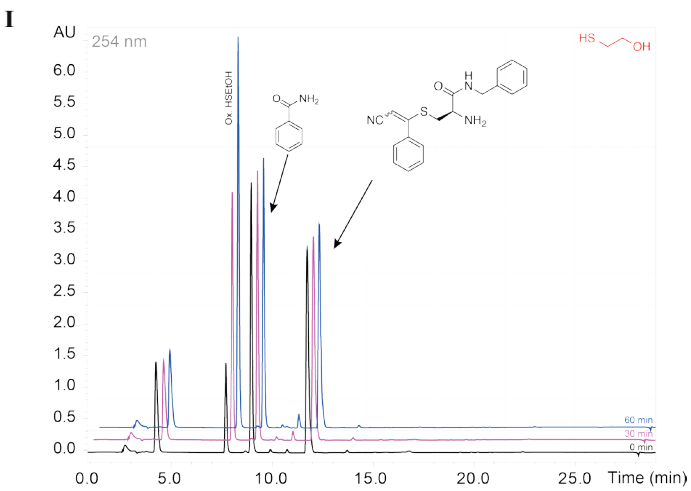
m/z	z	Abund
497.16872	2	320226.4
497.67058	2	188573.1
498.16898	2	69813.4
498.66598	2	16993

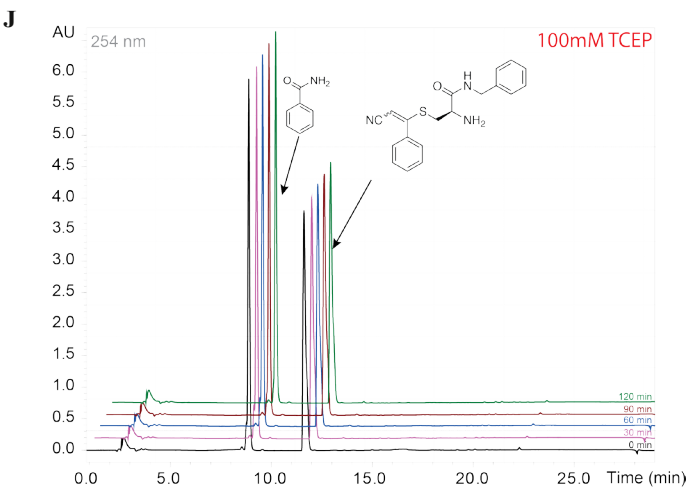
ANNEX 10: STABILITY OF ADDUCTS APN-CYSTEINE ADDUCT 33 (HPLC TRACES).

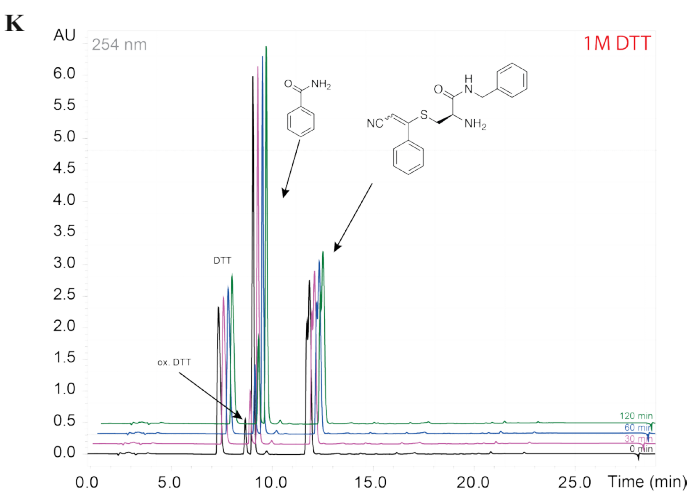



G
AU
254 nm
SH
PhSSPh
180 min
150 min
120 min
90 min
60 min
30 min
0 min
0.0 5.0 10.0 15.0 20.0 25.0 Time (min)


H
AU
254 nm
1M Cys
120 min
90 min
60 min
30 min
0 min
0.0 5.0 10.0 15.0 20.0 25.0 Time (min)

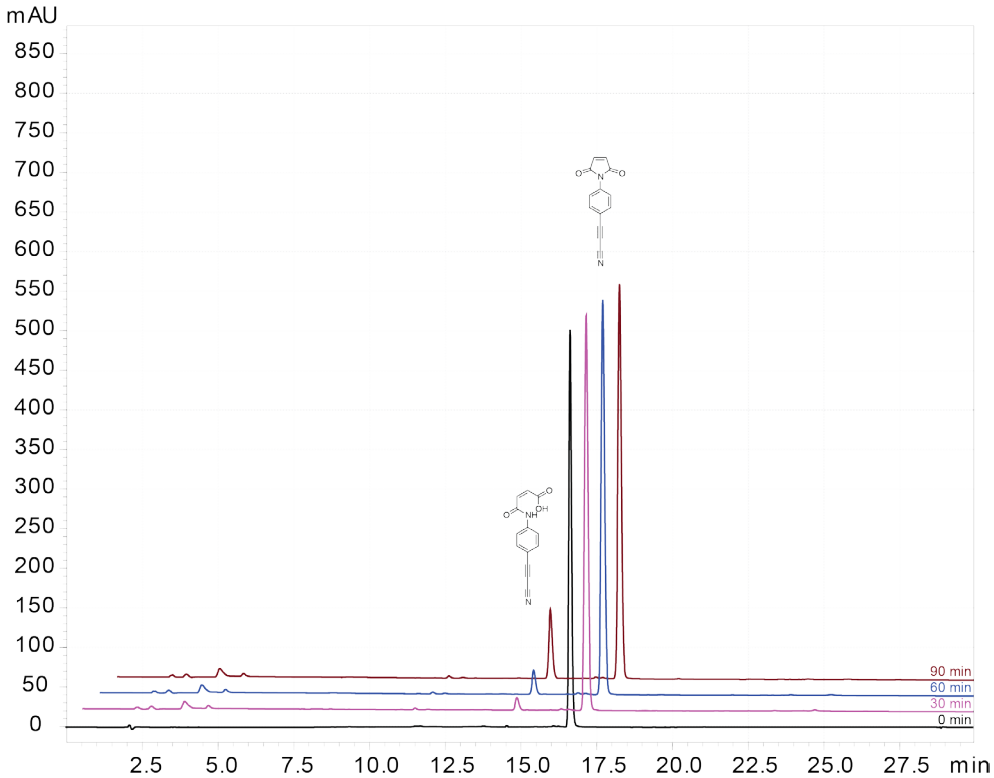

I
AU
254 nm
HS OH
Ox. HSEtOH
60 min
30 min
0 min
0.0 5.0 10.0 15.0 20.0 25.0 Time (min)


J
AU
254 nm
100mM TCEP
120 min
90 min
60 min
30 min
0 min
0.0 5.0 10.0 15.0 20.0 25.0 Time (min)

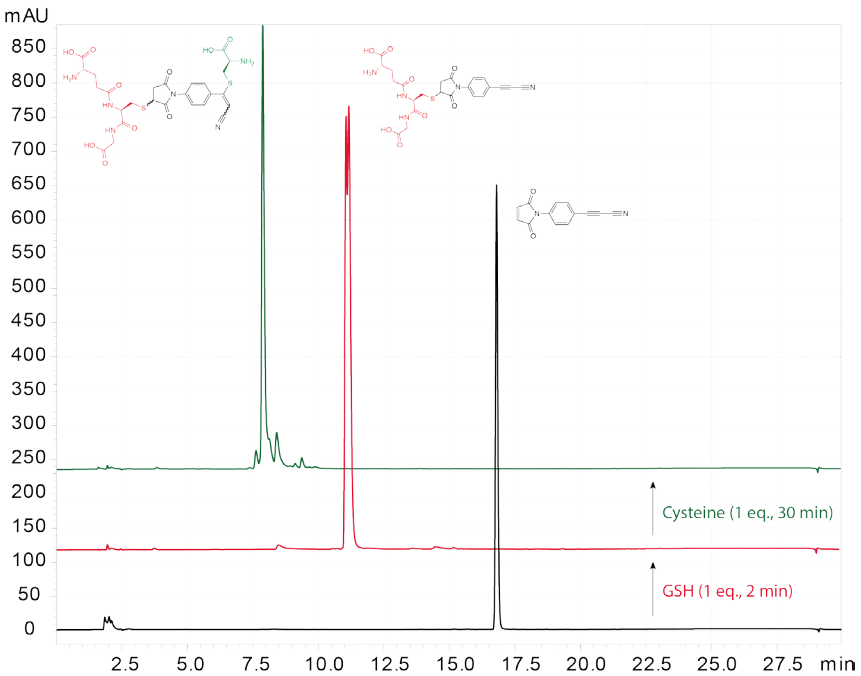

K
AU
254 nm
1M DTT
DTT
ox. DTT
120 min
60 min
30 min
0 min
0.0 5.0 10.0 15.0 20.0 25.0 Time (min)

ANNEX 11: APN-MALEIMIDE PROBE 71.

HYDROLYTIC STABILITY:

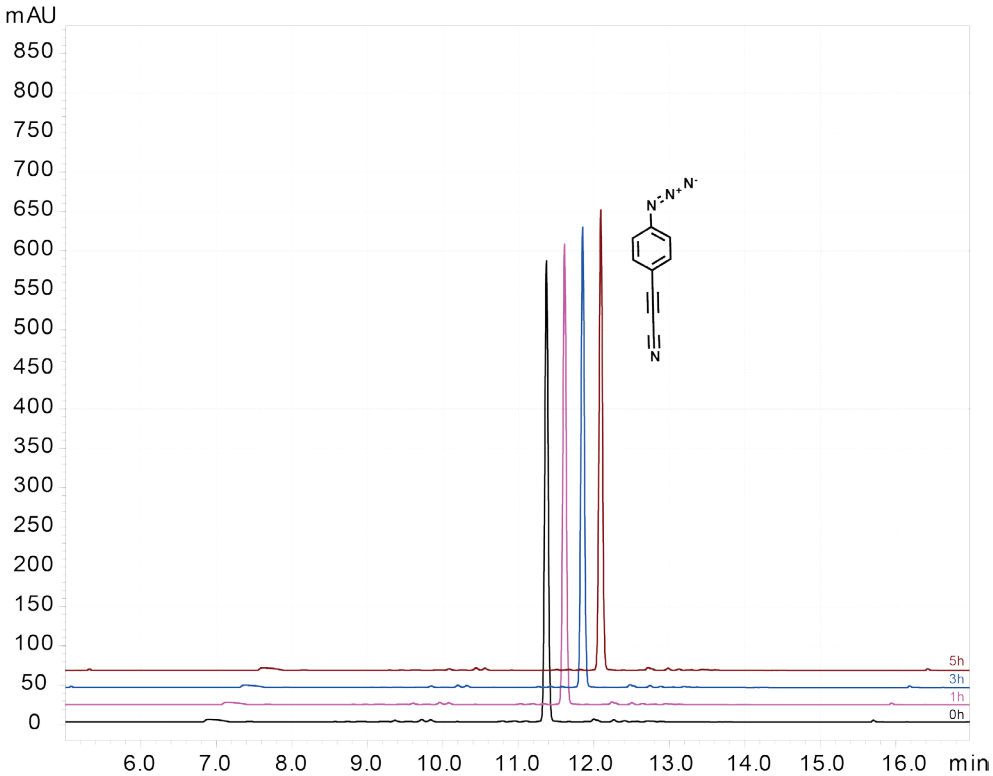


COUPLING OF GLUTATHIONE WITH CYSTEINE:

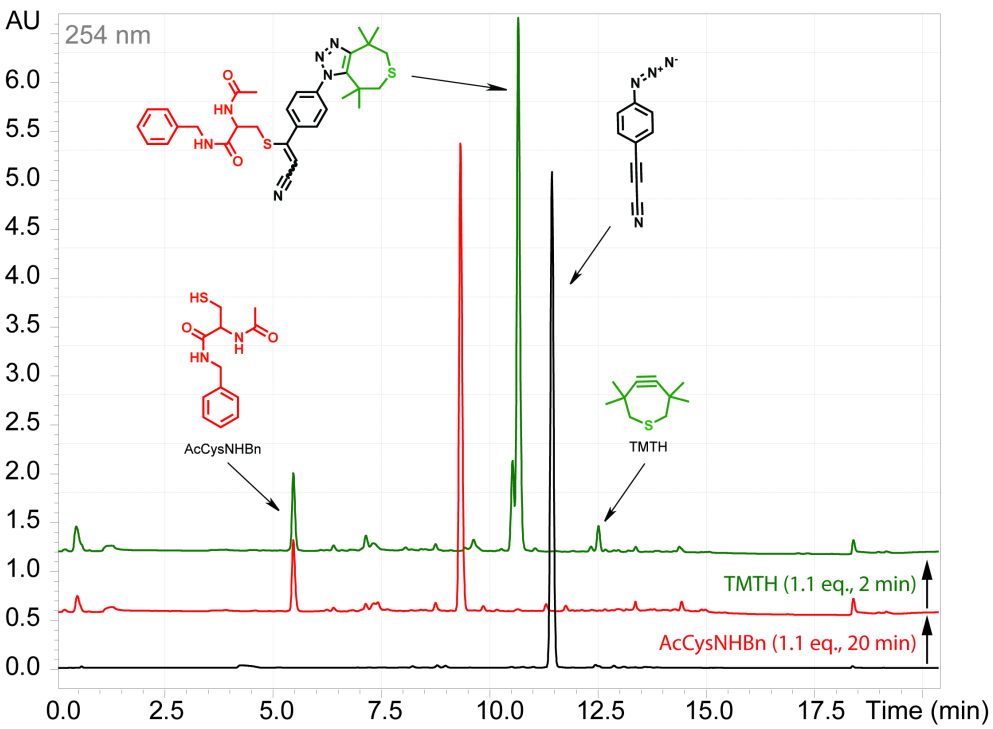


ANNEX 12: APN-AZIDE PROBE 76.

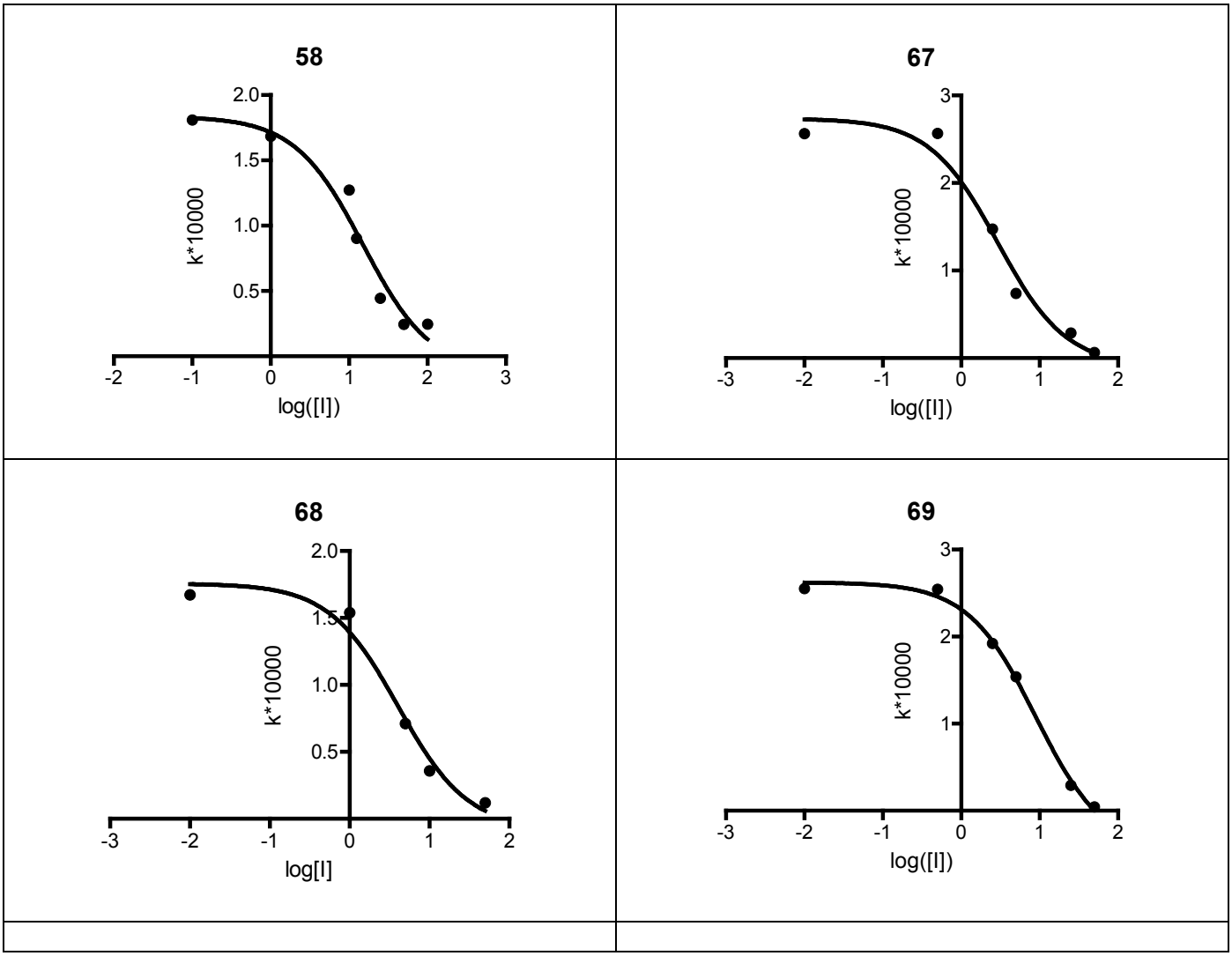
HYDROLYTIC STABILITY:



COUPLING OF TMTH WITH CYSTEINE DERIVATIVE 76:



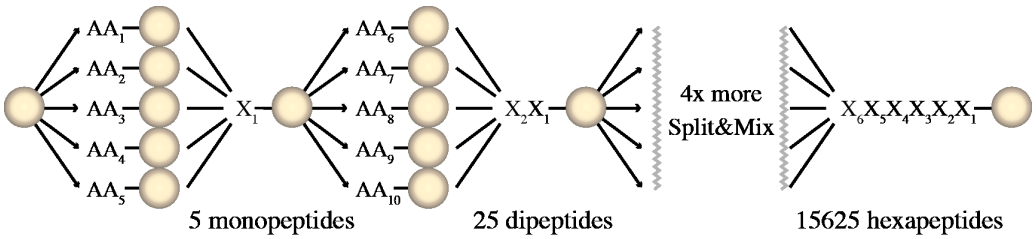
ANNEX 13: THIOREDOXINE REDUCTASE INHIBITION (IC50)



ANNEX 14: COMBIBATORIAL PEPTIDE LIBRARIES.

Hexapeptide libraries are synthesised following Split&Mix SPPS (see Section D 2.7, page 76) using TAGSFREE algorithm of "unique pairs".[837,838].

SCHEMATIC REPRESENTATION OF THE SYNTHESIS OF PEPTIDE LIBRARIES:



COMPOSITION OF THE SYNTHESISED LIBRARIES (CYS RESIDUES ARE SHOWN IN GREEN):

Library #1

X_1	X_2	X_3	X_4	X_5	X_6
Ala	Met	Arg	**Cys**	**Cys**	Lys
Tyr	Leu	Ser	Lys	Tyr	Ala
Pro	Phe	Pro	Arg	Ser	His
Leu	Gln	Asp	Phe	Met	Asp
Thr	Gly	Gly	Thr	His	Gln

Library #2

X_1	X_2	X_3	X_4	X_5	X_6
Lys	Asp	Ile	Ala	Ala	Ser
Arg	Gly	Gln	Ser	Arg	Lys
Phe	Thr	Phe	Ile	Glu	**Cys**
Gly	Tyr	His	Thr	Asp	His
Pro	Leu	Leu	Pro	**Cys**	Tyr

CONTROL OF THE BEADS:

Library #1 contained methionine that was found to represent several difficulties during its detection by amino acid analysis (AAA). The peak corresponding to Met residue is almost always has lower abundance than expected. This results in ambiguous data analysis and corresponding limitations of the method. **Library #2** is deprived from this drawback and was used in screening with cysteine-selective reagents.

Library #1

#	Composition by AAA	Sequence	Quality
1	Ser, His, Pro, Met, Trp, Tyr	SHMWPY	+
2	Gly, Arg, Ala, Cys, Ley, Lys	LCGKRA	+
3	Lys, Phe, Ala, Asn, Gly, Cys	DCKGFA or KCFDGA	+
4	Lys, Tyr, Ala, Thr, Gly, ***Met***	AYKG***M***T or KYTG***M***A	+/-
5	Glu, Arg, Tyr, Phe, Tyr, ***Met***	QYFR***M***Y or QY***C***RFY	+/-
6	Asn, Gly, His, Arg, Thr, Gln	QHRDGT or DHRGQT	+
7	Ser, Gly, His, Thr, Phe, ***Met***	HSCGFT or HSFGMT	+/-
8	His, Ala, Ala, Pro, Phe, Arg	AHRPFA	+
9	Asn, Asn, Tyr, Leu, Lys, Pro	DYKDLP	+
10	Gln, Arg, Ala, Pro, Tyr, Phe	QYRPFA or AYFRQP	+
11	His, Arg, Pro, Leu, Phe, ***Met***	H***M***FRLP or H***C***FRLP	+/-
12	Ser, Gly, Arg, Thr, Ala, Lys	ASKRGT or KSTRGA	+

Library #2

#	Composition by AAA	Sequence	Quality
13	Gly, Thr, Ala, Ile, Lys, Cys	CATIGK or KCAITG	+
14	Asp, Arg, Ala, Tyr, Leu, Leu	YNALLR	+
15	Asp, Ser, Arg, Thr, Leu, Lys	KNSLTA or SRTLNK	+
16	Gln, Ala, Tyr, Lys, Lys, Cys	KCAEYK	+
17	Ser, Gly, Gly, His, His, Ala	HASHGG	+
18	Arg, Thr, Tyr, Leu, Leu, Cys	YCTLLR	+
19	Asp, Gly, Arg, Pro, Tyr, Ile	YNPIGR or YRPING	+
20	Gly, Thr, Ala, Leu, Phe, Lys	KATLGF or KATFLG	+
21	Ser, Gly, Arg, Ala, Leu, Phe	SRAFLG or SRALGF	+
22	His, Arg, Pro, Tyr, Lys, Cys	CRPHYK or KCPHYR	+
24	His, Ala, Ala, Tyr, Ile, Phe	HAAIYF	+
25	Asp, Asp, His, Ala, Phe, Phe	HNAFNF	+