

**DESIGN, SYNTHESIS AND EVALUATION OF NEW
PHOSPHOTYROSINE MIMICS**

GE JINGYAN

**NATIONAL UNIVERSITY OF SINGAPORE
2012**

**DESIGN, SYNTHESIS AND EVALUATION OF NEW
PHOSPHOTYROSINE MIMICS**

GE JINGYAN
(B. Sc., Zhejiang University)

**A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DEPARTMENT OF CHEMISTRY
NATIONAL UNIVERSITY OF SINGAPORE
2012**

Acknowledgements

First of all, from the bottom of my heart, I would like to thank my supervisor and mentor - Prof. Yao Shao Qin for giving me the opportunity to work in his lab and always highly supporting my works. Because of his critical and comprehensive thinking, research is taking shape in my mind. He let me know the passion and aspiration on discovering things and pursuing ideas. His enthusiasm and guidance have led me to my accomplishments.

I would like to thank all members in Prof Yao group - both past and present, Mingyu, Junqi, Hongyan, Haibin, Lay Pheng, Wu Hao, Pengyu, Kalesh, Raja, Liu Kai, Candy, Liqian, Li Lin, Grace, Xiamin, Zhenkun, Su Ying, Chongjing, Xiaoyuan, Zhengqiu, Biwei, Jiaqi, Chelsea, Yu Miin, Zhu Ying, Ching Tian, Xiaohua, Mei Xuan, Mei Yin, David, Derek, Joo Leng, Yuhui, Choon Meng, Liang Xian, Kaijia , Farhana, Cindy, Pei Xin, Jigang, Cathy who made this time so special and enjoyable. I thank particularly the following people: Junqi who have been great impact on the research attitude working in the lab; Haibin, Mingyu and Pengyu for not only being mentored me on research but also helping me the life in Singapore; Lay Pheng and Wu Hao who taught me biological experiments; Xiamin who taught me mass related techniques and helped me synthesis of trifunctional tag; Dr Li Lin who help me do the bioimaging. Thank all of you for helping and accompanying these years, providing me a nice and friendly home - Yao lab in Singapore.

I would like to thank Prof. Lu Yixin and Dr Mahesh for writing the recommendation letters. I appreciate the support from National University of Singapore for providing me research scholarship.

Finally, I would like to thank my parents, my sister and Mr Tu Jinjun, for their understanding, supports and encouragement throughout these years.

Big thanks to all of you!

Table of Contents

ACKNOWLEDGEMENTS	i
TABLE OF CONTENTS	iii
SUMMARY	viii
LIST OF PUBLICATIONS	ix
LIST OF FIGURES	xi
LIST OF TABLES	xiv
LIST OF SCHEMES	xv
LIST OF ABBREVIATIONS AND SYMBOLS	xvi
LIST OF AMINO ACIDS	xxiii
CHAPTER 1 INTRODUCTION	1
1.1 Protein Tyrosine Phosphatases	3
1.1.1 PTP family	3
1.1.2 Structure of PTP catalytic domain and mechanism of PTP catalysis	4
1.2 SH2 domains	7
1.3 Mimics of <i>p</i> Tyr	9
1.3.1 Fluorogenic mimics of <i>p</i> Tyr	9
1.3.2 Caged <i>p</i> Tyr	11
1.3.3 Reversible non-hydrolyzable mimics	13
1.3.4 Irreversible <i>p</i> Tyr mimics	18
1.4 Activity-based protein profiling (ABPP)	19
1.4.1 Activity based probes for PTPs	23

1.5	Objectives	26
-----	------------	----

CHAPTER 2 AN UNNATURAL AMINO ACID THAT MIMICS

PHOSPHOTYROSINE 27

	Summary	27
--	---------	----

2.1	Introduction	27
-----	--------------	----

2.2	Synthesis of the unnatural amino acid	30
-----	---------------------------------------	----

2.3	Solid-phase synthesis of inhibitors	32
-----	-------------------------------------	----

2.4	Biological experiments and discussion	33
-----	---------------------------------------	----

2.4.1	Docking Stimulations	33
-------	----------------------	----

2.4.2	Fluorescence polarization (FP) experiments	33
-------	--	----

2.4.3	Inhibition assays of PTP1B	35
-------	----------------------------	----

2.4.4	Hydrolytic stability test	36
-------	---------------------------	----

2.4.5	Cell permeability test	37
-------	------------------------	----

2.4.6	Cell proliferation	39
-------	--------------------	----

2.5	Conclusion	41
-----	------------	----

CHAPTER 3 A SELF-IMMOBILIZING AND FLUOROGENIC

UNNATURAL AMINO ACID THAT MIMICS PHOSPHOTYROSINE 42

	Summary	42
--	---------	----

3.1	Introduction	42
-----	--------------	----

3.2	Synthesis of unnatural amino acid	45
-----	-----------------------------------	----

3.3	Solid-phase synthesis of three localization peptides	47
-----	--	----

3.4	Biological experiments and discussion	49
-----	---------------------------------------	----

3.4.1	UV-initiated uncaging experiment	49
-------	----------------------------------	----

3.4.2	Dephosphorylation reaction with uncaged compound	51
-------	--	----

3.4.3	Enzyme specificity	52
-------	--------------------	----

3.4.4	Labeling experiments with pure PTP1B	53
3.4.4	Labeling experiments with bacterial lysate	54
3.4.5	Competitive gel analysis	55
3.4.6	Mass spectrometry	56
3.4.7	Effects of the activities of PTP1B as a model	57
3.4.8	Comparison of enzyme activities before and after probe reaction	58
3.4.9	FACS experiments	59
3.4.10	Microscopy experiments	63
3.5	Conclusion	65

**CHAPTER 4 UGI REACTION-ASSISTED RAPID ASSEMBLY OF
AFFINITY-BASED PROBES AGAINST POTENTIAL PROTEIN
TYROSINE PHOSPHATASES**

		66
	Summary	66
4.1	Introduction	66
4.2	Synthesis of two components-aldehyde and isonitrile	70
4.3	Biological experiments and discussion	72
4.3.1	Inhibition assay	72
4.3.2	Fingerprint profiling of PTP1B and MptpB	75
4.3.4	<i>In vitro</i> labeling mammalian proteomes and pull-down/LC-MS/MS analysis	81
4.4	Conclusion	84

**CHAPTER 5 UNEXPECTED PHOSPHOTYROSINE MIMICKING
PROBES TARGET PROTEIN DISULFIDE ISOMERASE IN
DIFFERENT CELL LINES**

	Summary	85
--	---------	----

5.1	Introduction	85
5.2	Synthesis of four probes	89
5.3	Biological experiments and discussion	90
5.3.1	<i>In situ</i> proteomic profiling and in-gel fluorescence scanning	90
5.3.2	Co-localization bioimaging	93
5.3.3	Pull-down/LCMS/MS experiment	94
5.3.4	Immunofluorescence	96
5.3.5	Protein disulfide isomerase	98
5.4.5	<i>In situ</i> proteomic profiling and pull-down experiments of different cell lines	102
5.4.5	Cell proliferation-XTT assay	103
5.5	Conclusion	104
	CHAPTER 6 EXPERIMENTAL SECTION	105
6.1	General Information	105
6.2	Synthesis of unnatural amino acid 2-6' and relatives	106
6.3	Solid-phase synthesis of I1 and I2	111
6.5	Synthesis of unnatural amino acid 3-2 and relatives	115
6.6	Solid phase synthesis of three localization peptides	125
6.6.1	General information	125
6.6.2	General procedure for solid phase synthesis	125
6.7	Synthesis of aldehyde and isonitrile for MCR reaction	129
6.8	Synthesis of TER-azide	136
6.8	MCR reaction	136
6.8.1	General procedure for MCR reaction	136
6.8.2	LCMS Data	141

6.8.3	^1H NMR Data	149
6.9	Synthesis of four probes	154
6.10	Synthesis of trifunctional tag	163
6.11	Mass results	166
CHAPTER 7 CONCLUSION REMARKS		171
CHAPTER 8 REFERENCES		174

Summary

Reversible protein phosphorylation plays a fundamental role in signal transduction. Protein phosphatases, protein kinases and some modular domains (SH2, PDB, etc) form key components of the highly complex human phosphoproteome network. Dysregulation of this network is known to cause many human diseases including cancer. The phosphotyrosine (*p*Tyr) residue in a protein serves as one of the key recognition and binding elements for signaling proteins. Its hydrolytic instability and poor cell permeability has led to the development of many novel phosphate mimics. This thesis described several *p*Tyr mimics, most of which possess suitable N- and C-terminus for conjugation/incorporation into peptide/protein synthesis, including 1) isoxazole carboxylic acid-based mimics that possess better cell permeability and hydrolytic stability; 2) quinone methide-based mimics containing fluorogenic and self-immobilizing properties; 3) vinyl sulfonate-based mimics having irreversible covalent labeling reactivity. These small *p*Tyr mimics were subsequently incorporated into either small molecules or peptides to provide expanded chemical tools for further chemical biology applications, such as bioimaging, FACS and activity based-profiling. Furthermore, we successfully explored the multi-component Ugi reaction for rapid assembly of a panel of protein tyrosine phosphatase-targeting affinity-based probes (AfBPs).

List of Publications

(2007 - 2012)

- 1 Kalesh, K. A.; Shi, H.; Ge, J.; Yao, S. Q.* “The Use of Click Chemistry in the Emerging Field of Catalomics”, *Org. Biol. Chem.*, **2010**, 8, 1749-1762
- 2 Ge, J.; Wu, H.; Yao, S. Q.* “An Unnatural Amino Acid That Mimics Phosphotyrosine”, *Chem. Commun.*, **2010**, 46, 2980-2982
- 3 Wu, H.; Ge, J.; Yao, S. Q.* “High-Throughput Small Molecule Microarray Identify Inhibitory Binder Against 14-3-3 Protein”, *Angew. Chem. Intl. Ed.*, **2010**, 49, 6528-6532
- 4 Wu, H.; Ge, J.; Yang, P.-Y.; Yao, S. Q.* “A Peptide Aldehyde Microarray for High-Throughput Discovery of Cysteine Protease Inhibitors”, *J. Am. Chem. Soc.*, **2011**, 133, 1946-1954
- 5 Wu, H.; Ge, J.; Yao, S. Q.* “Small Molecule Microarray: The First Decade and Beyond”, *Chem. Commun.*, **2011**, 47, 5664-5670
- 6 Ge, J.; Li, L.; Yao, S. Q.* “A Self-Immobilizing and Fluorogenic Unnatural Amino Acid that Mimics Phosphotyrosine”, *Chem. Commun.*, **2011**, 47, 10939-10941

- 7 Ge, J.; Cheng, X.; Tan, L. P.; Yao, S. Q.* “Ugi Reaction-Assisted Rapid Assembly of Affinity-Based Probes against Potential Protein Tyrosine Phosphatase”, *Chem. Commun.*, **2012**, 48, 4453-4455

- 8 Li, L.; Ge, J.; Wu, H.; Xu, Q. -H.; Yao, S.Q.* “Organelle-Specific Detection of Phosphatase Activities with Two-Photon Fluorogenic Probes in Cells and Tissues”, *J. Am. Chem. Soc.*, **2012**, Submitted

- 9 Ge, J.; Cheng, X. -M.; Li L.; Yao, S. Q.* “Unexpected Phosphotyrosine Mimicking Probes Target Protein Disulfide Isomerase in Different Cell Lines”, **2012**, In preparation

List of Figures

Figures	Page
1.1 Phosphorylation and dephosphorylation on Ser/Thr/Tyr catalyzed by Kinases and Phosphatases.	2
1.2 Schematic representation of the catalytic mechanism of PTP1B.	6
1.3 The effect of SH2 domains in SHP2 protein.	8
1.4 Fluorogenic <i>p</i> Tyr mimics.	10
1.5 Caged <i>p</i> Tyr mimics.	12
1.6 Pmp and F ₂ Pmp based inhibitors.	14
1.7 PTP1B inhibitors with carboxylic acid-based <i>p</i> Tyr bioisosteres.	15
1.8 Different <i>p</i> Tyr mimics-based binders of Grb2 SH2 domain.	16
1.9 Improvement of the inhibition and cellular activity by discovering new <i>p</i> Tyr mimics.	17
1.10 Inhibitors containing <i>p</i> Tyr bioisosteres.	18
1.11 Irreversible <i>p</i> Tyr mimics.	19
1.12 Overview of some developed bioorthogonal reactions used in the tag-free approach.	20
1.13 Proteome profiling with activity-based probes (ABPs) and affinity-based probes (AfBPs).	22
1.14 Quinone methide-based approach.	24
1.15 Bromobenzylphosphonate and vinyl sulfone/sulfonate based probes.	25
2.1 Structures of <i>p</i> Tyr, its bioisosteres and the one in this study.	29
2.2 Molecular docking of I1 and I2 .	33
2.3 Coomassie gel of the purified STAT3 SH2 domain.	34
2.4 IC ₅₀ of two inhibitors determined by FP experiments.	35

2.5	IC ₅₀ graphs of I2 against PTP1B.	36
2.6	LC profiling of hydrolytic stability test.	37
2.7	T47D cell viability in the presence of the two inhibitors.	40
3.1	Representative examples of previously reported fluorogenic PTP substrates, 2-FMPT and Withers' sugar probe.	43
3.2	Overall working principle of the <i>p</i> Tyr mimic 3-2 and its peptide probes.	45
3.3	LC profiles and MS results of uncaging process.	50
3.4	Time-dependent fluorescence measurements of UV-irradiated 3-2 and pER .	52
3.5	Microplate-based enzymatic assay of PTP1B and other four enzymes.	53
3.6	Labeling of PTP1B with different concentrations of compound 3-2' .	54
3.7	Fluorescence gel of PTP1B-overexpressed bacterial lysates treated with uncaged 3-2 and 3-1 .	55
3.8	Indirect competition assay.	
3.9	MS profiling of PTP1B before and after incubation with uncaged 3-2 .	57
3.10	Comparison of PTP1B's activity, with and without compound 3-1' and 3-2' .	59
3.11	Histogram of flow cytometry analysis of dead cells after UV-irradiation.	60
3.12	Density-plot of FACS analysis of HeLa cells treated with 3-1 and 3-2 .	61
3.13	FACS analysis of five cell lines.	63
3.14	Fluorescence images of HeLa cells treated with different non-irradiated peptide probes.	64
4.1	Structures of representative PTP1B/MptpB bidentate inhibitors.	67
4.2	Chemical synthesis of the 25 A/BPs using Ugi reaction.	68
4.3	The A/BP work flow.	70
4.4	IC ₅₀ graphs of all 25 A/BPs against PTP1B and MptpB.	74

4.5	Optimization of labeling of pure PTPs.	78
4.6	Affinity-based profiling of recombinant PTP1B and MptpB against the 25 A/BPs.	80
4.7	Results of proteome profiling of P23 against MCF-7 lysate.	83
5.1	Overall strategy of probe design.	87
5.2	Synthetic schemes of four probes.	88
5.3	The structure of Tri-N ₃ .	89
5.4	<i>In situ</i> proteome-profiling of different probes against MCF-7.	91
5.5	Dose and time-dependent <i>in situ</i> proteome profilings.	92
5.6	Organelle co-localization bioimaging experiments.	94
5.7	Fluorescence image of MCF-7 pull-down experiments after 5-1 .	96
5.8	Immunofluorescence experiments were performed with anti-PDIA1.	97
5.9	IC ₅₀ curves for 5-1 against PDI.	99
5.10	Concentration-dependent labeling of pure PDI protein by 5-1 .	100
5.11	Labeling PDI with different pretreatment.	101
5.12	<i>In situ</i> labeling results of four cell lines.	102
5.13	<i>In situ</i> pull-down results of four cell lines.	103
5.14	Effects of 5-1 on the cell proliferation of four cell lines.	104
6.1	The analytical HPLC results of compound 2-6' .	111
6.2	The LCMS results of compound I1 .	112
6.3	The analytical HPLC results of compound I2 .	114

List of Tables

Table		Page
2.1	Optimization of isoxazole ring formation under Cu(I)-catalyzed, 1,3-dipolar cycloaddition conditions.	31
2.2	The results of cell permeability.	38
3.1	Kinetic data of compound 3-1' and 3-2' with PTP1B.	58
4.1	Summary of IC ₅₀ values of 25 AfBPs against PTP1B/MptpB.	75
5.1	List of proteins identified from mass spectrometry of three specific bands.	95
6.1	Summary of AfBPs and their characterizations.	137
6.2	The complete list of P23 labeled proteins identified from MCF-7 cell lysate.	166

List of Schemes

Scheme		Page
2.1	Synthesis of 2-6' - the Fmoc-protected form of 2-6 .	31
2.2	Solid-phase synthesis of I2 .	32
3.1	Chemical synthesis of 3-2 .	46
3.2	Chemical synthesis of 3-1 .	47
3.3	Fmoc-based SPPS of the three CPP-containing probes.	48
4.1	Chemical synthesis of 4-1 and 4-2 .	71
6.1	Synthetic scheme of trifunctional tag.	163

List of Abbreviations and Symbols

2-FMPT	2-Fluoromethyl phosphotyrosine
ABP	Activity based probe
ABPP	Activity based protein profiling
A _f BP	affinity-based probe
AIDS	Acquired immunodeficiency syndrome
ALS	Amyotrophic lateral sclerosis
Boc	<i>tert</i> -Butoxycarbonyl
br	Broad
BSA	Bovine serum albumin
CBB	Coomassie brilliant blue-stained gel
CC	Click chemistry
CuAAC	Cu(I)-catalyzed 1,3-dipolar cycloaddition
CPP	Cell-penetrating peptide
Cs ₂ CO ₃	Cesium carbonate
C-terminus	Carboxyl terminus
CuI	Copper(I) iodide
DAST	Diethylaminosulfur trifluoride
DBU	1,8-Diazobicyclo[5.4.0]undec-7-ene
DCM	Dichloromethane
dd	Doublet of doublets
DIC	<i>N,N'</i> -diisopropylcarbodiimide
DIEA	<i>N, N'</i> -diisopropylethylamine
DIFMU	6,8-Difluoro-7-hydroxy-4-methylcoumarin

DIFMUP	6,8-Difluoro-4-methylumbelliferyl phosphate
DMAP	4-Dimethylaminopyridine
DMEM	Dulbecco's Modified Eagle Medium
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DSP	Dual-specificity phosphatase
DTT	Dithiothreitol
E. coli	<i>Escherichia coli</i>
EA/EtOAc	Ethyl acetate
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide HCl
EDT	1,2-ethanedithiol
EDTA	Ethylenediamine tetraacetic acid
EGF	Epidermal growth factor
equiv	Equivalent
ER	Endoplasmic Reticulum
ESI	Electron Spray Ionization
Et	Ethyl
Et ₂ O	diethyl ether
EtOH	Ethanol
F2Pmp	Phosphonodifluoromethyl phenylalanine
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
FL	Fluorescent gel
Fmoc	9-Fluorenylmethoxycarbonyl
FMPP	2-Fluoromethyl phenyl phosphate

FMT	Fluoro- <i>O</i> -malonyltyrosine
FP	Fluorescence polarization
g	Gram
Grb2	Growth factor receptor-bound protein 2
HATU	O-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium
HBSS	Hank's balanced salt solution
HBTU	<i>O</i> -Benzotriazole- <i>N,N,N',N'</i> -tetramethyl-uronium- hexafluorophosphate
HCl	Hydrogen chloride
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human immunodeficiency virus
HOAT	1-Hydroxy-7-azabenzotriazole
HOBT	<i>N</i> -Hydroxybenzotriazole
HPLC	High performance liquid chromatography
Hz	Hertz
IC ₅₀	Half the maximal inhibitory concentration
IF	Immunofluorescence
K ₂ CO ₃	potassium carbonate
K _d	Dissociation constant
K _M	Michaelis constant
KHCO ₃	potassium hydrogen carbonate
LC	Liquid chromatography
LC-MS	Liquid chromatography-mass spectrometry
LiOH	Lithium hydroxide
LMW	Low molecular weight

M	Molar
m	Multiplet
MCR	multi-component reaction
MeOH	Methanol
mg	Milligram
min	Minute
mM	Millimolar
mM	Micromolar
mmol	Millimole
MptpB	<i>Mycobacterium</i> protein tyrosine phosphatase B
MS	Mass
MUP	Methyl Umbelliferone Phosphate
MW	Molecular weight
Na ₂ SO ₄ .	Sodium sulfate
NaBH ₄	Sodium borohydride
NaCl	Sodium chloride
NaHCO ₃	Sodium bicarbonate
NaN ₃	Sodium azide
NaOH	Sodium hydroxide
Ni-NTA	Nickel-nitrilotriacetic acid
nM	Nanomolar
NMR	Nuclear magnetic resonance
NPE	1-(2-nitrophenyl)ethyl
N-terminus	Amino terminus
OMT	<i>O</i> -malonyltyrosine

OTs	<i>p</i> -toluenesulfonyl
P loop	phosphate-binding loop
PBS	Phosphate buffered saline
pCAP	Phosphocoumaryl amino propionic acid
pCAPF	8-fluorocoumarin deviation of pCAP
PDIA1	Protein disulfide isomerase A1
PDIA4	Protein disulfide isomerase A4
PDIA6	Protein disulfide isomerase A6
PEG	Polyethylene glycol
<i>p</i> H	Negative logarithm of the hydroxonium ion concentration
Ph	Phenyl
PK	Protein kinase
Pmp	Phosphonomethyl-phenylalanine
PMS	Phenazine methosulfate
PPI	Protein-protein interaction
ppm	Parts per million
PP	Protein phosphatase
PTB	Phosphotyrosine binding
PTK	Protein tyrosine kinase
PTM	Post-translational modification
PTP1B	Protein tyrosine phosphatase 1B
PTP	Protein tyrosine phosphatase
<i>p</i> Tyr.	Phosphotyrosine
PVS	Phenyl vinyl sulfonate
q	Quartet

s	Singlet
S.D.	Standard deviation
sat.	Saturated
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SH2	Src homology domain
SHP1	Src homology region 2 domain-containing phosphatase-1
SHP2	Src homology region 2 domain-containing phosphatase-2
SPPS	Solid phase peptide synthesis
STAT	Signal transducer and activator of transcription
STS	Staurosporine
t	Triplet
TCPTP	T-cell protein tyrosine phosphatase
TB	Tuberculosis
TBAF	Tetrabutylammonium fluoride
TBST	Tris buffered saline with Tween-20
TBS/TBDMS	<i>tert</i> -Butyldimethylsilyl
TBTA	tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine
tBu	<i>tert</i> -Butyl
<i>t</i> BuOH	<i>tert</i> -Butyl alcohol
TCEP	Tri(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TIS	Triisopropylsilane
TLC	Thin layer chromatography
TMS	Trimethylsilyl

TOF	Time of flight
Tris	Tris(hydroxymethyl) amino methane
UV	Ultraviolet
XTT	2,3-Bis-(2-methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide
Å	angstrom
s	seconds
min	minutes
h	hours
n	nano
mol	mole
°C	degree celsius
λ_{em}	Emission wavelength
λ_{ex}	Excitation wavelength
λ_{abs}	Absorption wavelength

List of Amino Acids

One Letter	Three Letter	Amino Acid
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
r	D-Arg	D-Arginine
F _x	-	Cyclohexylalanine

Chapter 1 Introduction

The human genome project is regarded as a milestone in the journey toward deciphering the entire story of life. DNA is a unit of heredity in a living organism, but it could not fully interpret the whole life and living processes. Functional annotation and characterization of proteins have received much attention in recent years, since proteins play a critical role in virtually every biological process, including cell-cell communication, cell death and cell division. The large-scale study of proteins, particularly their structures and functions is called “proteomics”. Of special importance are enzymes, which as biocatalysts play fundamental roles in regulating almost every cellular processes and metabolic transformation. Dysfunctions of enzymes are known to cause major human diseases including cancers, AIDS and Parkinson diseases. Consequently, it is not surprising that enzymes account for more than 25% of all current drug targets.¹

Enzymes catalyze most post-translational modification (PTM) events in biological processes. The PTMs are regulatory processes by the chemical modification of a protein after its translation, usually through the covalent addition of a small molecule to one of the amino acid residues.² Among different posttranslational events involved in cellular signaling pathways, phosphorylation/dephosphorylation of a protein is a key reversible modification that regulates protein activity, localization, degradation and complex formation. The identification and characterization of proteins containing phosphate groups are named phosphoproteomics^{2c}, a branch of proteomics. The phosphoproteomics network is extremely complex with a composition of around 500 mammalian protein kinases³, 100 protein phosphatases⁴

and hundreds of proteins containing modular signaling domains⁵ (SH2⁶, PTB⁷, 14-3-3 domains⁸, etc.), which interact with phosphorylated proteins. More than 30% of all human proteins are estimated to be phosphorylated. Abnormal levels of protein phosphorylation are linked to a wide range of diseases. Consequently, proteins involved in the phosphoproteomics network (e.g. kinases, phosphatases, phosphoserine/threonine/tyrosine binding proteins) have become attractive targets for therapeutic developments.^{3,4} The tyrosine residue is one of the important sites which could be tagged with a phosphate group by protein kinases, referred to as phosphotyrosine. This chapter will provide a brief overview of some phosphotyrosine (*p*Tyr) related proteins and mainly focus on the development and utilities of *p*Tyr mimics.

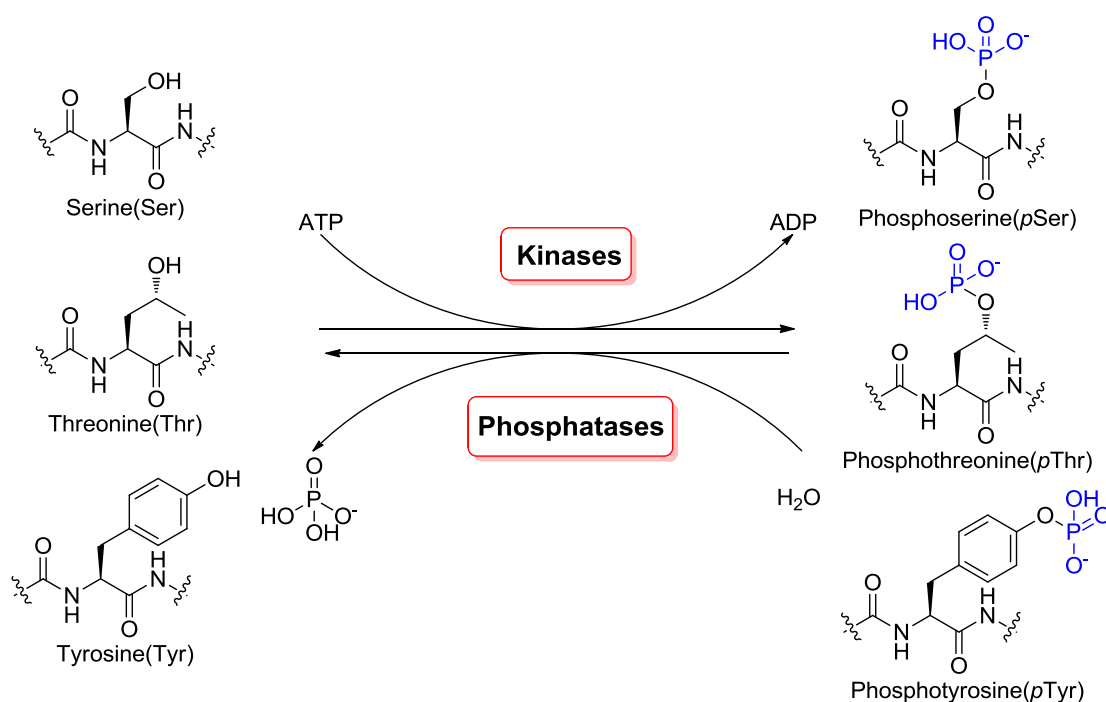


Figure 1.1 Phosphorylation and dephosphorylation on Ser/Thr/Tyr catalyzed by Kinases and Phosphatases.

1.1 Protein Tyrosine Phosphatases

Reversible protein phosphorylation and dephosphorylation reactions (shown in Figure 1.1) are catalyzed by protein kinases (PKs), which are responsible for phosphorylation, and protein phosphatases (PPs), which govern protein dephosphorylation, respectively. There are two main classes of protein phosphatases based on their structure and substrate specificity, namely protein serine/threonine phosphatases (PSTPs) and protein tyrosine phosphatases (PTPs). Among these phosphatases, PTPs, which catalyze the dephosphorylation of *p*Tyr residues in proteins, are a structurally diverse and highly specific family of enzymes. There are a total of 107 PTPs encoded in the human genome. Evidences have shown that most PTPs are related to human diseases⁹, for example, PTP1B acts as a negative regulator of the insulin and leptin receptor pathways, CD45 is associated with autoimmune diseases and LAR downregulates signaling from multiple receptor tyrosine kinases. Consequently, the development of inhibitors against PTPs is a proven approach in the discovery of new therapeutic agents.^{9,10}

1.1.1 PTP family

The PTP super-family could be broadly categorized into at least four subfamilies: 1) classical PTPs, 2) dual-specificity phosphatases (DSPs), 3) Cdc25 phosphatases, and 4) low-molecular-weight (LMW) PTPs.¹⁰ The classical PTPs (~ 38 members) are specific for phosphotyrosine substrates and are sub-divided into non-receptor and receptor-like forms based on their subcellular localizations. The non-receptor PTPs such as PTP1B, TCPTP, SHP1 and SHP2 are non-transmembrane PTPs which have other extra sequences for specific subcellular localization or binding target proteins. Receptor-like PTPs have a single transmembrane domain and other

extracellular domains. These kinds of PTPs may have one or two catalytic domains.^{9a} The second catalytic domain is normally inactive but plays a role in regulating the activity, specificity and stability of the PTPs. In contrast to the classic PTPs, dual-specificity phosphatases dephosphorylate not only on *p*Tyr but also on *p*Ser/*p*Thr residues in proteins.¹¹ Cdc25 phosphatases which could also dephosphorylate both *p*Tyr and *p*Thr residues, have different primary and tertiary structures and are hence grouped differently.¹² The overexpression of Cdc25 has been detected in multiple cancers and correlated with poor prognosis. LMW PTPs are a family of 18 kD proteins involved in cell growth regulation and share no significant differences except the PTP catalytic domain when compared with other members.¹³

1.1.2 Structure of PTP catalytic domain and mechanism of PTP catalysis

One of the key features to define the PTP family is their conserved catalytic domain (active site) which contains signature motif (H/V)C(X)₅R(S/T).^{4d, 10b, 14} This motif is located in the phosphate-binding loop (P loop), where the cysteine residue acts as a nucleophile and reacts with the phosphate head during catalysis. The arginine residue functions in both binding to the substrate and stabilizing the phosphoenzyme intermediates. There are other conserved loops, such as the *p*Tyr recognition loop and WPD (Trp-Pro-Asp) loop. The recognition loop is created to selectively target *p*Tyr as its substrate and not *p*Ser and *p*Thr residues in the target protein. The WPD loop contains a general acid-base catalyst (Asp) which takes part in the dephosphorylation reaction and could also affect the conformation change of PTPs. Upon binding to the substrate, the WPD loop closes over the phenyl ring of the tyrosine residue, holding it as a tight binding pocket to facilitate subsequent reactions. Mutations in this loop,

such as the W354A mutation in *Yersinia* PTP and the P425A mutation in SHP1, are known to attenuate or even disable the PTP's enzymatic activity.

The general mechanism of PTP catalysis has been explained with the aid of PTP structural information.¹⁵ PTP1B, one of the best studied PTPs and a novel target for type 2 diabetes, is used as a model to illustrate the mechanism (Figure 1.2). The phosphate head of *p*Tyr is first stabilized by the positive charges in Arg221 in the P loop. Upon binding, as mentioned above, PTP1B changes its conformation and closes the WPD loop to further stabilize the interaction. Asp181 in the WPD loop forms a hydrogen bond with the phenolic oxygen of *p*Tyr. At this moment, the Cys215 serves as a nucleophile and attacks *p*Tyr in the P loop, with Asp181 acting as a general acid by protonation of the leaving group on the substrate.¹⁶ Subsequently, a water molecule attacks the thiophosphoryl intermediate, liberating an inorganic phosphate and regenerating the active enzyme.¹⁶ In this step, Asp181 residue acts as a general base and deprotonates the water molecule.¹⁷

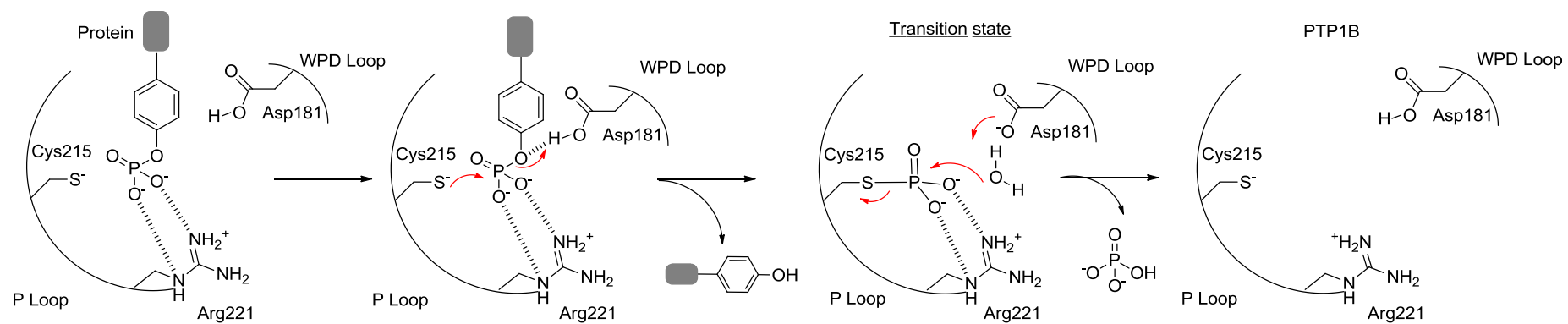


Figure 1.2 Schematic representation of the catalytic mechanism of PTP1B.

1.2 SH2 domains

The Src homology domains (SH2 domains) originally identified in numerous oncoproteins, such as Src and Fps, recognize *p*Tyr and serve as the prototype for phosphorylation-dependent protein-protein interaction (PPI) modules that organize the multi-protein complexes in the cell signaling pathway.¹⁸ In recent years, SH2 domains have become attractive targets for the development of new PPI inhibitors and potential therapeutic drugs.¹⁹

SH2 domains are relative small protein modules with a sequence of approximately 100 amino acids. The human genome encodes 116 SH2 domains in approximately the same number of proteins. The basic structure of SH2 domains contains a central anti-parallel β -sheet surrounded by two α -helices, like a sandwich. They have a conserved pocket²⁰ which could specifically recognize *p*Tyr. Furthermore, researchers found that there is the so-called “specificity determining region” which recognizes several additional residues, three to five amino acids to the C-terminus of *p*Tyr and causes specificity-based recognition of different phosphorylated ligands.²⁰

In the cell, SH2 domains are usually a part of a protein possessing other domains such as the catalytic domain and other binding modules. SH2 domain-containing proteins participate in a wide variety of intracellular processes (SAP with signal regulation²¹, c-Cbl with ubiquitination²², SHP1 and SHP2 with dephosphorylation²³, Src and Fps with phosphorylation^{20, 24}). Mutations within the SH2 domain of several proteins have been implicated with defects in signaling pathways and cause diseases such as Noonan’s syndrome (SHP2) and the X-linked lymphoproliferative syndrome (SAP). For example, SHP2 (which contains a PTP domain and two SH2 domains) is a PTP with a proven oncogenic function. As

illustrated in Figure 1.3, SHP2 is normally endogenously expressed in mammalian cells in a low-activity state masked by the N-terminal SH2 domain which binds intramolecularly to the PTP domain. When cell membrane receptors are activated, two of the phosphorylated tyrosines within the SHP2-binding domain of these proteins interact with the two SH2 domains in SHP2. Subsequently the conformation of SHP2 is changed, which leads to complete restoration of the full enzymatic activity in SHP2. As a consequence, activated SHP2 dephosphorylates substrates which play important roles in various signal transduction processes. Mutation of SHP2, especially in the N-terminal SH2 domain, could cause hyperactivation of its catalytic activity (Figure 1.3c). Activated mutations in SHP2 are known to cause leukemia.

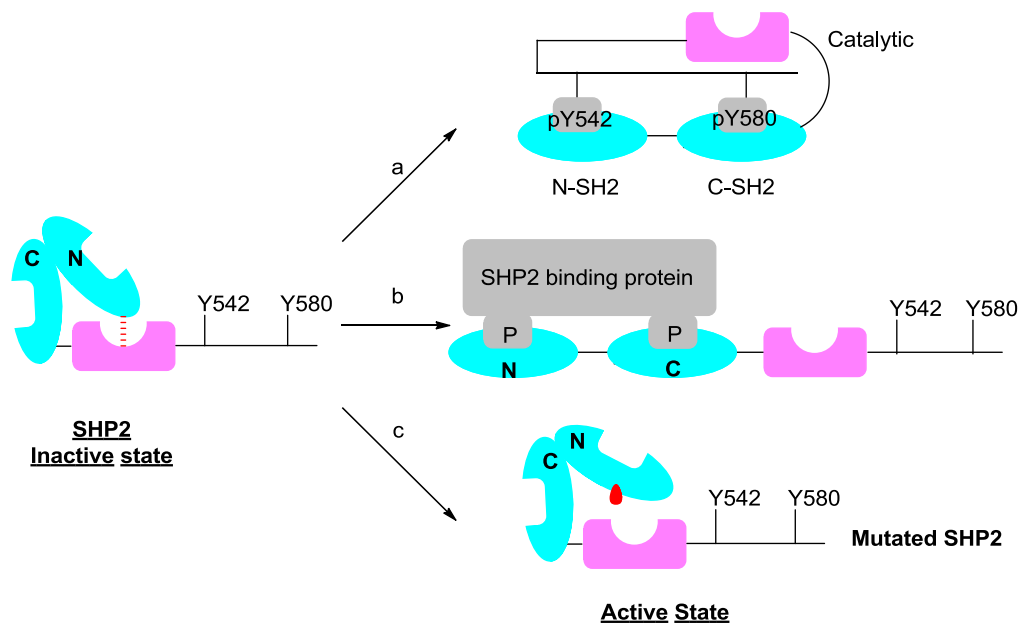


Figure 1.3 The effect of SH2 domains in SHP2 protein. In the inactive state, SHP2 is kept silence by interaction of the N-terminal SH2 domain with the PTP catalytic domain. a) Self-control releasing. In the C-terminal tail, both Y542 and Y580 are phosphorylated. Subsequently, the phosphorylated tail binds to two SH2 domain and it leads to the activation of SHP2; b) SHP2-binding protein. Activation of surface receptors and subsequent tyrosine phosphorylation of binding sites in receptor-tyrosine kinases or some scaffolding adapter

proteins for the two SH2 domains, leads to recruitment of SHP2; c) Mutation of N-terminal SH2 domain. In Noonan syndrome, mutations in SHP2 impair the inhibitory interaction of two domains, leading to decrease the binding affinity of two domains and increase activity of SHP2.

1.3 Mimics of *p*Tyr

As discussed above, the addition of a phosphate group to a protein or other organic molecules is important in many biological processes. Specifically, the *p*Tyr moiety serves as a critical affinity element as it is the key recognition and binding part for the phosphoproteomic network. Up to date, various chemical biology tools have been developed for the study of *p*Tyr-based proteins. One common strategy is the creation of novel types of *p*Tyr mimics. The potential mimics could be used in the rational design of inhibitors/binders, with greater pharmaceutical impacts, to achieve specific controls of intracellular pathways and to investigate the roles of targets. In addition, *p*Tyr mimics could possess both the N- and C-terminus similar to natural amino acids, thus enabling their easy incorporation into peptides and even proteins using native chemical ligation²⁶ or nonsense suppression technology²⁷. In the next sections, four types of *p*Tyr mimics, namely the fluorogenic *p*Tyr mimics, caged *p*Tyr, reversible non-hydrolyzable *p*Tyr mimics and irreversible *p*Tyr mimics, will be discussed. Subsequently, we will highlight the application of activity-based protein profiling (ABPP), a relatively new method and its successful contributions in the discovery of new probes for detecting PTPs.

1.3.1 Fluorogenic mimics of *p*Tyr

Fluorogenic compounds generally exhibit low or no fluorescence until enzymatic reaction occurs. The common fluorophores used in the design of

fluorogenic compounds include coumarin, fluorescein and rhodamines, which could form ethers, amides or chemically reduced dyes. Such derivatives have a very weak fluorescent intensity relative to their hydrolyzed/oxidized products, which are phenol, anilines or the oxidized fluorophores. Taking PTPs fluorogenic substrates as an example (Figure 1.4), the enzyme recognizing phosphate head group is conjugated with the phenol moiety of the fluorophore. This modification could reduce the availability of lone pairs in the fluorophore's π system. Upon PTPs' hydrolysis, the head is released from the fluorophore, leading to concomitant the emission of the fluorescence. The changes in the intensity of fluorescence at certain wavelength could be used to monitor the enzyme activity. Therefore, they are widely used in the detection of enzyme activity, high-throughput inhibitor screening and bioimaging.²⁸

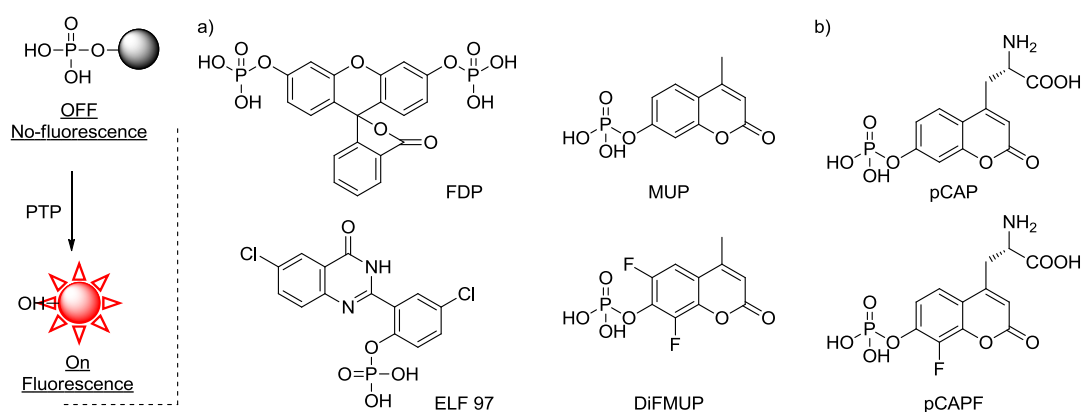


Figure 1.4 Fluorogenic *p*Tyr mimics. left: principle of PTP-mediated fluorogenic substrates; right: a) Some known PTP fluorogenic substrates; b) Some fluorogenic *p*Tyr mimics.

Many PTPs' fluorogenic substrates such as FDP, ELF 97, MUP²⁹, and DiFMUP³⁰ (shown in Figure 1.4) are developed based on the hydrolysis of the phosphate head of tyrosine residue by PTPs. In particular, DiFMUP ($pK_a = 4.9$, $\Phi = 0.89$), which is widely used in inhibition assay, has a low pK_a and higher fluorescent

quantum yield than that of MUP ($pK_a = 7.8$, $\Phi = 0.63$), thus making it a suitable substrate under acidic pH conditions. Additionally, it is a non-fluorescent compound and only generates the fluorescent DIFMU ($\lambda_{ex} = 358$ nm, $\lambda_{em} = 450$ nm) upon hydrolysis by PTPs. DIFMUP-based assay provides a general and sensitive solution assay for fast and accurate exploration of the binding modes between enzymes and inhibitors. However, DIFMUP does not work well on certain members of the PTP family, such as SHP2 ($K_m = 104$ μ M, $K_{cat} = 1.3$ s⁻¹). Barrios and co-workers reported a fluorogenic phosphocoumaryl amino propionic acid (pCAP), which is a coumarin based fluorogenic *p*Tyr mimic.³¹ Subsequently, the incorporation of pCAP into different peptide-based PTP substrates allows enzymatic detection of most PTPs with high specificity and good sensitivity

To overcome the pH sensitivity of simple coumarins, the same group designed a fluorinated analog of pCAP (pCAPF).³² Additionally, they replaced propionic acid in the parental compound with butanoic acid to have a better fit for the large catalytic pockets in PTPs. The pCAPF-based peptides are more sensitive upon hydrolysis when compared to the original version, but have smaller turnover numbers. These fluorogenic mimics have potential applications in large-scale profiling of PTP activities using a combinatorial peptide library and to provide the fingerprint of each PTP.

1.3.2 Caged *p*Tyr

In chemical biology terms, a “caged molecule” is a compound whose function is masked by a chemical protecting group.³³ The uncaging process triggered by chemical or physical means removes the protecting group and restores its intrinsic activity. Caged compounds allow the spatial and temporal control over the release of

reactive functions in target molecules, thus enabling the study of complex cellular processes in “real time” for the elucidation of their pathways more accurately. The photo-protecting group is often used as the cage as lights could be precisely and easily manipulated. Furthermore, with the development of lasers and microscopes, the uncaging process by light directed to a specific area, duration and intensity of irradiation can be controlled.^{33b, 33c} The straightforward design of caged *p*Tyr involves protecting one or two phosphate oxygens with photo-removable protecting groups. Upon removal of the caging group by photolysis, the masked phosphorylated tyrosine residue is exposed, and hence becomes susceptible to phosphatases and other domains.

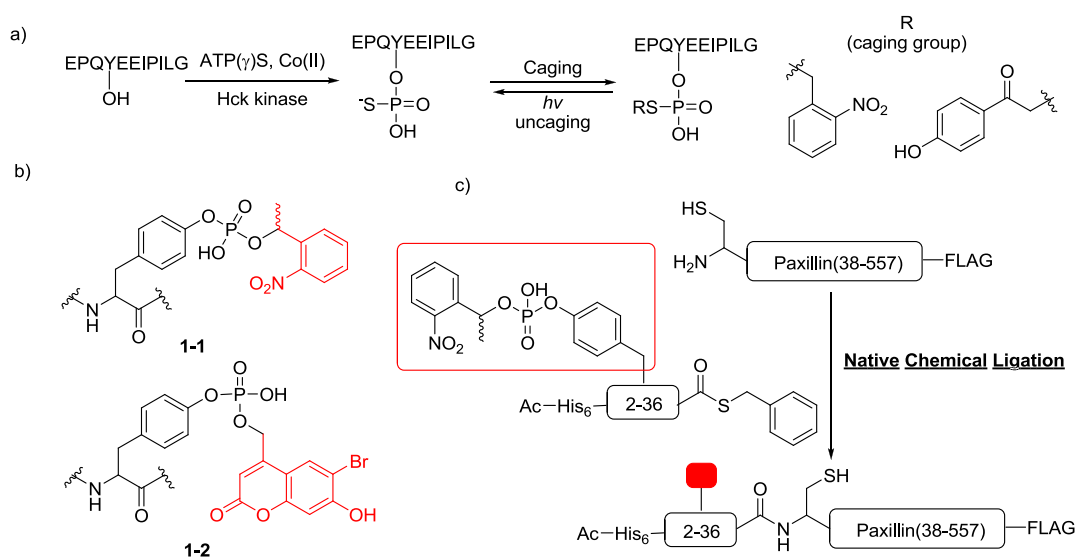


Figure 1.5 Caged *p*Tyr mimics. a) Caging and uncaging of thiophosphotyrosine peptides; b) Two different caging groups; c) Native chemical ligation joins fragments of Paxilin (38-557) with a paxillin thioester containing a caged *p*Tyr at position 31 to generate full length paxillin.

In one of the earliest examples using the caged system, Bayley and co-workers reported the synthesis of nitrobenzyl- and *p*-hydroxyphenacyl caged thiophosphotyrosine incorporated peptides (Figure 1.5a).³⁴ The peptides were first thiophosphorylated by their cognate tyrosine kinases under divalent transition metal

buffer. This is, followed by the reaction of the thiophosphate sulfur with electrophilic reagents (2-nitrobenzyl bromide or *p*-hydroxyphenacyl bromide) to generate new photocaged substrates (Figure 1.5a). The caged peptides showed no binding to SH2 domain, however, upon by UV irradiation, the initial affinity is restored.

Imperiali and co-workers demonstrated the successful synthesis of 1-(2-nitrophenyl)ethyl (NPE)-based caged *p*Tyr (**1-1**) building block for the use in Fmoc-based solid-phase peptide synthesis.³⁵ NPE caging group could be removed with far UV light which is less harmful to cells. Semi-synthesis (native chemical ligation) was then used to install the caged *p*Tyr residue into paxillin which could serve as a tool to investigate the influence of Tyr31-paxillin phosphorylation on cellular migration.^{36a} Similar attempts were made by the Becker group to incorporate the caged *p*Tyr in recombinant STAT6, and demonstrating that the movement and translocation of STAT6 is regulated by the phosphorylation of its Tyr641.^{36b} The Nagamune group has recently developed a new mimic (**1-2**) with a two-photon uncaging group (6-bromo-7-hydroxycoumarin-4-ylmethyl).³⁷ The two-photon uncaging approach is simple, economic and provides improved the spatial resolution.³⁸

1.3.3 Reversible non-hydrolyzable mimics

The phosphate group is tetrahedral in geometry and dianionic at physiological *p*H. The charged phosphate group is not suitable and has poor bioavailability due to (1) the low cell permeability and (2) the ease of hydrolysis of the phosphoryl group by endogenous phosphatases. Hence the development of phosphate mimics has been focused on the introduction of a non-hydrolyzable bond and the use of groups with fewer charges, yet maintaining the affinity to target proteins.

One of the most obvious approaches in the design of bioisosteres is to replace the benzylic oxygen atom of *p*Tyr with a methylene group, to achieve a hydrolytically stable phosphonomethyl-phenylalanine (Pmp) derivative.³⁹ The resulting phosphonate has a higher ionization constant ($pK_{a2} = 7.1$) compared to the normal phosphonate ($pK_{a2} = 5.7$), thereby resulting a different ionization state in neutral *p*H. However, the absence of the hydrogen bonding ability in the methylene group in turn causes Pmp to show less binding affinity.

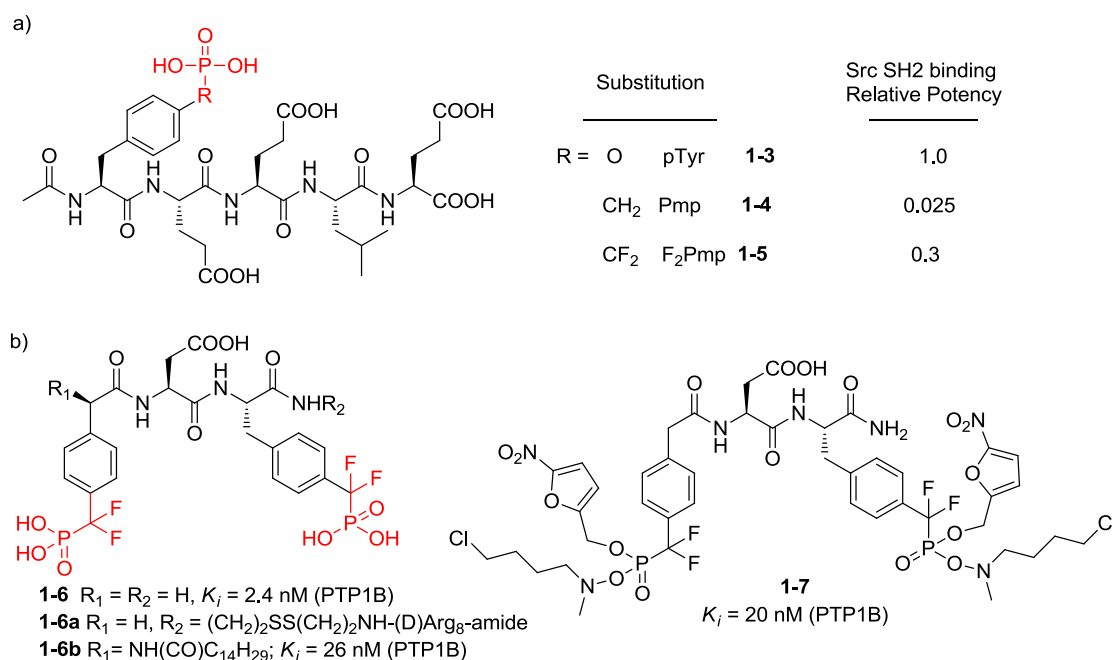


Figure 1.6 Pmp and F₂Pmp based inhibitors.

Subsequently, the Burke group in 1992 synthesized a new Pmp analog. They introduced two fluorine atoms at α -methylene position.⁴⁰ The fluoro substituents lower phosphonate pK_a values and potentially afford hydrogen-bonding interactions similar to the phosphoryl ester oxygen. This increases the binding affinity by several orders of magnitude. As shown in Figure 1.6a, phosphopeptide Ac-*p*Tyr-Glu-Glu-Ile-

Glu-OH is a known Src SH2 domain binder ($IC_{50} = 0.64 \pm 0.04 \mu\text{M}$). The investigation on several *p*Tyr replacements found that the F₂Pmp mimic (**1-5**) maintains the binding affinity (3-fold less potent compared to the parent peptide **1-3**), while Pmp based **1-4** exhibits 40-fold decrease binding potency.^{19a, 41} It is worth to noted that, to date, the Lawrence and Zhang groups reported the use of a combinatorial high-throughput screening method and furnished a F₂Pmp based small inhibitor **1-6** exhibiting notable selectivity in favor of PTP1B ($K_i = 2.4 \text{ nM}$) against a panel of twelve other PTPs including TCPTP.⁴² However, the bioavailability of **1-6** due to the charged phosphate group is still a limitation. To enhance the cellular activity, derivatives of **1-6** were developed, including coupling a cell permeable peptide tag (D-Arg₈, **1-6a**), incorporating a highly lipophilic fatty acid (**1-6b**) and synthesizing a prodrug (**1-7**). These modifications have some success to introduce phosphorus-based therapeutics, while, non-phosphorus-based mimics could also provide an approach to solve the problem.

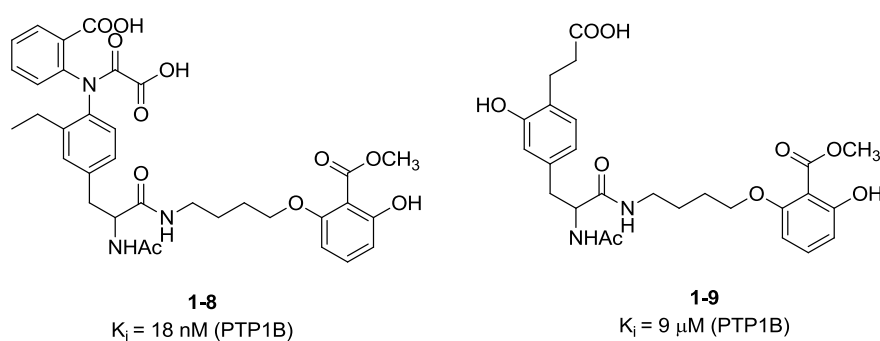


Figure 1.7 PTP1B inhibitors with carboxylic acid-based *p*Tyr bioisosteres.

Consequently, *p*Tyr mimics with combinations of polar groups including mono- and di- anionic groups were generated to maximally maintain phosphate-like electrostatic interactions and still obtain higher cell bioavailability. For example,

compound **1-8** was found to be a potent inhibitor of PTP1B and has two carboxylic acids as a phosphate bioisostere.⁴⁴ However, its low cell permeability limits its future therapeutic development. The same group further developed a singly charged compound **1-9** with reasonable inhibitory ($K_i = 9 \mu\text{M}$), good selectivity for PTP1B over most other phosphatases and more importantly, higher cell permeability.⁴⁵

The Burke group have spearheaded the synthesis of numerous mono and dicarboxylic acid-based *p*Tyr replacements as shown in Figure 1.8, including *O*-malonyltyrosine (OMT, **1-10**), fluoro-*O*-malonyltyrosine (FMT, **1-17**) and F₂Pmp (**1-12**).⁴⁶ Compounds (**1-14**, **1-16**) with mono carboxylic acid due to lower the binding affinity show less potency than dicarboxylic acid-based analogues (**1-10**, **1-15**, **1-17**). Compound **1-15** with the dicarboxylic acid modification (2.5-fold decreases) maintains similar hydrogen-bonding interactions with Grb2 SH2 domain compared to the corresponding Pmp derivative.

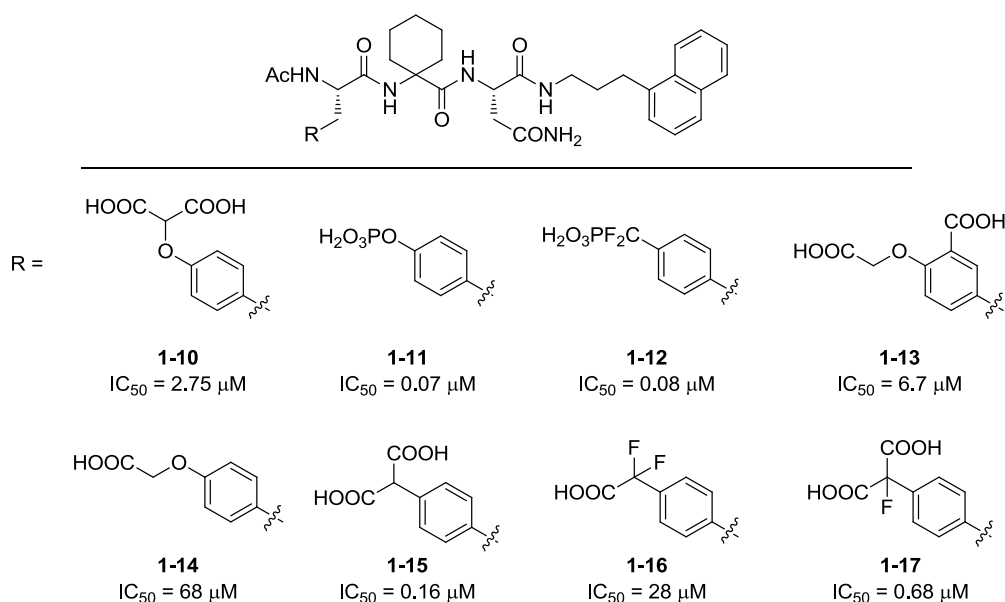


Figure 1.8 Different *p*Tyr mimics based binders of Grb2 SH2 domain.

Another study reported various modifications to a small molecular peptidomimetic PTP1B inhibitor with O-carboxymethyl salicylic acid moiety **1-18** by the Bleasdale group.⁴⁷ Due to the low cell permeability, one carboxylic acid moiety was replaced by a tetrazole unit as a carboxyl group. The resulting analogue **1-19** showed the comparable potency to the dicarboxylic acid analogue against PTP1B. Both compounds showed a similar selectivity over LAR and SHP-2. The X-ray crystallographic studies also confirmed that the tetrazole-containing **1-19** bound in a similar manner in PTP1B active site as the dicarboxylic acid analogue **1-18**. Furthermore, **1-19** showed the modest enhancement of insulin-stimulated 2-deoxyglucose uptake by L6 myocytes.

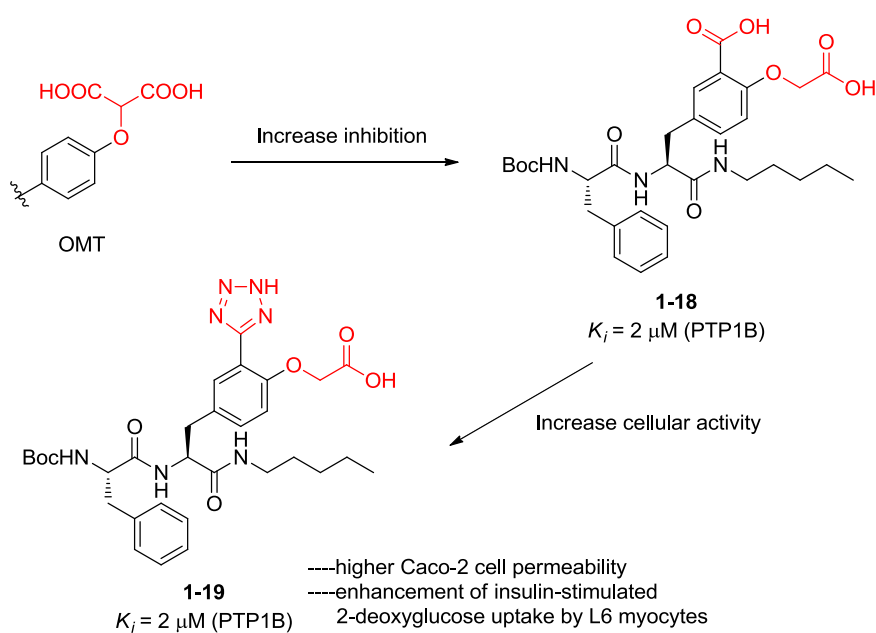


Figure 1.9 Improvement of PTP inhibition and cellular activity by discovering new *p*Tyr mimics.

In a more recent analysis, from the structure-activity relationship study, 1,2,5-thiadiazolidin-3-one-1,1-dioxide was identified as a novel phosphomimic.⁴⁸ The

isothiazolidinone group was used in the synthesis of PTP1B inhibitors **1-20** ($IC_{50} = 190 \text{ nM}$)^{48a} and **1-21** ($IC_{50} = 40 \text{ nM}$)^{48b}, demonstrating that the utility of this scaffold could serve as a powerful *p*Tyr mimic.

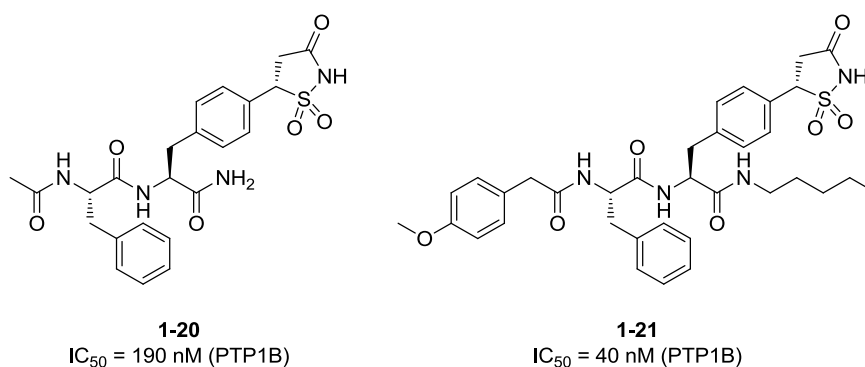


Figure 1.10 Inhibitors containing *p*Tyr bioisosteres.

1.3.4 Irreversible *p*Tyr mimics

The active site of the PTP family is highly conserved and contains a highly nucleophilic Cys residue as described above. Irreversible *p*Tyr mimics always have a reactive group that could be attacked by the Cys residue, therefore resulting in irreversible inhibitions. There are few examples that the mimics contain N- and C-terminal modifications. Other *p*Tyr mimics without the properties of amino acids have also been developed and widely used in activity-based probe profiling which will be discussed later (Chapter 1.4).

Zhang and colleagues have replaced the phosphate of *p*Tyr with a seleninic acid motif ($-\text{CH}_2\text{SeO}_2\text{H}$).⁴⁹ The RSeO_2^- is also a potential bioisosteric match for biological *O*-phosphate (ROPO_3^{2-}). This mimic reacts effectively and irreversibly with the PTPs by formation of a covalent selenosulfide adduct. It inactivates YopH as well as PTP1B, VHR and VHX. It gives the potential application for seleninate based

inhibitors to study the physiological functions of PTPs. The Yao group developed the quinone methide-based unnatural amino acid (Figure 1.11, the mechanism was described in Figure 1.14) containing a 2-fluoromethyl phenyl phosphate (**2-FMPP**) and incorporated it into several peptide-based substrates of PTPs. The peptide-based probes are able to profile the PTPs present in the mammalian cell lysates.⁵⁰

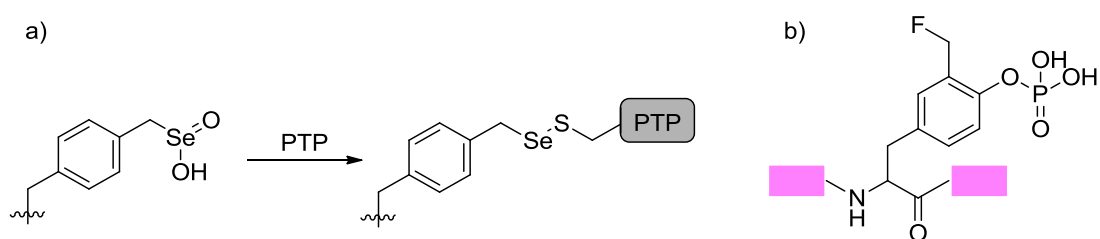


Figure 1.11 Irreversible *p*Tyr mimics. a) Seleninate based mimics b) Quinone methide based mimics.

1.4 Activity-based protein profiling (ABPP)

Activity based protein profiling, pioneered by the Cravatt group, is one of most efficient ways to better understand both *in vitro* and *in vivo* enzymatic activities of a class of enzymes from a crude proteome.⁵¹ The strategy uses active site-directed mechanism-based probes, which normally have three key elements: 1) an active site-targeting reactive group (warhead) that directs the probe to the active site and covalently reacts with the protein; 2) a reporter group, typically a fluorophore or biotin for visualization and identification of labeled proteins; 3) a linker which gives the space between reporter and warhead (Figure 1.12). To date, many ABPs have been developed against different classes of enzymes, including serine hydrolases⁵², cysteine proteases⁵³ and so on.

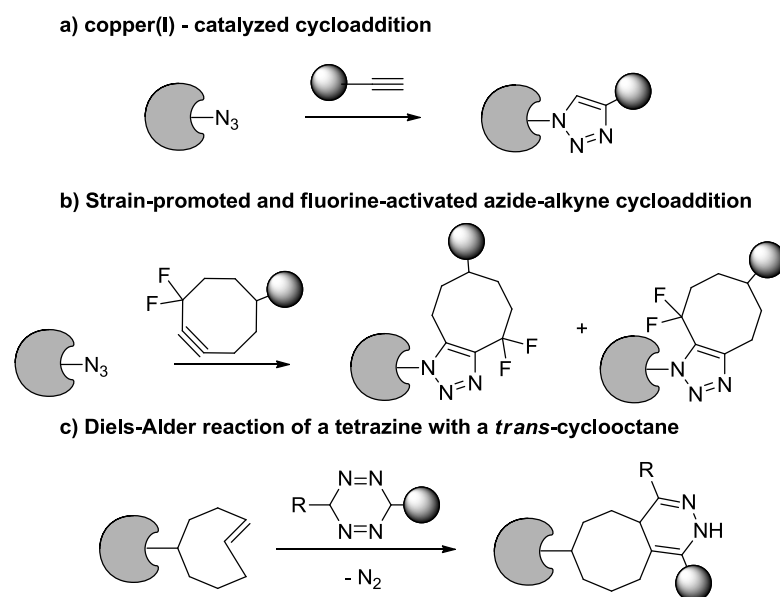


Figure 1.12 Overview of some developed bioorthogonal reactions used in the tag-free approach.

However, the bulky reporter group could affect the probe's cellular distribution and decrease its cell permeability. The groups of Cravatt and Overkleeft independently developed strategies which used copper(I)-catalyzed azide-alkyne [3+2] cycloaddition (CuAAC, click chemistry) to incorporate the reporter group only after the labeling of the enzyme has occurred so as to avoid the influence of the reporter in the enzyme-probe interaction.⁵⁴ This cycloaddition reaction is one of the several types of bioorthogonal ligation reactions (a term coined by C. R. Bertozzi in 2003^{57a}) that exist in the literature. These reactions normally have functional groups which could be inserted into a probe structure while minimizing interference with native biochemical processes and avoiding cross-reactivity with other endogenous biomolecules. Ideally, the reactions are biocompatible and not toxic. The Bertozzi group has developed mono and difluorinated cyclooctyne reagents possessing ring strains and electron-withdrawing fluorines to promote the rate of the cycloaddition reaction with an azide without the normally cytotoxic copper catalyst.⁵⁵ With these

new substituted cyclooctynes, the authors were able to successfully image the dynamic processes of protein glycosylation in animals.⁵⁶ New bioorthogonal ligation reactions⁵⁷, with faster rates and less toxicity, including those based on the Diels-Alder reaction between a tetrazine and a *trans*-cyclooctene are currently being developed as well as being applied in tag-free protein profiling experiments.⁵⁸

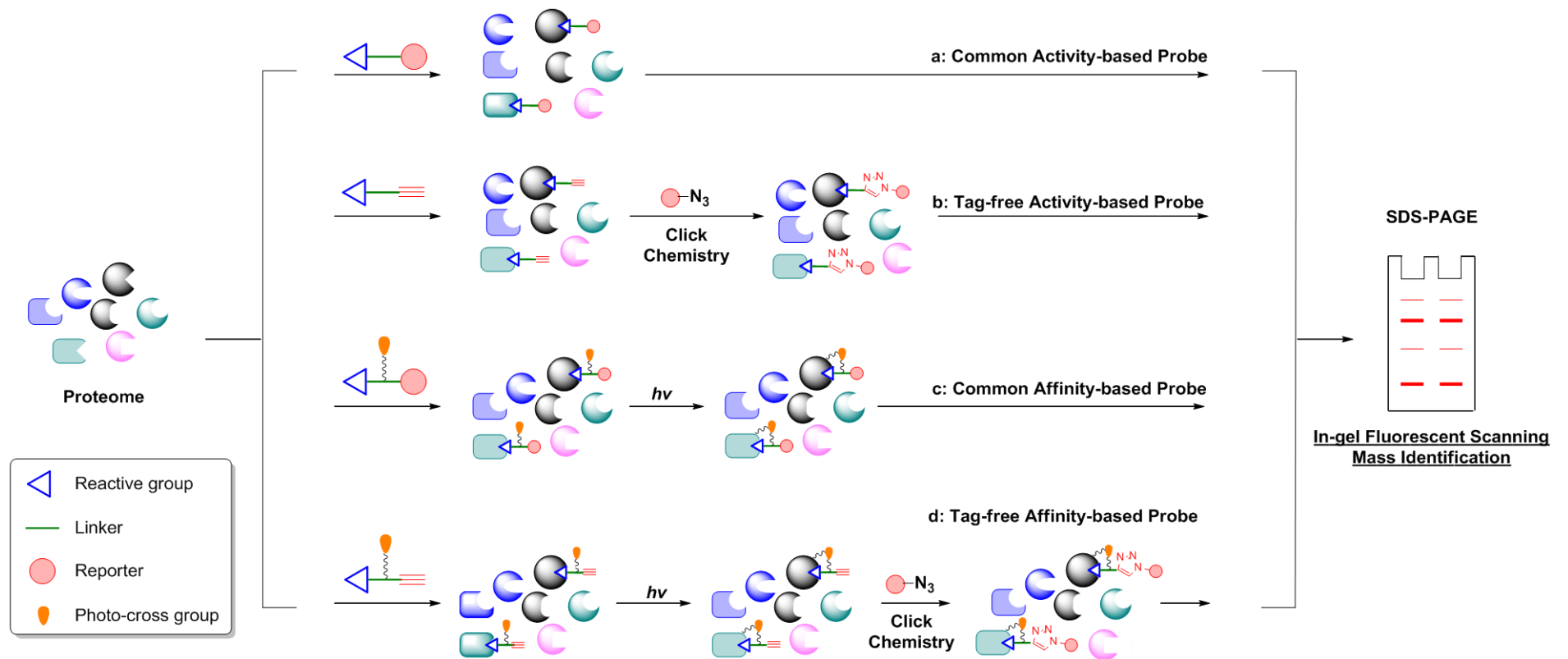


Figure 1.13 Proteome profiling with activity-based probes (ABPs) and affinity-based probes (A/BPs). a) Original activity-based probe, it contains warhead, linker and reporter, after covalent labeling, could be directly analyzed by SDS-PAGE. b) First generation of a), tag-free activity-based probe, it replaces the big report group with a small alkyne. After covalent labeling, click chemistry subsequently is carried on adding the reporter group. c) A/BPs, another branch of ABPs, it contains warhead, reporter and photo-cross group. After binding with targets, probes are then covalently bond with targets under UV-irradiation. d) Tag-free version of A/BPs.

Furthermore, instead of using known reactive groups-based scaffolds for the development of probes against proteins, an alternative approach, so-called affinity-based probes (AfBP, Figure 1.13) based on reversible inhibitors and binders was generated.⁵⁹ The key components of AfBPs include an inhibitor/binder as a core group, which provides tight and specific binding to the target proteins and a photo-cross unit⁶⁰, including benzophenone, aryl azide and diazirine group, which, upon UV irradiation, provides a covalent linkage between the probe and the target. This approach is potentially more appealing and much more powerful since there are various inhibitors and binders of interested proteins readily available.⁶¹ Furthermore, this approach could be amenable to the study of the cellular off-targets of many existing reversible drugs/drug candidates. The Yao group reported AfBP-like modification of Dasatinib, a dual Src/Abl inhibitor, which is capable of proteome-wide profiling of potential cellular targets of Dasatinib.^{61g}

1.4.1 Activity based probes for PTPs

In recent years, there has been a growing trend to develop activity based PTP probes. Lo *et al.* first explored small molecules containing a **2-FMPP** moiety which serves as a phosphate mimic.⁶² Upon hydrolysis by PTPs, the probe generates a highly reactive quinone methide intermediate which subsequently alkylates the nucleophiles present in the active site of PTP (Figure 1.14). Furthermore, they found that six cysteine residues of PTP1B could be labeled by the probes thereby suggesting a lack of specificity of the reactive intermediate. The Yao group further developed **1-22** and **1-23** and demonstrated that these probes specifically label phosphatases (both PTP and alkaline phosphatase) over other enzymes like proteases and lipases in an activity-based manner.^{63b} Although quinone methide based probes have been successfully used to profile PTPs, the diffusible and highly reactive nature of the intermediates

lead to poor PTP specificity and cross-reactivity in the crude proteome, making these probes unsuitable for ABPP application.⁶³ The issue has been partially solved by Li and colleagues who reported a probe **1-24** with a short peptide based on EGF receptor (a known substrate of PTP1B.).^{63d} As only PTP1B was tested in this literature, it remains unclear whether other PTPs could also be targeted. The Yao group synthesized a new quinone methide based unnatural amino acid and incorporated it into peptide-based probes which were chosen from the known substrates of PTPs.⁵⁰ This design is better than **1-24** since the generation of quinone methide is directly based on PTPs' activity, and not from other sources. From the result, the conjugation of the additional PTP substrate sequences improves the probe's binding affinity as well as specificity and decreases the possibility of diffusion and cross-reactivity.

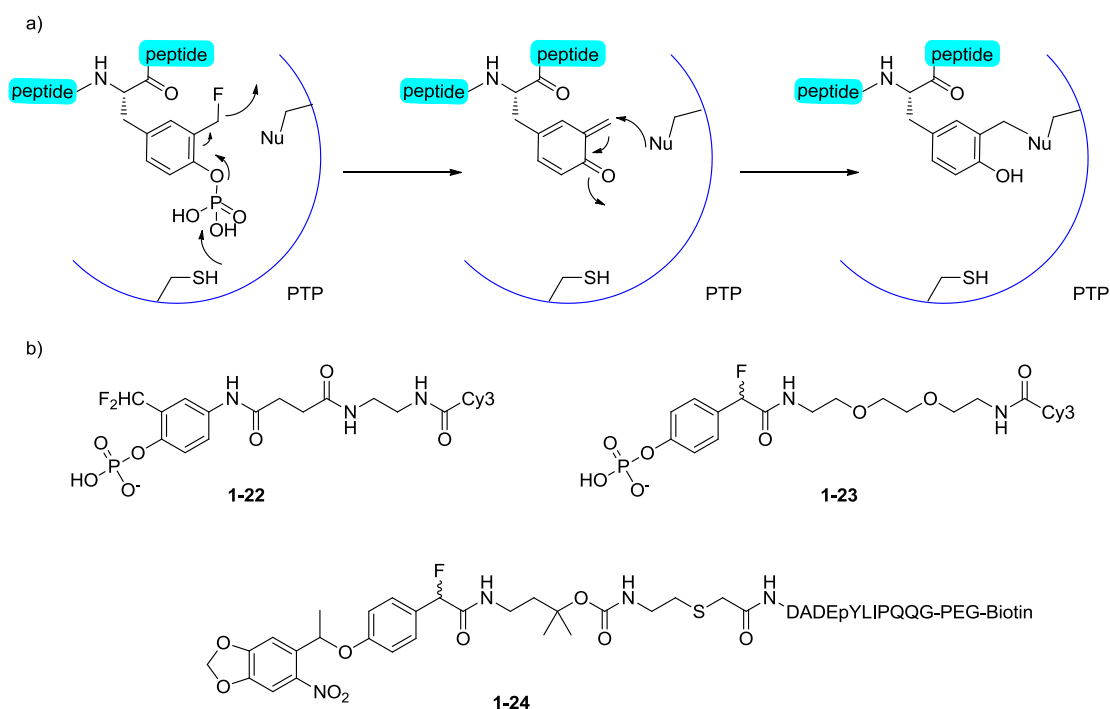


Figure 1.14 Quinone methide-based approaches. a) Schematic representation of formation of quinone methide intermediate and subsequent enzyme labeling.; b) Quinone methide-based probes.

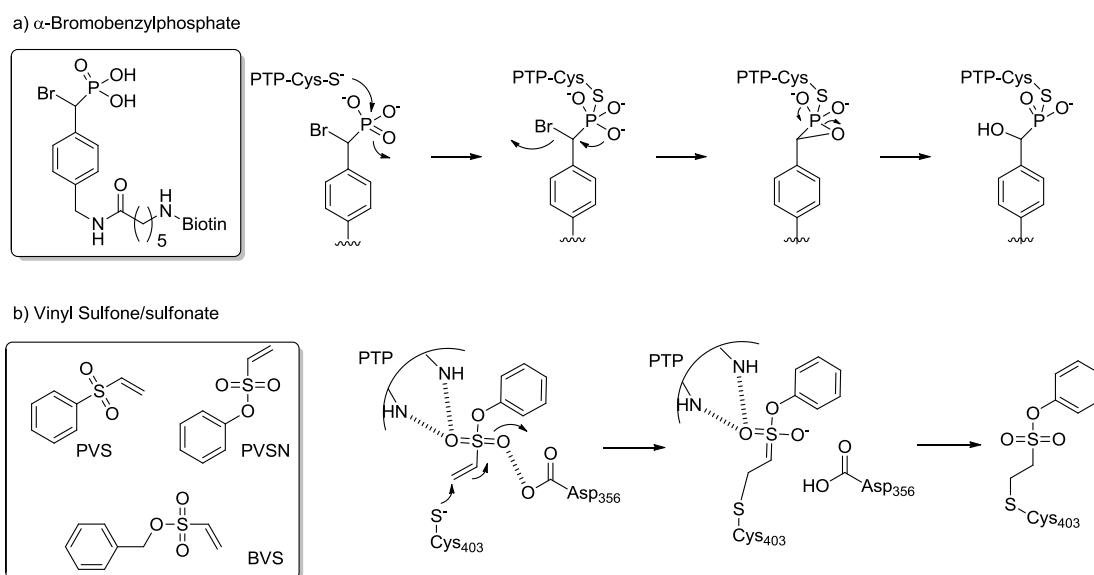


Figure 1.15 Bromobenzylphosphonate and vinyl sulfone/sulfonate based probes. a) The structure of a biotinylated bromobenzylphosphonate probe and its proposal mechanism.; b) The structure of three vinyl sulfonate/sulfone based probes and the proposal mechanism.

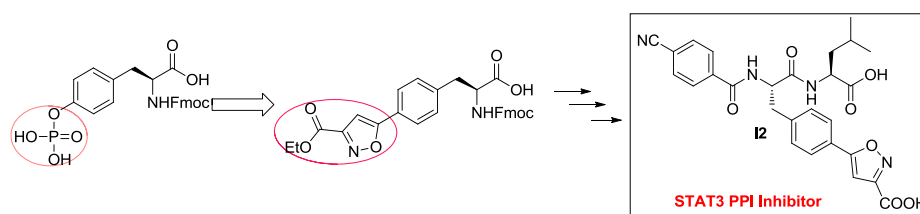
Zhang et al. developed a probe with α -bromobenzylphosphate as a non-hydrolyzable phosphate mimic (Figure 1.15a).⁶⁴ This mimic is sufficiently electrophilic to react covalently with the Cys residue in the PTP active site (one possible mechanism was shown in Figure 1.15a). The probe was shown to target PTPs in a crude proteome and more specific than the quinone methide-based probes. However, the highly unstable nature of the probes renders it impractical for widespread applications in ABPP. Recently, the same group developed a new version of probes based on phenyl vinyl sulfone/sulfonate motif (Figure 1.15b).⁶⁵ Through the Michael 1,4-addition, cysteine with low pK_a in active site nucleophilic attacks on the vinylic position, resulting in the formation of a stable thioether bond. In the design, the phenyl ring also plays an important role in recognition. However, vinyl sulfone/sulfonate based probes are well known to targeting cysteine-utilizing proteins, such as cysteine proteases. The cross-reactivity problem still needs to be addressed.

1.5 Objectives

The design of new *p*Tyr mimics and investigation of PTPs in cellular process is still a challenge. In this thesis, the main aim is to develop new *p*Tyr mimics and to apply them in the development of new inhibitors or activity-based probes. In addition, a new high throughput reaction is carried out to assemble PTPs probes for the profiling of PTPs in cellular system.

Chapter 2 An unnatural amino acid that mimics phosphotyrosine

Summary



This chapter, we describe the successful design and chemical synthesis of novel unnatural amino acid which mimics phosphotyrosine (*p*Tyr). By replacing the phosphate group in *p*Tyr with an isoxazole carboxylic acid, a novel unnatural amino acid has been successfully synthesized. Subsequently, its incorporation into a known PPI (protein-protein interaction) inhibitor of STAT3 protein (**ISS 610**) generated **I2** which showed reasonable anti-STAT3 activity in both fluorescence polarization and cell-proliferation experiments.

2.1 Introduction

Protein-protein interaction (PPI) mediates a large number of important regulatory pathways in cellular processes.⁶⁶ Inappropriate interactions are known to cause many human diseases including cancer and obesity. Small molecule inhibitors capable of disrupting protein-protein interaction therefore offer unique opportunities in drug discovery, especially against less common therapeutic targets such as transcription factors.⁶⁷ PPI inhibitors of STAT3 are such examples.⁶⁸ The signal transducer and

activator of transcription (STAT) family of proteins, including STAT3, are mediators of cytokine and growth factor responses, and control cellular growth and differentiation, survival, development and inflammation.⁶⁹ They are members of the Src homology 2 (SH2) domain-containing proteins, which also include many other adaptor proteins, phosphatases (i.e. SHP1 and SHP2) and kinases (i.e. Src and Abl).⁷⁰ SH2 domains are phosphotyrosine (*p*Tyr)-binding protein domains made of ~ 100 amino acids.^{69,71} In recent years, they have become attractive targets for developing new PPI inhibitors and potential therapeutic drugs.^{19, 72} In the case of STAT3 inhibitors, it is known that the activation of STAT3 is initiated by the phosphorylation of a key tyrosine residue near its C-terminus, which triggers protein dimerization via SH2-*p*Tyr interaction, resulting in subsequent nuclear translocation, binding to specific DNA-response elements and inducing gene transcription.⁷³ Although there are several proteins in the STAT family, so far only STAT3 is shown to be expressed constitutively in a variety of cancer cells, making it a good target for cancer therapy and the corresponding PPI inhibitors potential anti-tumor agents.^{68,74}

Most SH2 domain-targeting inhibitors are *p*Tyr containing peptides or peptidomimetics, in which the phosphotyrosyl group is the most critical element for molecular recognition/tight binding to the SH2 domain. However, its hydrolytic instability (mainly due to enzymatic hydrolysis by cellular protein tyrosine phosphatases or PTPs) and predominantly anionic nature at physiologic *p*H practically render the corresponding *p*Tyr containing inhibitors to possess poor cell permeability and bioavailability. Numerous *p*Tyr bioisosteres have been developed (Figure 2.1a),^{19c, 72b,75} most of which, however, are either still highly charged (**2-2a-c**), or poor mimics ((**2-3**) - (**2-5**); therefore possess low SH2-binding affinity). Herein, we report the design and chemical synthesis of a novel unnatural amino acid, **2-6** (Figure 2.1a), in which the

phosphate group in *p*Tyr was replaced with a non-hydrolyzable isoxazole carboxylic acid. The isoxazole carboxylic acid was previously shown to be a good cell-permeable bioisostere of the phosphate, and has been incorporated into small molecule inhibitors of PTPs.^{76, 77} Our finding here represents the first example in which this moiety has been successfully introduced as the side-chain of an amino acid. We further show this unnatural amino acid could be readily incorporated into SH2 domain inhibitors, i.e. **ISS610** (a previously reported STAT3 PPI inhibitor^{68a,68c}), to give a new inhibitor **I2** which possesses more desirable pharmacological properties (cell permeability and biostability) but still retains most of its inhibitory activity against STAT3 (Figure 2.1b).

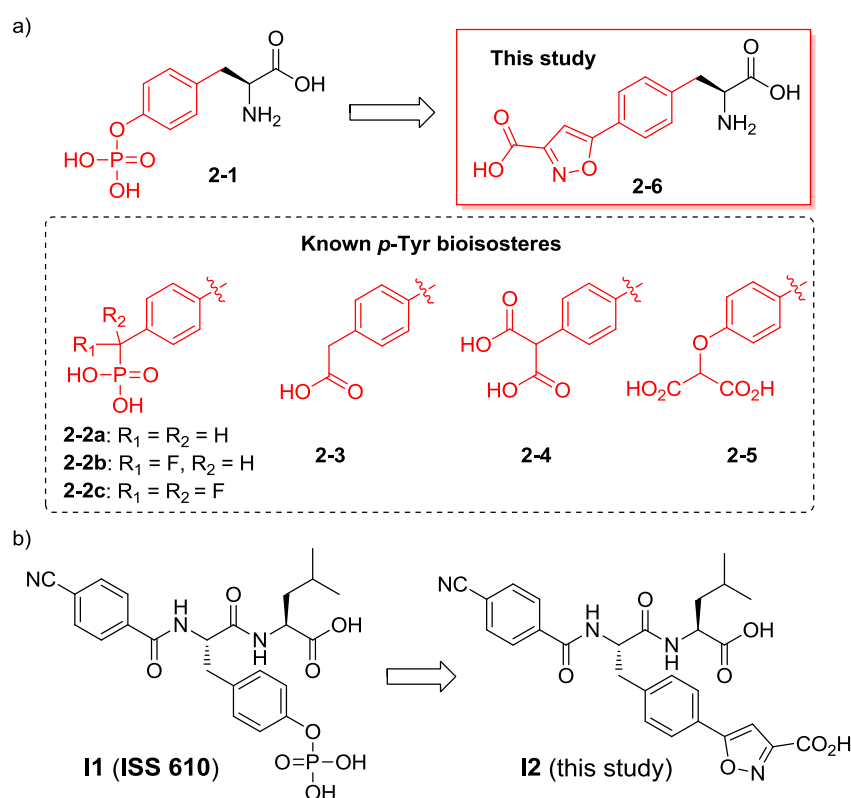
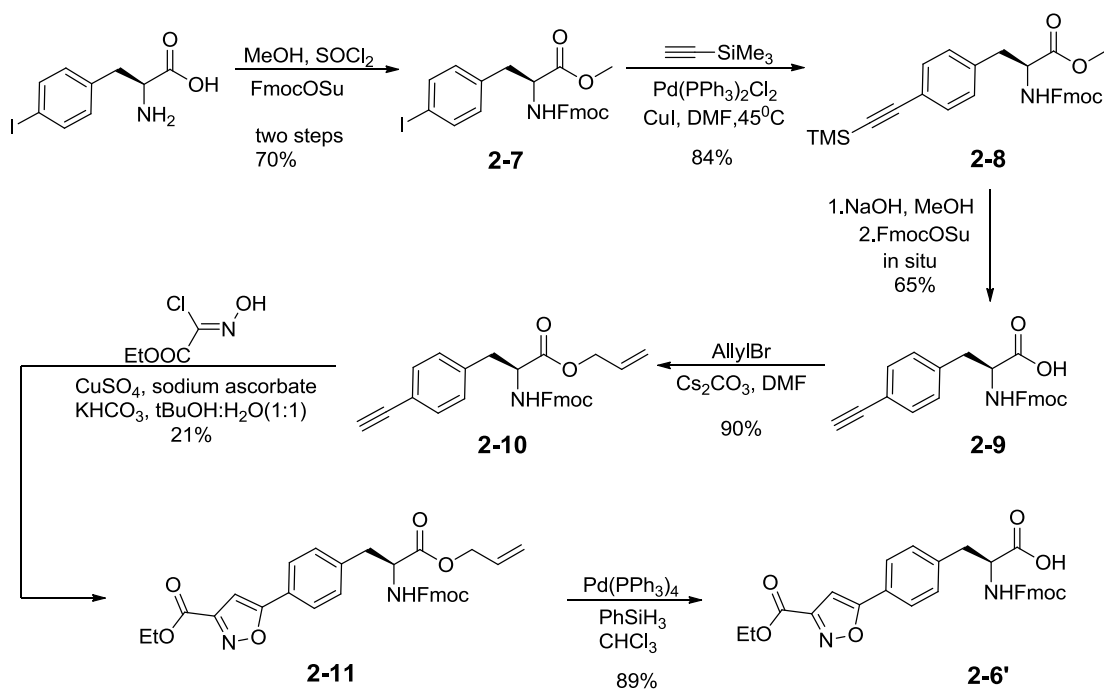


Figure 2.1 Structures of *p*Tyr, its bioisosteres and the one in this study. (a) Structures of *p*Tyr (**2-1**), representative examples of previously reported *p*Tyr bioisosteres (**2-2** to **2-5**), and the one developed from this study (**2-6**). (b) Structures of **ISS 610** (a previously reported STAT3 inhibitor^{68a,68c}), and **I2** (an improved version with incorporation of **2-6**).

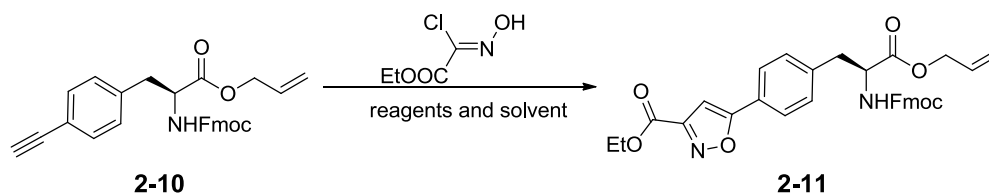
2.2 Synthesis of the unnatural amino acid

The Fmoc-protected form of the unnatural amino acid, **2-6'**, was synthesized in several steps from the commercially available 4-Iodophenylalanine⁷⁸, as shown in Scheme 2.1. Fmoc was chosen as the protecting group for the α -amino group in **2-6'** so that it's compatible with standard Fmoc-based solid-phase peptide synthesis. The isoxazole carboxylic acid side-chain was protected in the form of an ethyl ester as it is orthogonal to Fmoc chemistry and could be deprotected by LiOH.⁷⁷ Briefly, the Fmoc-protected methyl ester of 4-Iodophenylalanine, **2-7**, was subjected to Sonogashira coupling conditions with Pd(PPh₃)₂Cl₂/CuI in DMF, in the presence of TMS-protected acetylene, to give **2-8** in 84% yield. All three protecting groups were then removed by NaOH treatment, followed by Fmoc protection of the *N*- α -amino group to give **2-9** (65% yield in two steps). Subsequent protection of **2-9** in the form of an allyl ester gave **2-10** (90% yield). Next, the key isoxazole ring in **2-11** was formed using a recently published procedure,^{79, 80} under the Cu(I)-catalyzed 1,3-dipolar cycloaddition conditions between the terminal alkyne in **2-10** and ethyl chlorooximidoacetate. Previously, it was found electron-deficient nitrile oxides in general react sluggishly.⁷⁹ Attempts were therefore made to optimize the reaction conditions (Table 2.1); the best yield obtained was with CuSO₄/sodium ascorbate in KHCO₃ with (1:1) *t*BuOH/H₂O cosolvent, giving **2-11** in 21% and 23% yield in 1 d and 5 d, respectively. Finally, deprotection of the allyl group by Pd(PPh₃)₄/PhSiH₃⁸¹ afforded the Fmoc-protected unnatural amino acid precursor **2-6'**. No sign of racemization was observed throughout the whole synthesis and in the final unnatural amino acid **2-6'**.



Scheme 2.1 Synthesis of **2-6'** - the Fmoc-protected form of **2-6**.

Table 2.1 Optimization of isoxazole ring formation under Cu(I)-catalyzed, 1,3-dipolar cycloaddition conditions.

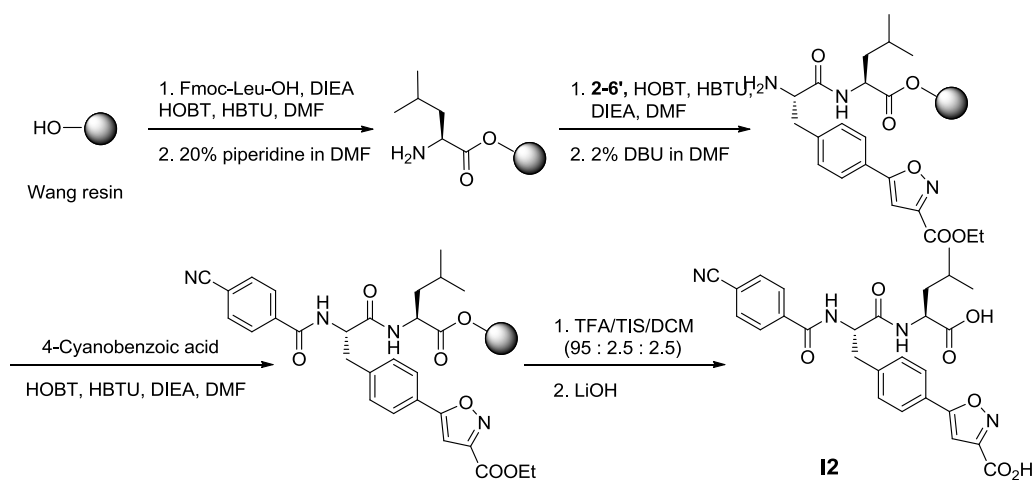


entry	Reagents (A or B) ^a	solvent	time (day)	yield (%)
1	A	<i>t</i> BuOH:H ₂ O (1:1)	1	21
2	A	<i>t</i> BuOH:H ₂ O (1:1)	5	23
3	A	DCM:H ₂ O (1:1)	1	20
4	A	THF:H ₂ O (1:1)	1	19
5	B	THF	1	15

^a A: CuSO₄/Sodium ascorbate/KHCO₃ (0.05/0.1/4 eq.); B: triethylamine (2 eq.)

2.3 Solid-phase synthesis of inhibitors

Next, the two STAT3 PPI inhibitors, **I1** and **I2** (Figure 2.1), were synthesized. **I1**, or **ISS 610**, is a peptidomimetic inhibitor derived from the STAT3 SH2 domain-binding *p*Tyr (Y*) peptide PY*LKTK.^{68a} **ISS 610** inhibits STAT3:STAT3 dimer (IC₅₀ = 42 μM from a DNA-binding assay^{68a}) and was shown to induce cell growth inhibition and apoptosis of human cancer cells that harbor constitutively active STAT3, including MDA-MB-435 and MDA-MB-231 (human breast), NIH3T3 (fibroblast) and T47D (human breast) cells.^{68c} **ISS 610**, however, has limited cell permeability and hydrolytic stability due to its *p*Tyr group. We reasoned that replacing *p*Tyr in **ISS 610** with our newly developed unnatural amino acid **2-6** (giving **I2**) should improve both aspects, making **I2** a potentially better STAT3 inhibitor for therapeutic intervention. Both inhibitors were synthesized conveniently using Fmoc-based solid-phase peptide synthesis protocols (see Scheme 2.2 for synthesis of **I2**). Upon resin cleavage, the ethyl group was removed by LiOH treatment, delivering the final inhibitor **I2** which was further purified by preparative HPLC to homogeneity (as judged by ¹H NMR) before biological screenings.



Scheme 2.2 Solid-phase synthesis of **I2**.

2.4 Biological experiments and discussion

2.4.1 Docking Stimulations

At first, we did docking stimulations to gain insight view of the binding similarity between two inhibitors. Docking was carried out on an SGI IRIX 6.5 workstation using the SYBYL7.2 suite installed with the FlexX docking software. The mol2 format of the two inhibitors **I1** and **I2** was prepared by Sybyl v7.2 (Tripos, Missouri, USA). STAT3 were obtained from the Protein Data Bank, entry 1BG1⁸². The docking sphere was set at 10 Å and centered on the Arg 609 residue as binding site. The docking was performed for 100 interactions and Figure 2.2 shows the most preferred conformations. Both phosphate of **I1** and isoxazole carboxylic acid of **I2** interact with Arg595 of STAT3 SH2 domain.

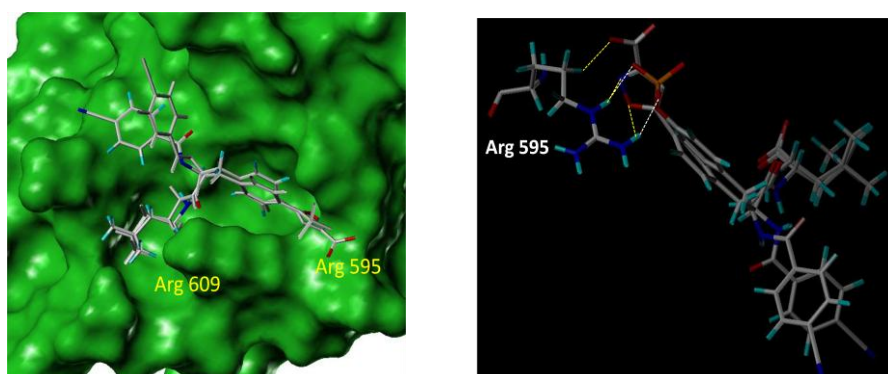


Figure 2.2 Molecular docking of **I1** and **I2**. Right shown the hydrogen bond between Arg595 and phosphate (**I1**, white dash bond) , isoxazole carboxylic acid (**I2**, yellow dash).

2.4.2 Fluorescence polarization (FP) experiments

2.4.2.1 Expression and purification of STAT3 SH2 domain

The clone containing the mammalian STAT3 SH2 domain was purchased from Open Biosystems (USA). The construct contains the SH2 domain cloned into a

modified pET28 bacterial expression vector. A 6x His tag is encoded for expression at the N-terminus which gives a final 68 kD fusion protein. The protein was expressed and purified according to protocols provided by the vendor.⁸³ Figure 2.3 shows the coomassie gel of fractions containing the purified protein (shown as a band at the expected 68 kD position). Only sufficiently pure fractions were collected and used for the following FP experiments.

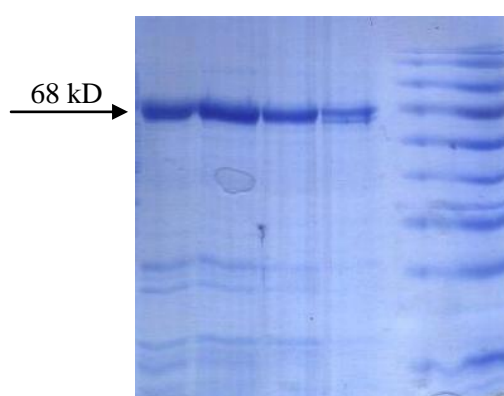


Figure 2.3 Representative coomassie gel of the purified STAT3 SH2 domain (From left to right, different elution fractions 1-4 and MW ladder). Only sufficiently pure fractions were pooled for subsequent FP experiments.

2.4.2.2 Fluorescence polarization experiments

Fluorescence polarization (FP) experiments were carried out following the published procedures.⁸⁴ Different concentrations of the inhibitors **I1** and **I2** (1 mM - 7.8 μ M) were first incubated with the protein (150 nM) for 30 min. Subsequently, the fluorescently labeled STAT3 peptide (Fluorescein-GY*LPQTV-NH₂) was then added into each well (final conc: 10 nM). The fluorescence readings were then taken with a fluorescence plate reader (Tecan, USA) installed with 2 pairs of polarizers (λ_{ex} : 485 \pm 20 nm, λ_{em} : 520 \pm 10 nm). From the data, each IC₅₀ plot was generated by averaging

the duplicates obtained from two independent assays. As shown in Figure 2.4, **I2** ($IC_{50} = 144 \pm 36 \mu\text{M}$) was able to retain most of the binding affinity of **I1** ($IC_{50} = 83 \pm 21 \mu\text{M}$) to STAT3,⁸⁵ indicating our unnatural amino acid indeed serves as a good mimic of *p*Tyr. Although **I2** (similar to **I1**) is a rather poor inhibitor, future improvement may be made by modifications of other proteins in **I1**, either rationally or combinatorially.⁶⁸

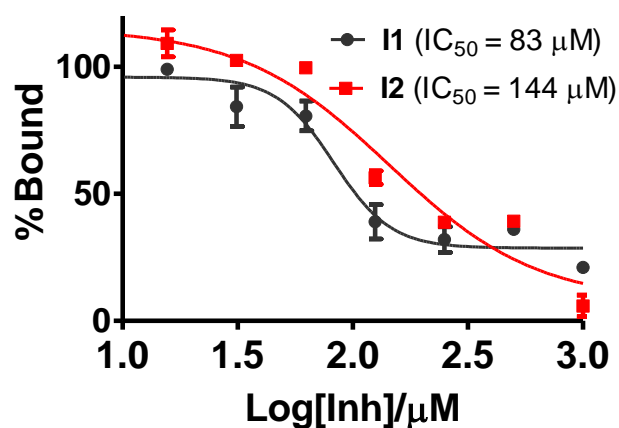


Figure 2.4 IC_{50} of two inhibitors determined by fluorescence polarization experiments

2.4.3 Inhibition assays of PTP1B

Next, we carried on the inhibition assay of PTP1B by **I2** to test the specificity of the inhibitor. The inhibition of **I2** to PTP1B was assayed by measuring the rate of hydrolysis of the fluorogenic substrate, 6,8-difluoromethylumbellifery phosphate (DIFMUP, Invitrogen, USA) in 25 μL reaction volumes in black polypropylene flat-bottom 384-well microtiter plates (Greiner, Germany).⁷⁷ Briefly, a two-fold dilution series of **I2**, from approximately 2 mM to 125 μM (final concentrations) was prepared. PTP1B's concentration is the same as STAT3 SH2 domain, 150 nM. The substrate concentration is 25 μM . As shown in Figure 2.5, little or no inhibition of

PTP1B by up to 2 mM of **I2** was observed, indicating **I2** is likely a specific inhibitor (binder) of the STAT3 SH2 domain.

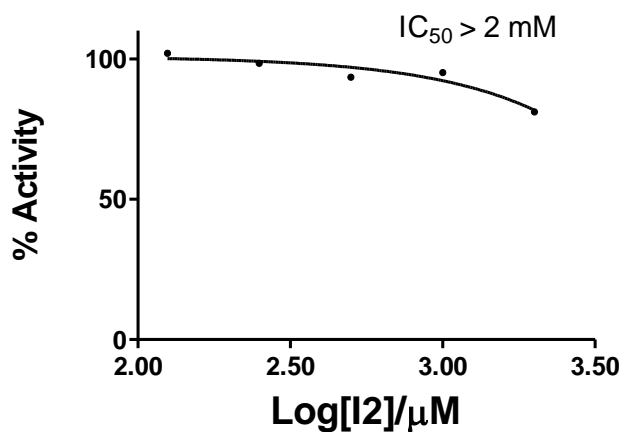


Figure 2.5 IC₅₀ graphs of **I2** against PTP1B

2.4.4 Hydrolytic stability test

Since **I1**'s phosphate head is known to be hydrolysis by PTPs, we tested the stability of both compounds toward cellular PTPs. First, T47D cell lysate was prepared in PBS, pH 7.5 by disrupting their cell membranes using needle and syringe. 1 µL of 5 mM **I1** and **I2** in DMSO was added into 100 µL 2 mg/mL cell lysate, respectively. After 15 h, the lysate was pipetted into Ultracel YM-3 (Millipore) and centrifuge (13000 rpm, 50 min) to get the filtrate and remove most of proteins (larger than 3 kD). Then, the composition of filtrate was analyzed by LCMS (Figure 2.6). From LCMS, after 15 h incubation, **I2** was quite stable and still not hydrolyzed by cellular proteases or others. However **I1** has been almost completely dephosphorylated.

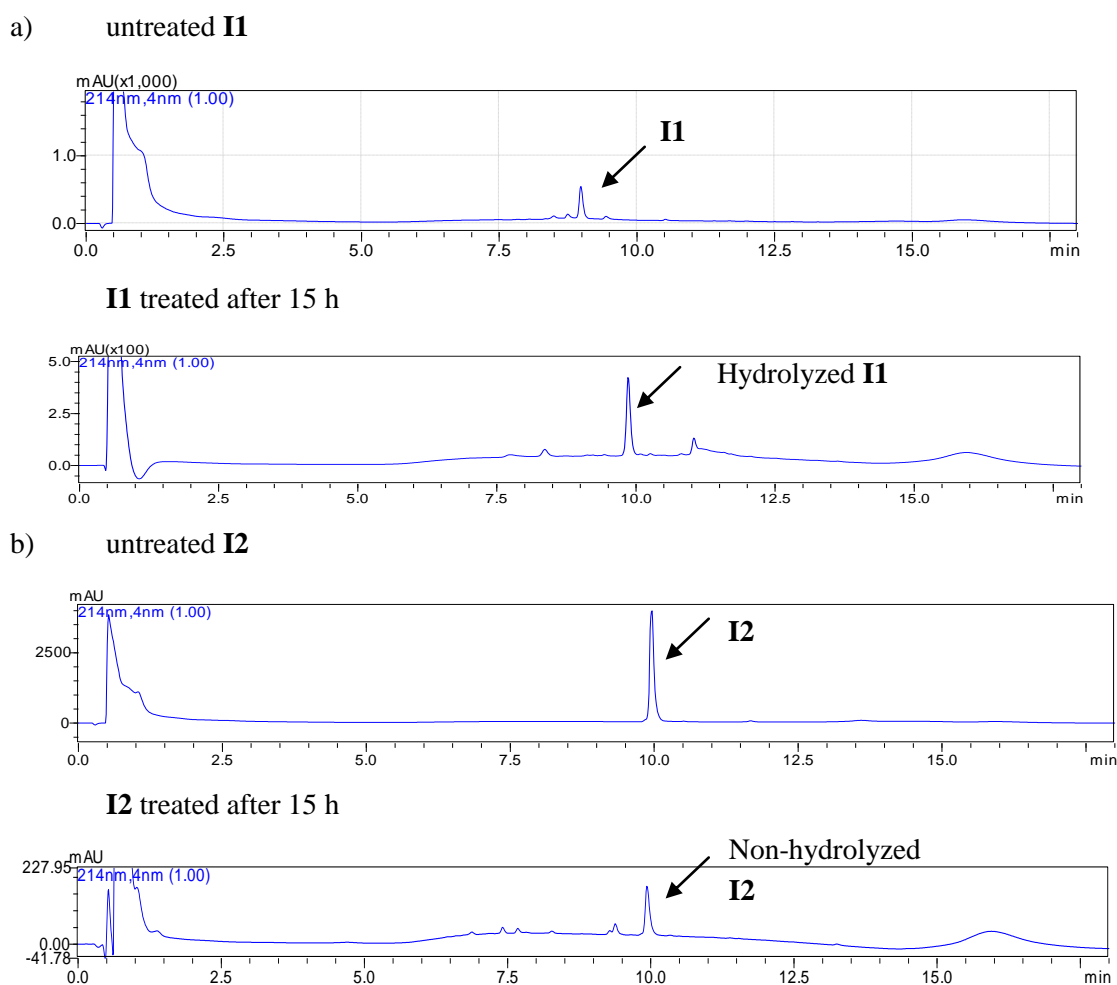


Figure 2.6 LC profiling of hydrolytic stability test. a) The LC profiling of treated **I1**. b) The LC profiling of treated **I2**.

2.4.5 Cell permeability test

As phosphate head has another limitation (low cell permeability) and we modified the head to have one carboxylic acid, cell permeability test was continued to test whether the modification could improve the cell permeability. MDCK (Madin-Darby canine kidney) cells⁸⁶ were used for testing the cell permeability. MDCK was seeded with 600,000~700,000 cells per cm^2 (0.11 cm^2 per well insert, Millipore® #PSHT004R1) and cultured for 3 days before test. When cultures are confluent, remove media and rinse inserts with Hanks' Balanced Salt Solution (HBSS, Gibco®

#14025). Transport assay donor solutions consisted of 50 μM **I1**, **I2** and caffeine in transport medium containing 60 μM Lucifer Yellow (LY) and 1% DMSO. Transport assays were conducted using 75 μL of apical (AP) donor solution and 235 μL of basolateral (BL) acceptor solution (transport medium, pH 7.4, following manufacture descriptions of Millipore® #PSHT004R1). Monolayer was incubated with donor and acceptor solutions for 60 min at 37 °C, 95% humidity. After 60 min, the donor and acceptor solutions of each compound are collected and quantified by HPLC. Lucifer Yellow (LY) was quantified using a fluorescence 96-well plate reader (BioTek® Synergy 4 fluorescence plate reader at Ex = 485 nm and Em = 539 nm). P_{app} (apparent permeability) values were calculated according to the following equation:

$$P_{app} = \left(\frac{dQ}{dt} \right) \times \frac{1}{C_0} \times \frac{1}{A}$$

where dQ/dt is the permeability rate, C_0 is the initial concentration in the donor compartment, and A is the surface area of the filter. Permeability rates were calculated by plotting the percent of initial AP drug mass (peak area) found in the BL compartment versus time and determining the slope of the line. Lucifer yellow (LY) results were used as an internal control for each monolayer to verify tight junction integrity during the entire assay period. Accordingly, LY P_{app} values were quantified from 60 min basolateral samples after background subtraction. As a quality control, results from MDCK monolayer with LY $P_{app} > 30$ nm/s were not used. From Table 2.2, because **I1** and the hydrolytic product have different cell permeability, the data of **I1** is not available. **I2** is shown quite reasonable cell permeability.

Table 2.2 The results of cell permeability.

	Positive control	Negative control	I1	I2
--	------------------	------------------	-----------	-----------

Caffeine / 50 μM	+	-	-	-
Lucifer Yellow / 60 μM	+	+	+	+
Compound / 50 μM	-	-	+	+
DMSO	1%	1%	1%	1%
P_{app} / nm s^{-1}	1217	n.d.	N/A	876

n. d. : not determined. N/A: not available.

2.4.6 Cell proliferation

Subsequently **I1** and **I2** were tested for their inhibition activity against the T47D cancer cell line in a cell proliferation assay. T47D cancer cell line was from the National Cancer Institute Developmental Therapeutics Program (NCI60 cell line panel). Cells were cultivated in a T25 flask in RPMI (Invitrogen, Carlsbad, CA) medium with 10% Fetal Bovine Serum (FBS, Gibco Invitrogen) and 100 U/mL Penicillin-Streptomycin (Thermo Scientific, Rockford, IL) and maintained in a humidified 37 °C incubator with 5 % CO₂. Subsequently, trypsin with EDTA were used to detach cells from the flask. Cells were counted using Hemocytometer and seeded equally into each well of a 96-well plate. Each well contains around 3000~4000 cells in culture medium with 10% FBS. After overnight culture, the medium was aspirated and then treated with different concentrations of **I1** and **I2** in DMSO (500 μM , 250 μM , 125 μM). Staurosporine (STS) was used as a positive control. The same volume of DMSO (final concentration in each well is 0.56%) was used as a negative control. After 2 days, the tetrazolium salt XTT solution (0.2 mg/mL XTT, 250 μM phenazine methosulfate (PMS), Roche) was added to the cells. After incubating at 37 °C for 20 h, the absorbance was measured at 450 nm and background absorbance was measured at 650 nm using a Biotek™ plate reader. Data was presented as the following:

$$viability = \frac{A_{450nm} - A_{650nm}}{a_{450nm} - a_{650nm}}$$

A_{450nm} and A_{650nm} are the absorbance at 450 nm and the background absorbance at 650 nm of the sample wells; a_{450nm} and a_{650nm} are the absorbance at 450 nm and the background absorbance at 650 nm of the control wells (which only contained DMSO). In the cell proliferation assay (Figure 2.7), we observed a corresponding dose-dependent inhibition of **I2** towards T47D cells. More importantly under the same concentration (250 μ M), **I2** was at least as effective as **I1** (see bars labeled with * in Figure 2.7), suggesting that although **I2** may have weaker binding towards the SH2 domain of STAT3 *in vitro* than **I1**, its better cell permeability and hydrolytic stability have apparently led to an improved *in vivo* anti-tumor activity.

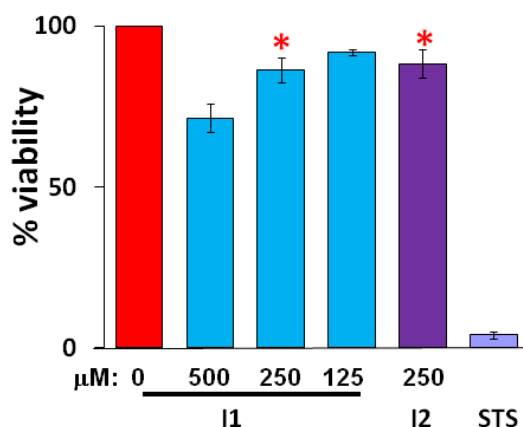


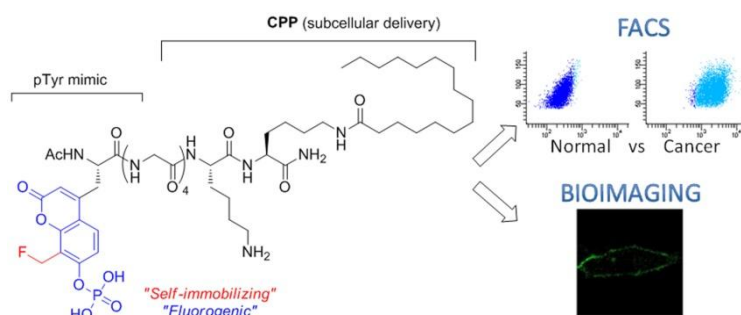
Figure 2.7 Relative % of T47D cell viability in the presence of the two inhibitors, **I2** (500 μ M, 250 μ M & 125 μ M) and **I1** (250 μ M). Negative control: DMSO only (0.56% final conc.). Positive control: staurosporine (STS, 500 nM). \pm SD was based on average of two independent experiments.

2.5 Conclusion

In summary, we have synthesized a novel unnatural amino acid which mimics phosphotyrosine but possesses better cell permeability and hydrolytic stability. Its incorporation into a known STAT3 SH2 domain inhibitor has confirmed our design principle. Future work will focus on incorporation of this amino acid into other *p*Tyr-containing biologically active compounds. This new *p*Tyr mimic, together with other newly developed approaches,^{50,87} will provide expanded chemical tools for future studies of PTP biology.

Chapter 3 A self-immobilizing and fluorogenic unnatural amino acid that mimics phosphotyrosine

Summary



This Chapter summarizes the designation and synthesis of the first self-immobilizing and fluorogenic unnatural amino acid that mimics phosphotyrosine (*pTyr*). By using solid-phase peptide synthesis, it was subsequently incorporated into peptide-based probes which found applications in bioimaging and fluorescence-activated cell sorting (FACS) of endogenous PTP activities in both organelle- and cell-specific manners.

3.1 Introduction

Reversible protein phosphorylation plays a fundamental role in signal transduction. Protein tyrosine phosphatases (PTPs; > 100 members in human), together with protein tyrosine kinases (PTKs), have an essential role in controlling this key cellular process and maintaining the intricate balance of the phosphoproteome network. Defective regulation of these enzymes has significant implications in human diseases.^{4, 9} Elevated levels of PTP activities are found in numerous tumor-associated cells and tissues.⁹ PTPs dephosphorylate proteins with a high degree of specificity inside the cell, although *in vitro* they have shown only moderate specificity towards synthetic peptide substrates.⁸⁸

Consequently, chemical and biological methods that report not only the global PTP activities, but more importantly the precise enzymatic activity of PTPs at single cellular and subcellular levels (and ultimately at the molecular level), may offer unprecedented views on how these enzymes work under physiological settings.⁸⁹ Existing chemical tools have so far focused on the detection of PTP activities either *in vitro* or *in situ* using recombinant PTPs or cellular lysates.^{62, 63, 64} More recently, a small molecule-based imaging probe capable of detecting PTP activities in live cells has appeared.⁹⁰ Notwithstanding, most probes are not able to detect PTP enzymatic activities on the basis of their substrate, organelle or cell specificity.⁹⁰ Herein, we have developed the first self-immobilizing and fluorogenic unnatural amino acid, which, upon incorporation into suitable peptide sequences, gave rise to chemical probes suitable for cell-based analysis (e.g. imaging and FACS) of endogenous PTP activities, thus successfully addressing some of the above-mentioned issues.

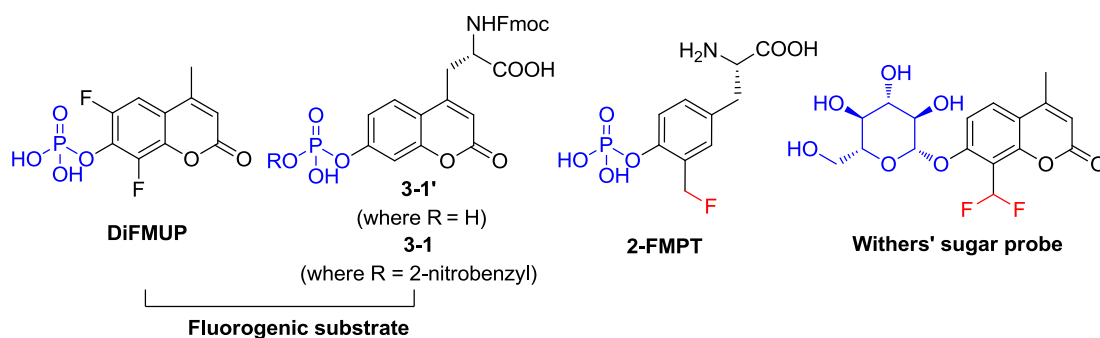


Figure 3.1 Representative examples of previously reported fluorogenic PTP substrates, **2-FMPT** and Withers' sugar probe.

Existing fluorogenic substrates, such as difluoromethylumbelliferone phosphate (DiFMUP; Figure 3.1), are widely used to detect PTP activities *in vitro* but are not suitable cell-based reagents due to their tendency to diffuse from the site of the

reaction.⁹¹ Activity-based probes of PTPs have been developed with varied degrees of success, but they generally inhibit PTPs irreversibly and do not provide Turn-ON fluorescence readouts, and thereby are not suitable for cell-based imaging and FACS experiments.^{62, 63, 64} An alternative strategy to minimize product diffusion is to, in the probe design, couple fluorescence release to an enzyme-probe covalent linkage.⁹⁰ Based on this principle, Withers and co-workers recently reported self-immobilizing fluorogenic imaging agents based on modified derivatives of coumarin glycosides (Figure 3.1) for histological and FACS studies of glycosidases.⁹² Inspired by this work, our newly developed unnatural amino acid (**3-2** shown in Figure 3.2), is a modified version of the phosphorylated coumaryl amino acid **3-1'** (and the “caged” version **3-1**; Figure 3.1), which was recently shown by Barrios and co-workers to be a close mimic of phosphotyrosine (*p*Tyr).³¹ We had previously shown in a separate study of another novel *p*Tyr mimic, **2-FMPT** (Figure 3.1), that the introduction of a 2-fluoromethyl group in the aromatic ring of *p*Tyr does not disrupt PTP recognition, but upon dephosphorylation generates a highly reactive quinone methide intermediate which subsequently attaches itself covalently to the nearby PTP.⁵⁰ **2-FMPT**, however, was not fluorogenic and cannot report PTP activities in real time in cells. Thus, our new *p*Tyr mimic **3-2** not only is self-immobilizing and fluorogenic (i.e. like Withers’s sugar probe), also possesses both N- and C-terminus (as in the case of naturally occurring amino acids), allowing essential PTP-recognizing peptide elements to be introduced when needed (Figure 3.2).

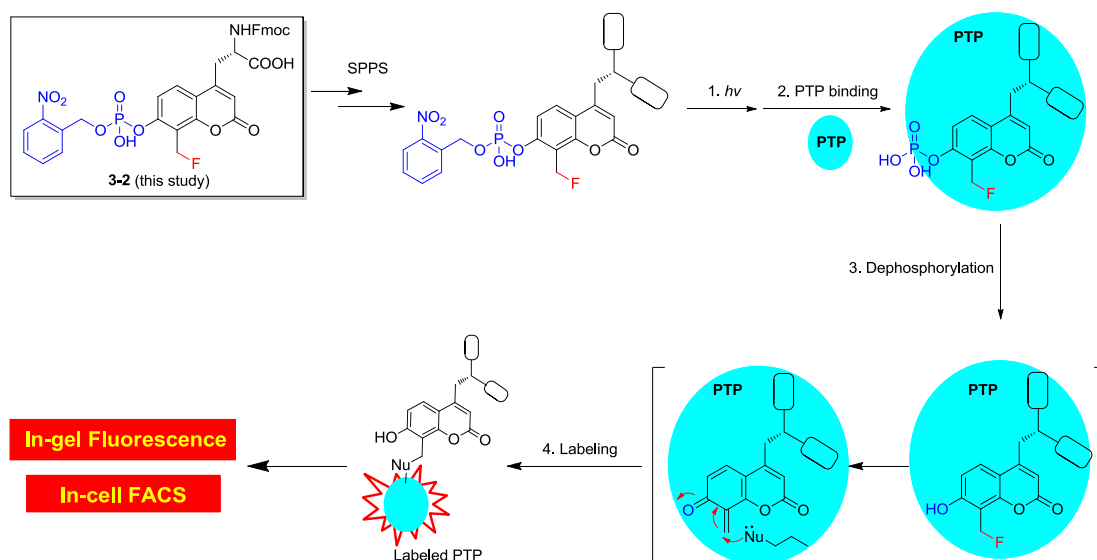
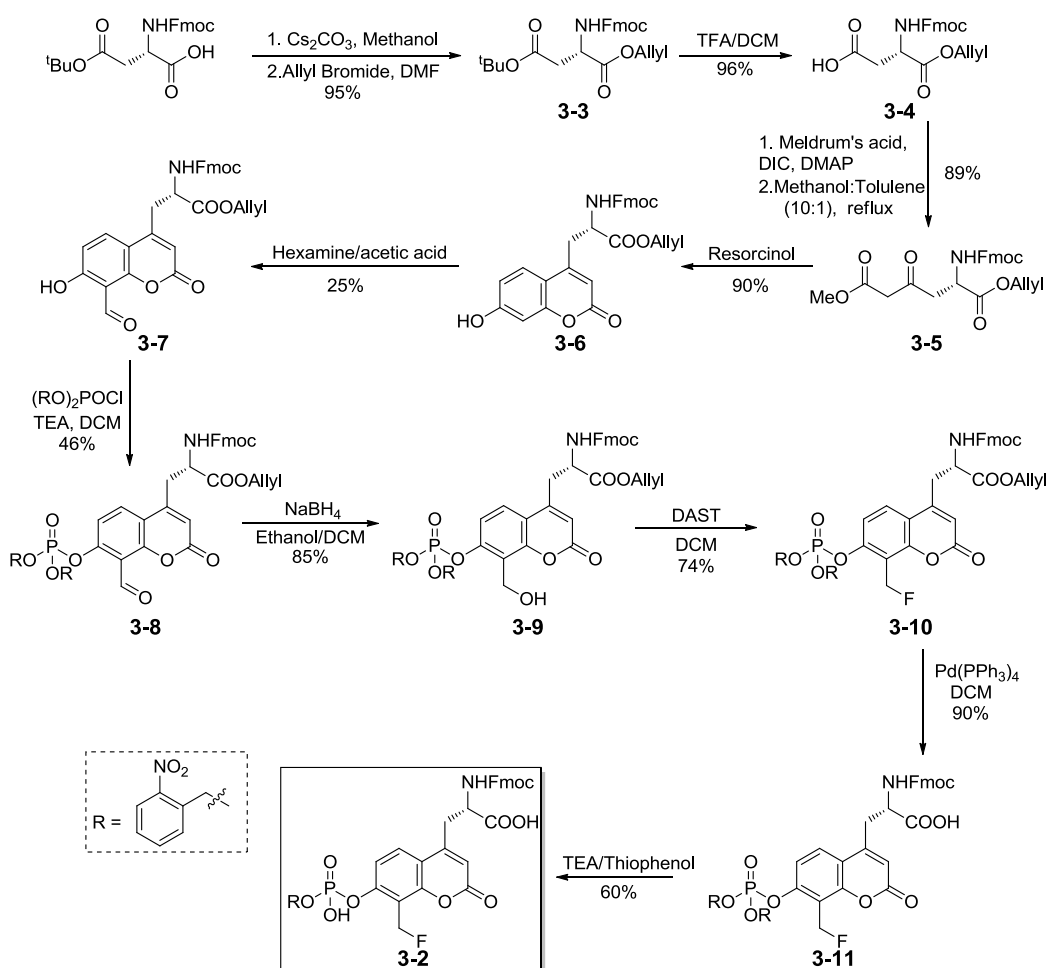


Figure 3.2 Overall working principle of the *p*Tyr mimic **3-2** and its peptide probes. With SPPS, PTP-recognizing elements or CPPs may be introduced into the probe. Upon UV-irradiation to uncage the phosphate, the peptide binds to its intended PTP, and gets dephosphorylated. The ensuing charge delocalization followed by spontaneous elimination of $-F$ generates the highly reactive quinone methide intermediate which, diffuses away from the enzyme active site before forming either covalent PTP-probe complex, or covalent complexes with nearby proteins (not shown).

3.2 Synthesis of unnatural amino acid

Compound **3-2**, the Fmoc-protected form of the unnatural amino acid, was synthesized in several steps from commercially available *N*- α -Fmoc-*L*-aspartic acid as shown in Scheme 3.1. A 2-nitrobenzyl photolabile protecting group was introduced in **3-2** to facilitate chemical synthesis and at the same time provide temporal control over PTP-probe recognition.^{50, 90, 93} Briefly, the first four steps in the synthesis were based on previously reported procedures,³¹ with some modifications, giving the allyl-protected coumaryl amino acid **3-6**. Subsequently, formylation of **3-6** by Duff reaction gave **3-7** in moderate yield (25%).⁹⁴ Next, phosphorylation with di(2-nitrobenzyl) chlorophosphate under basic conditions provided **3-8** (46% yield), which

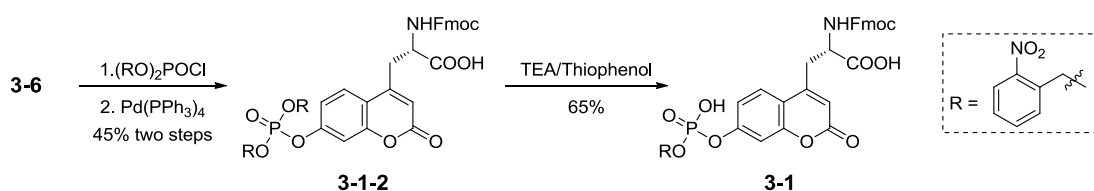
was reduced with NaBH₄ (giving **3-9**; 85% yield), followed by fluorination with DAST to give **3-10** (74% yield). Finally, the allyl group in **3-10** was removed by treatment with Pd(PPh₃)₄, giving **3-11** (90% yield). To obtain the new unnatural amino acid **3-2**, compound **3-11** was treated with TEA/thiophenol to remove one of 2-nitrobenzyl groups.



Scheme 3.1 Chemical synthesis of **3-2**

Furthermore, we also synthesized **3-1** as a control of **3-2**, since it is a fluorogenic substrate of PTP. There is no modification on the aromatic ring, it could not generate a reactive intermediate and continue on self-immobilizing function. The

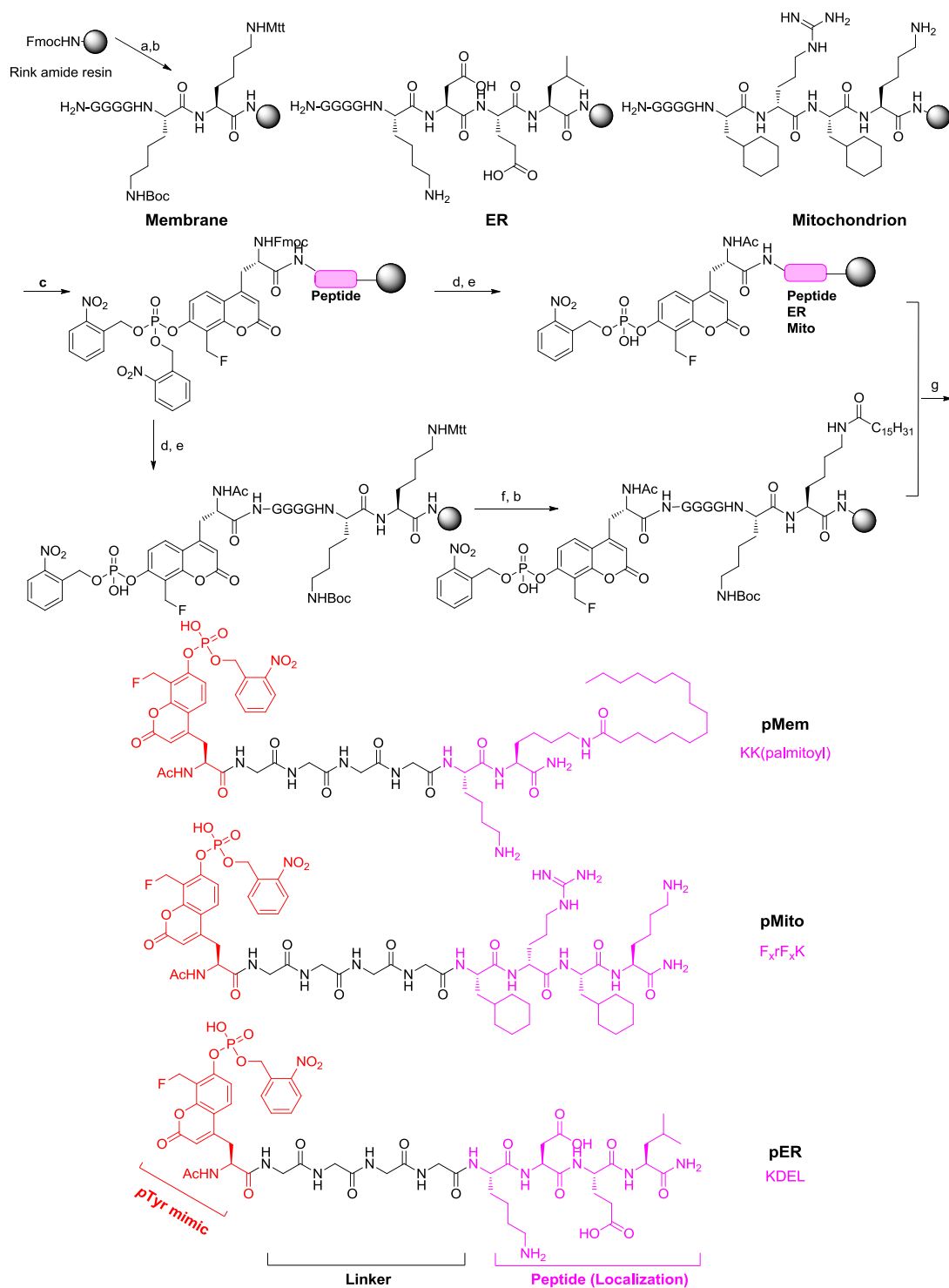
synthesis (Scheme 3.2) of **3-1** began with **3-6**. After phosphorylation with di(2-nitrobenzyl) chlorophosphate under basic conditions and deprotection of allyl group provided **3-1-2** (45% yield). Followed by removing one of 2-nitrobenzyl group was obtained the desired compound **3-1**.



Scheme 3.2 Chemical synthesis of **3-1**

3.3 Solid-phase synthesis of three localization peptides

To incorporate the unnatural amino acid into peptide-based probes, the precursor **3-11** was directly used in standard solid-phase peptide synthesis (SPPS) using Fmoc chemistry (Scheme 3.3). We chose to incorporate the self-immobilizing fluorogenic *p*Tyr mimic into three different cell-penetrating peptides (CPPs), giving the resulting peptide-based probes (**pMem**⁹⁵, **pER**⁹⁶, and **pMito**⁹⁷; Scheme 3.3). These localization peptides had previously been used to successfully deliver enzyme inhibitors into subcellular organelles of live cells.⁹⁸ A flexible (Gly)₄ linker was inserted between the CPPs and the unnatural amino acid to minimize potential interferences of PTP recognition by the CPPs. In the present work, we chose to focus on the prospect of these peptide-based probes as agents to study endogenous PTP activities through subcellular imaging and FACS experiments. In future, additional PTP-binding peptide fragments may be similarly introduced, by using the chemistry developed herein, to generate agents that address the substrate specificity of PTPs (Figure 3.2)



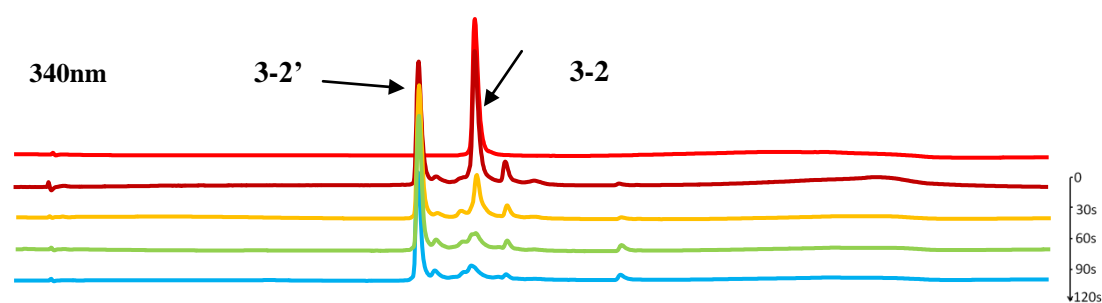
Scheme 3.3 Fmoc-based SPPS of the three CPP-containing probes. Reaction conditions: a) 20 % piperidine in DMF; b) Fmoc-protected amino acid or acid (Palmitic acid), HOBT, HBTU, DIEA, DMF; c) compound **3-11**, HOAT, HATU, DIEA, DMF; d) 2% DBU in DMF; e) acetic anhydride, DIEA, DCM; f) 1% TFA in DCM; g) TFA/TIS/H₂O (95:2.5:2.5).

3.4 Biological experiments and discussion

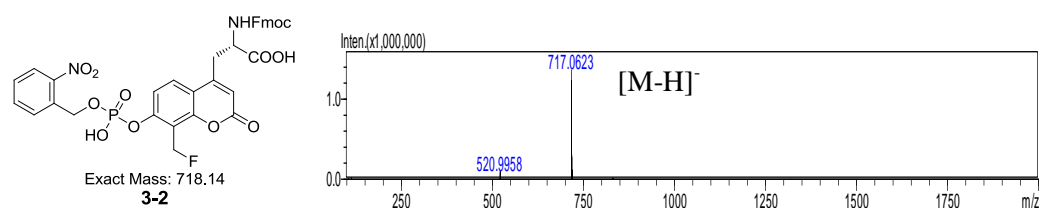
3.4.1 UV-initiated uncaging experiment

All the compounds and peptides we made have a photolabile protecting group, which could give better cell permeability and at the same time provide temporal control over PTP-probe recognition. So we first examined the photo- property of compound **3-2** and **pER** peptide (as a representative example). Briefly, compound **3-2** (50 μM) and **pER** peptide (50 μM), prepared in 25 mM Tris buffer with 50 mM NaCl at $\text{pH} = 7.2$, were photolyzed using a UVP CL-1000 Ultraviolet Crosslinker (500 $\mu\text{J}/\text{cm}^2$ at 254 nm). The reaction was monitored by LCMS. It was established that nearly complete uncaging ($> 95\%$) of the 2-nitrobenzyl group, giving the corresponding PTP-responsive adducts, could be achieved after 2 min under 500 $\mu\text{J}/\text{cm}^2$ UV exposure (Figure 3.3).

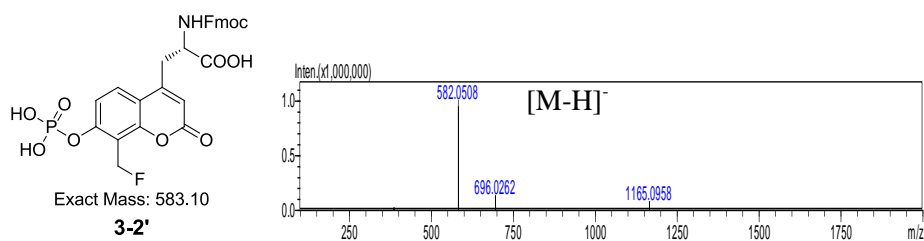
a) Compound **3-2**



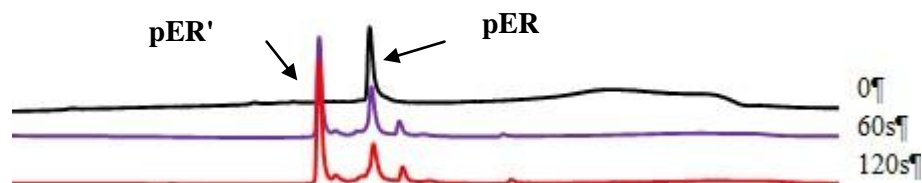
Before UV- irradiation



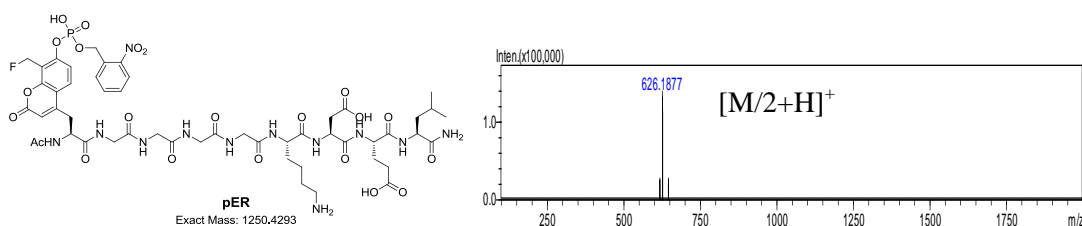
After UV-irradiation



b) **pER** peptide



Before UV-irradiation



After UV-irradiation

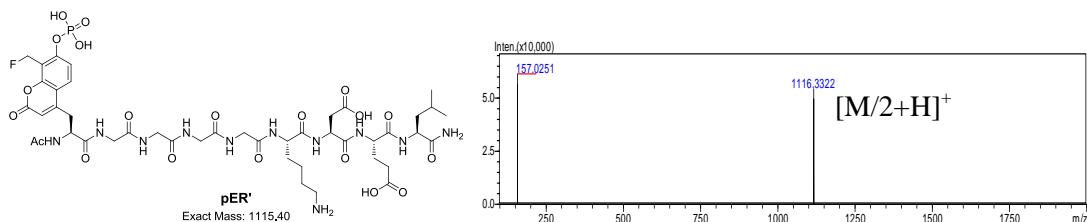
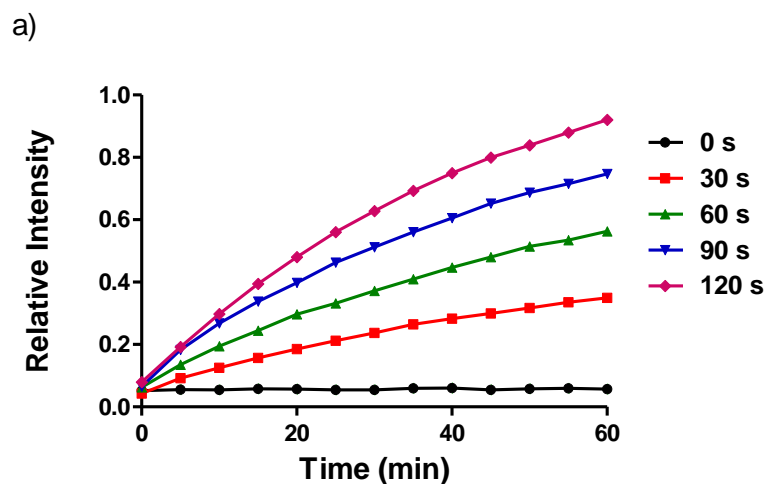


Figure 3.3 LC profiles and MS results of uncaging process. a) The LC profiles of **3-2** (50 μ M) upon UV-irradiation for different lengths of time (0, 30 s, 60 s, 90 s and 120 s). All peaks were identified and confirmed based on LCMS results. b) HPLC profiles of 50 μ M **pER** before (0 s) and after UV-irradiation (120 s). All peaks were identified and confirmed based on LCMS results.

3.4.2 Dephosphorylation reaction with uncaged compound

Compound **3-2** (4 μM in 25 mM Tris buffer, 100 mM NaCl, $\text{pH} = 7.26$; final volume is 25 μL), which were first UV-irradiated, or “uncaged”, for different lengths of time. Subsequently, recombinant PTP1B (a human PTP involved in diabetes and obesity, 80 nM final conc.)^{87b} was added and the dephosphorylation reaction was monitored continuously over 2 h using a BioTek multi-mode fluorescence microplate reader ($\lambda_{\text{ex}} = 360 \pm 10 \text{ nm}$; $\lambda_{\text{em}} = 460 \pm 10 \text{ nm}$) (Figure 3.4a). Microplate experiments were also performed to compare the dephosphorylation of caged and uncaged **pER** peptide (Fig. 3.4b). Briefly, HeLa cells were lysed using a buffer containing 50 mM Tris, 150 mM NaCl 0.02 % NP-40, at $\text{pH} = 8.0$. The total lysate was collected and its protein concentration was determined using Bradford assay. In each reaction, 15 μg (protein content) of the lysates was used, together with **pER** (4 μM final conc.), either non-irradiated or UV-irradiated (120 s), in a 25 μL final reaction volume. The reaction was then continuously monitored over 2 h in the BioTek microplate reader as above described. The results are shown that the addition of recombinant PTP1B or lysates of HeLa cells (a cancer cell line with elevated PTP activities) led to a time-dependent increase in fluorescence, with a concomitant release of the phosphate group.



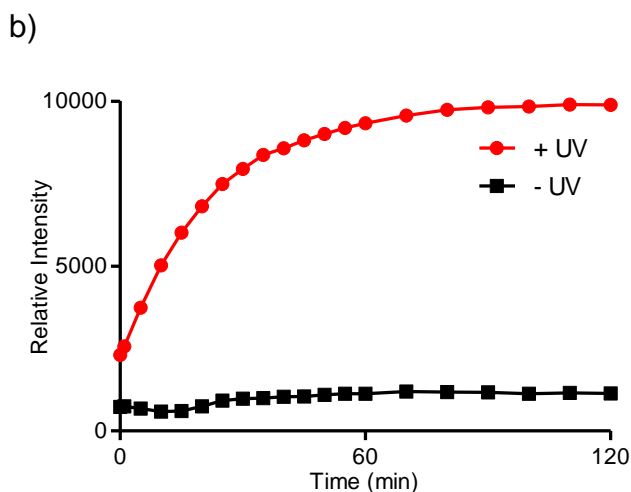


Figure 3.4 Time-dependent fluorescence measurements of UV-irradiated **3-2** and **pER**. a) Fluorescence measurements of **3-2**, upon UV-irradiation (for 0, 30, 60, 90, 120s), then treated with PTP1B. b) **pER** (4 μ M) in HeLa cell lysates. Control (\blacksquare): non-irradiated **pER** + lysates. The measurements were taken under coumarin channels ($\lambda_{\text{ex}} = 360 \pm 10$ nm; $\lambda_{\text{em}} = 460 \pm 10$ nm)

3.4.3 Enzyme specificity

Next, we carried out the microplate assays to test how selective the probe is towards PTPs. The microplate assays were performed in 384 wells. Compound **3-2** (20 μ M final conc.), which were uncaged by UV-irradiation. Subsequently, five enzymes (final 0.02 mg/mL) including recombinant PTP1B, recombinant glycosidase, recombinant Src (expressed and purified using the same protocol as reference 50), esterase (Fluka, No. 46069) and trypsin (Fluka, No. 93610) was added and the reaction was monitored continuously over 30 min using a BioTek multi-mode fluorescence microplate reader ($\lambda_{\text{ex}} = 360 \pm 10$ nm; $\lambda_{\text{em}} = 460 \pm 10$ nm). Results are shown in Figure 3.5, indicating our probe is only responsive towards PTPs as designed, not other kinase, proteases and glycosidase.

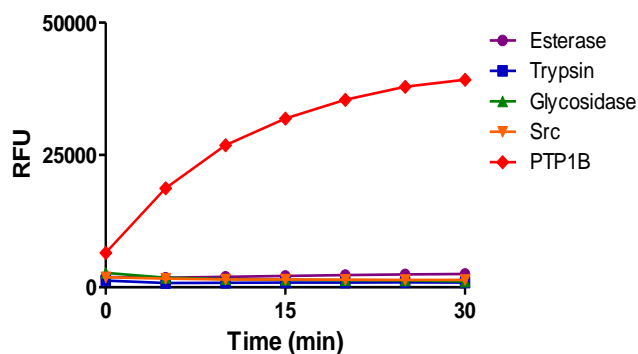


Figure 3.5 Microplate-based enzymatic assay of PTP1B and other four enzymes. The assay was done with uncaged compound **3-2** as the fluorogenic enzyme substrate.

3.4.4 Labeling experiments with pure PTP1B

In a recent study, it was found that the quinone methide generated in live cells underwent both self-quenching (> 90%; with aqueous media) and self-immobilizing (> 10%; with nearby proteins) pathways,⁹⁰ where the former produced a highly fluorescent diffusible dye molecule and the latter generated localized fluorescence.

First, we wondered whether the UV-irradiated **3-2**, when treated with PTP1B, behaves similarly and produces covalent adducts with proteins nearby (Figure 3.2).⁵⁰ Labeling with recombinant PTP1B was carried out. In a 20 μ L reaction, around 4 μ g PTP1B was incubated with the uncaged compound **3-2** (final conc. 1 mM, 500 μ M, 200 μ M and 100 μ M) at room temperature for 2 h. The reactions were stopped by addition SDS loading dye, followed by heating at 95 $^{\circ}$ C for 10 min. The mixture was resolved on a 10% SDS-PAGE gel and fluorescently labeled bands were photographed under UV light using CHEMI GENIUS Bio imaging system (Syngene). Results are shown in Figure 3.6. Most of **3-2**' upon treatments with PTP1B was converted to a diffusible small molecule dye (> 90%, a huge band at the front of the gel), while a small portion of it got self-immobilized to PTP1B. It indicated that

compound **3-2'** after hydrolysis by PTP1B was both self-quenching and self-immobilizing.

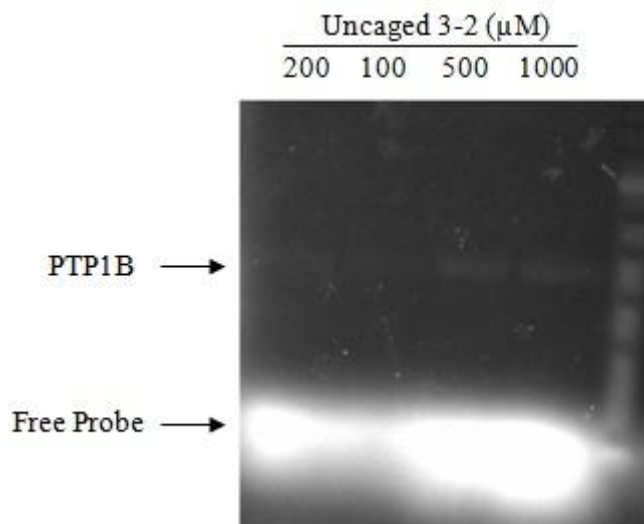


Figure 3.6 Labeling of PTP1B with different concentration of uncaged **3-2'**

3.4.4 Labeling experiments with bacterial lysate

Next, labeling with bacterial lysate was continued. Both with and without inducing PTP1B expression *Escherichia coli* BL21 DE3 cell lysate was prepared in 50 mM Tris buffer, 150 mM NaCl, 1% Triton X-100 pH = 7.26. Different concentrations of uncaged compound **3-1** and **3-2** were incubated with 30 μg lysate (protein content) for 2 h. Then the reaction was quenched by addition of SDS loading dye followed by heating at 95 °C for 10 min, then resolved on a 10% SDS-PAGE gel and photographed under UV light using CHEMI GENIUS Bio imaging system (Syngene). From the result, the majority of **3-2'** was still self-quenching (see free probe at the gel front; lane 1), a small amount of PTP1B and other cellular proteins were also labeled (caused by self-immobilization of **3-2'**). Bariros' coumaryl amino

acid **3-1'**, which is a non self-immobilizing analog of **3-2'**, was used as a negative control; no labeled protein was detected (lane 2).

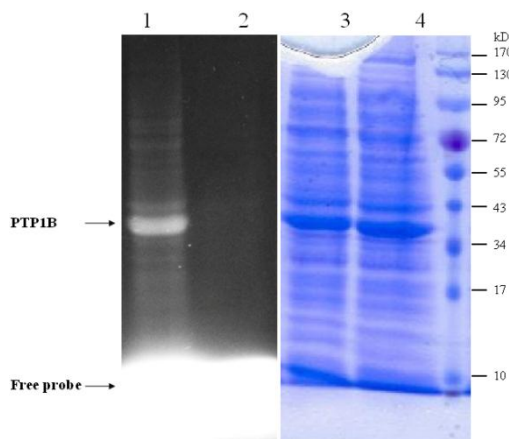


Figure 3.7 Fluorescence gel (imaged under a UV transilluminator) of PTP1B-overexpressed bacterial lysate treated with uncaged **3-2** (1 mM; lane 1) and **3-1** (0.5 mM; lane 2). The corresponding coomassie gel was shown on the left. A small amount of covalent labeling between **3-2'** and PTP1B and other cellular proteins was observed. Most of **3-2'** was released as the highly fluorescent free probe (arrowed).

3.4.5 Competitive gel analysis

Since the coumarin fluorescence in the probes could not be detected with our fluorescence gel scanner, we also used an indirect competition assay by labeling pre-treated PTP1B (with either different amounts of uncaged **3-2**, or the same amount of uncaged **3-2** incubated for different lengths of time) with a previously reported PTP1B peptide probe **K3** (Rhodamine-GGKAVDGXVKPQ-NH₂; X = **2-FMPT**⁵⁰). We still used uncaged **3-1** as a control for comparison; both dose- and time-dependent inhibitions of labeled PTP1B-**K3** band were observed, but only at a very high concentration of the compound (Figure 3.8). In addition, similar inhibitory effect was observed for both compounds **3-1'** and **3-2'**. These results indicated that although

some amount of uncaged **3-2** might have competed for covalent labeling of PTP1B with **K3** to a certain extent, the majority of it failed to do so (likely as a result of self-quenching).

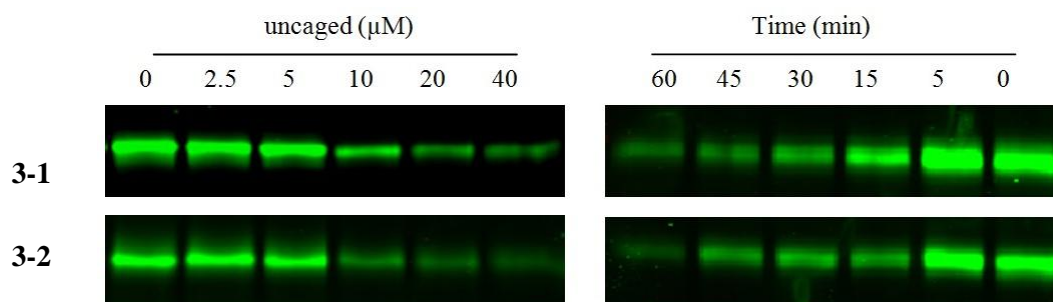


Figure 3.8 Indirect competition assay. Covalent labeling of PTP1B with **K3**, after PTP1B was pre-incubated with uncaged **3-1** or **3-2**. (Left) Dose-dependent experiments: different concentrations of compound **3-1** and **3-2** (upon 300s UV-irradiation) were incubated with PTP1B (1 μg) for 1.5 h in 25 mM Tris buffer (containing 50 mM NaCl at $\text{pH} = 7.26$). Subsequently, **K3** (final conc.: 20 μM) was added to the mixture and the reaction was further incubated for 1 h. Finally, the reaction was quenched by addition of 6x SDS loading dye followed by heating at 95 $^{\circ}\text{C}$ for 10 min, then resolved on a 12% SDS-PAGE gel and visualized by in-gel fluorescence scanning with Typhoon 9200 fluorescence gel scanner. (Right). Time-dependent experiments: 80 μM of compound **3-1** and **3-2** (upon 300s UV-irradiation) were incubated with PTP1B (1 μg) for different lengths of time (0, 5, 15, 30, 45, 60 min) in a 25 mM Tris buffer (containing 50 mM NaCl at $\text{pH} = 7.26$). Subsequently, **K3** (final conc.: 20 μM) was added to the mixture and the reaction was further incubated for 15 min. Finally, the reaction was quenched by addition of 6x SDS loading dye followed by heating at 95 $^{\circ}\text{C}$ for 10 min, then resolved on a 12% SDS-PAGE gel and visualized by in-gel fluorescence scanning.

3.4.6 Mass spectrometry

PTP1B treated with uncaged **3-2** was further analyzed by mass spectrometry. 25 μg of recombinated PTP1B was labeled by uncaged compound **3-2** (1:10, PTP1B/**3-2** molar ratio) overnight at room temperature. This sample, together with

unlabelled PTP1B as a control, was analyzed with a micromass Q-TOF 2 tandem mass spectrometer (Figure 3.9); results indicated most of protein still retained its original molecular (MW = 39344), therefore was mostly not covalently modified.

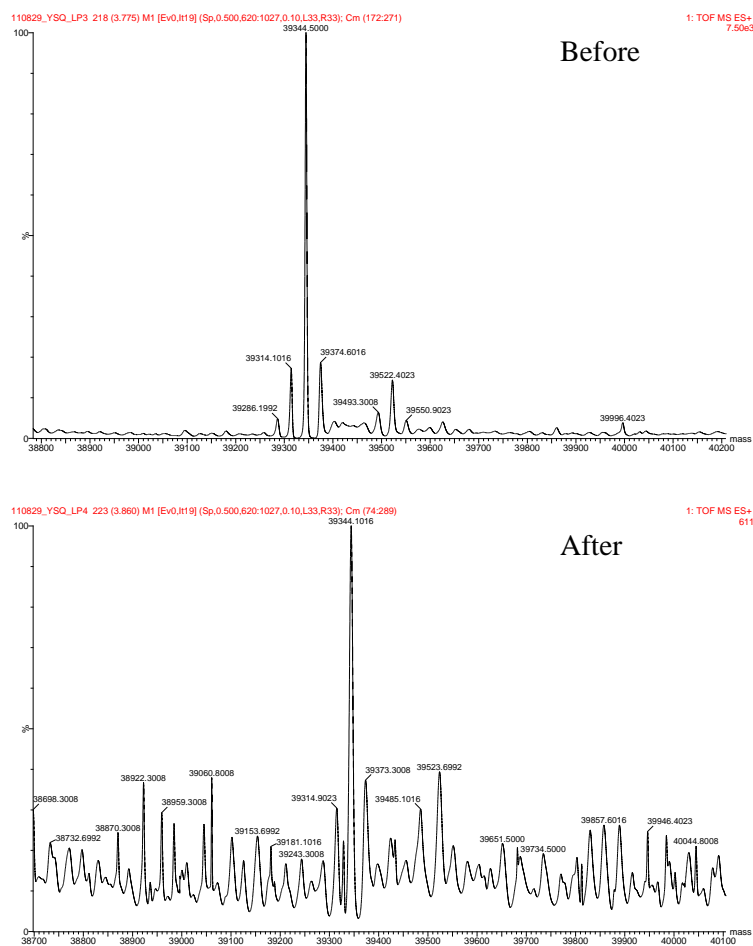


Figure 3.9 MS profiling of PTP1B before and after incubation with uncaged **3-2**.

3.4.7 Effects of the activities of PTP1B as a model

The assays were performed in HEPES buffer (25 mM HEPES, 0.05 M NaCl, 2.5 mM EDTA and 0.02% Triton X-100, pH 7.5). The enzyme was incubated with different dilutions of compound **3-1'** and **3-2'** (final concentration: 250, 125, 62.5, 50, 25, 12.5, 6.25, 1.25 μ M). The dephosphorylation reaction was monitored continuously over 2 h using a BioTek multi-mode fluorescence microplate reader ($\lambda_{ex} = 360 \pm 10$

nm; $\lambda_{em} = 460 \pm 10$ nm). Kinetic constants were computed by direct fits of the data to the Michaelis-Menton equation using a non-linear regression via GraphPad Prism software. The value was taken in mean \pm S.D. in two sets of data with duplicate. From Table 3.1, K_{cat} and K_M of the two compounds against PTP1B are similar, indicating small structural changes from **3-1'** to **3-2'** did not affect enzyme recognition.

Table 3.1 Kinetic Data of compound **3-1'** and **3-2'** with PTP1B.

	K_M (mM)	k_{cat} (s^{-1})	K_{cat}/K_M ($mM^{-1}\cdot s^{-1}$)
Compound 3-1'	0.77 ± 0.20	11.8 ± 2.4	15.2 ± 0.2
Compound 3-2'	1.1 ± 0.4	8.8 ± 2.6	7.9 ± 0.1

3.4.8 Comparison of enzyme activities before and after probe reaction

Lastly, PTP1B treated with uncaged **3-2** was further analyzed by enzymatic assay to compare the activity of PTP1B before and after probe reaction. The assays were performed in Hepes buffer (25 mM Hepes, 0.05 M NaCl, 2.5 mM EDTA and 0.02% Triton X-100, pH 7.5). PTP1B (2 μ g/mL final conc.) was incubated with compound **3-1'** and **3-2'** (0.2 μ M final conc.). The control reaction is adding the same amount of water replacing compound **3-1'** or **3-2'**. After 1 h incubation, DiFMUP (4 μ M final conc., Invitrogen, No. D6567) was then added. The dephosphorylation reaction was monitored continuously over 10 min using a BioTek multi-mode fluorescence microplate reader ($\lambda_{ex} = 360 \pm 10$ nm; $\lambda_{em} = 460 \pm 10$ nm). The kinetic curve was given after subtracting the background due to the dephosphorylation of compound **3-1'** and **3-2'**. The results showed that PTP1B's activity did not changed and still retained its original enzymatic activity after incubation, which also indicated

most of PTP1B is the reaction was not covalently labeled and the enzyme can still carry out multi-turn over conversion and achieve signal amplification.

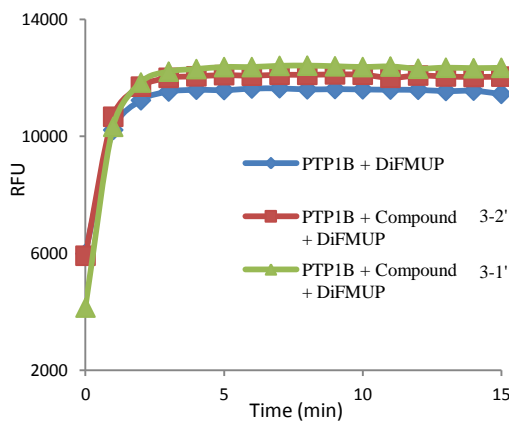


Figure 3.10 Comparison of PTP1B's activity, with and without compound **3-1'** and **3-2'** labeling, indicates that, no significant decrease of activity was observed.

3.4.9 FACS experiments

The self-immobilizing Turn-ON fluorescence properties of these new PTP probes make them suitable for FACS experiments.⁹² Flow Cytometry analysis was performed on a BD LSR II Bioanalyzer equipped with a 355 nm UV air-cooled laser. The data was analyzed using BD FACSDiva software. Cells were seeded in Nunc™ (Cat. No. 153066) 3.5 cm dishes and grown till ~90% confluence. Two sets of cells were prepared. One set of cells was UV-irradiated for 1 min. After pelleting the cells by centrifugation, discarding the supernatant and adding a fresh PBS buffer, Propidium iodide (PI, 2 μM final conc.) was added and the cells were stained for 10 min at room temperature. Then the cells were rinsed with 1x PBS and directly analyzed by flow cytometry. The results indicated that UV-treated cells and untreated cells did not appear to show any difference. Manual inspection of the cells are showed

1-min UV irradiation on cells did not appear to have any noticeable morphology effect (data was not shown).

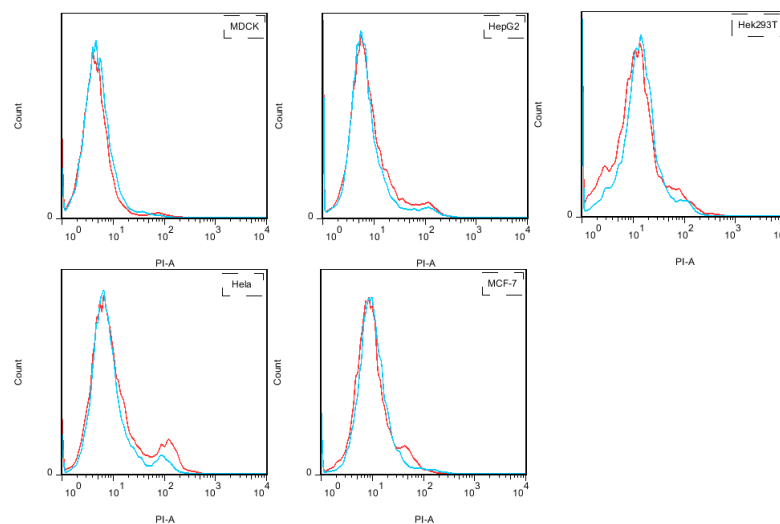


Figure 3.11 Histogram of flow Cytometry analysis of dead cells after 1 min UV-irradiation (five cell lines, MDCK, HEK293T, HepG2, HeLa, MCF-7). Red line: cells stained by PI; Blue line: UV-irradiated cells stained by PI.

HeLa cells were first incubated with **3-2** for 30 min, allowing sufficient time for the compound to enter the cells. Cells were subsequently UV-irradiated for 1 min then further incubated for 2 h. Upon washing with PBS (2x), the cells were detached and finally fixed with 1% paraformaldehyde in PBS before being subjected to FACS analysis. Compound **3-1** was used as a negative control. As shown in Figure 3.12, we observed a noticeable increase in the number of highly fluorescently labeled cells, over both non-irradiated cells and cells treated with **3-1**. The results clearly indicated that **3-2** was able to strongly and permanently “stain” the cells through its self-immobilizing property. No leakage of **3-2** from the cells was observed (data not shown) even after repeated washing of cells. On the other hand, cells stained with **3-1**

(the non self-immobilizing analog of **3-2**) showed much weaker fluorescence, which rapidly disappeared upon repeated washing (not shown).

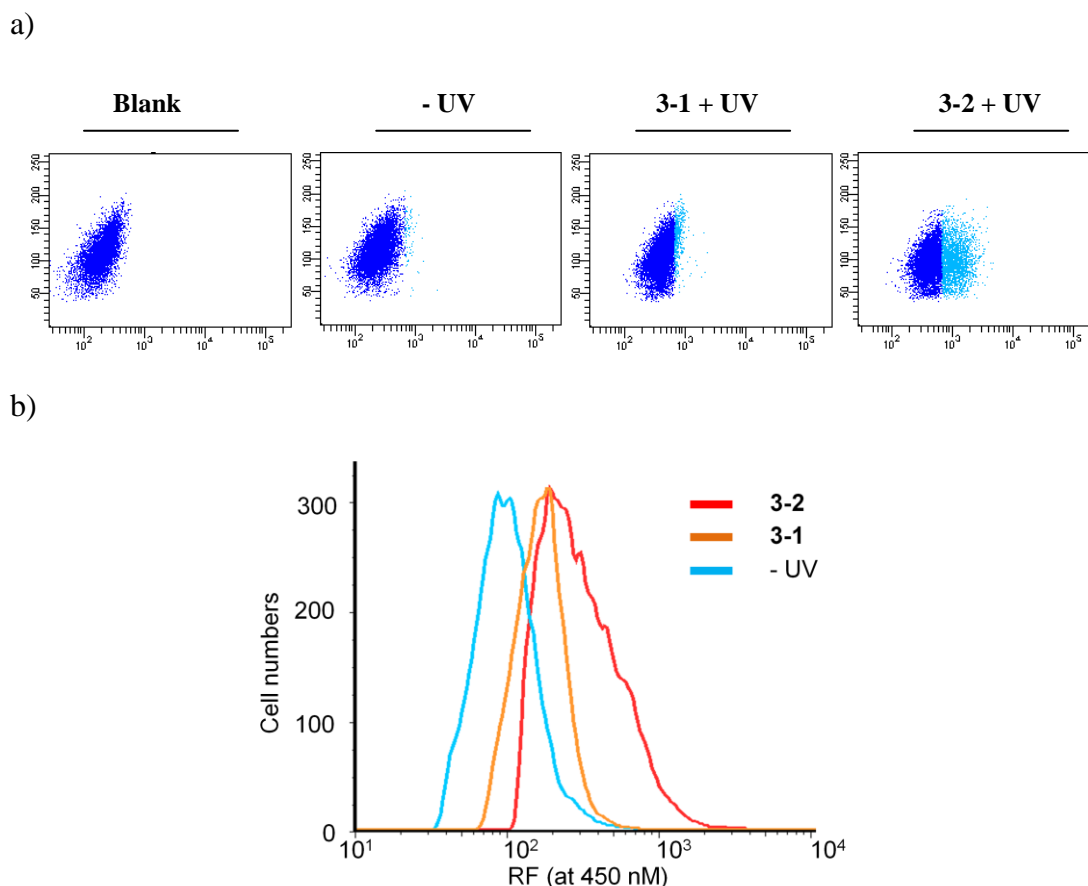
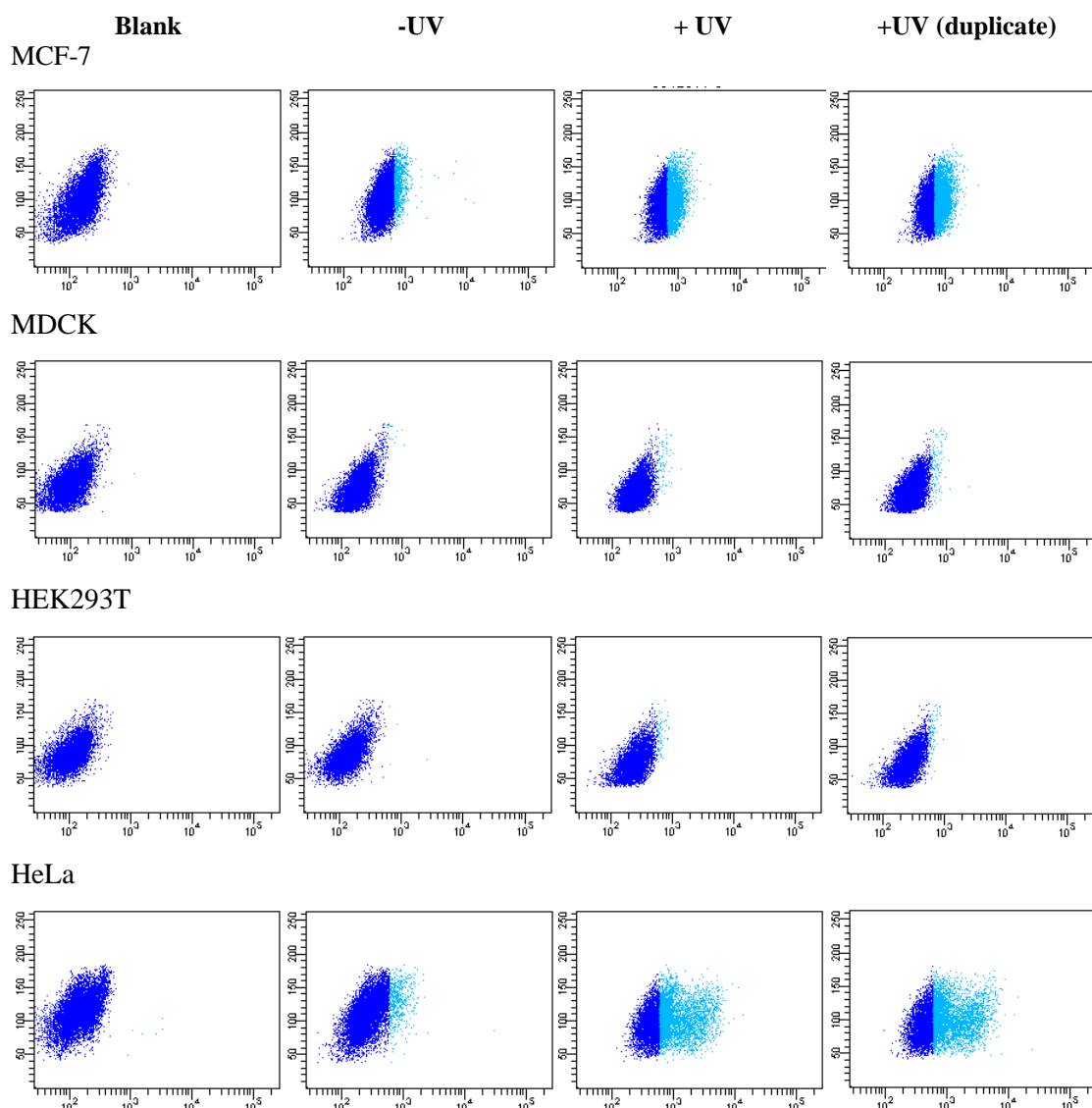


Figure 3.12 a) Density-plots of FACS analysis of HeLa cells treated with 80 μM of **3-1** and **3-2**. Blank: only cells (without any added compounds). “- UV”: cells treated with compound **3-2** but without UV-irradiation. UV + **3-1/3-2**: cells treated with compound **3-1/3-2** followed by UV-irradiation. The boundary between two colors corresponds to the maximum fluorescence in the blank sample. b) FACS histogram of HeLa cells treated with 80 μM of uncaged **3-1** or **3-2** (“-UV” represents cells treated with non-irradiated **3-2**, a negative control).

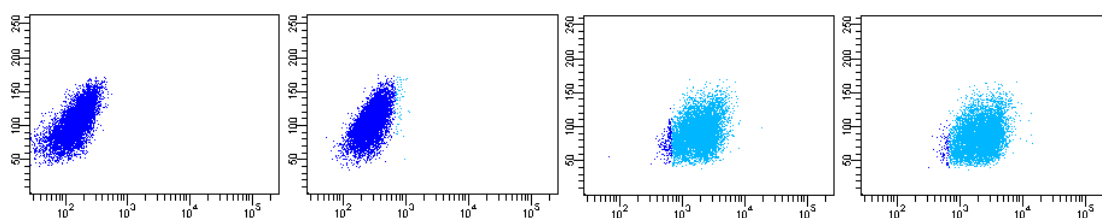
We next used **3-2** to detect PTP activities in a cell type-specific manner, in anticipation that most cancers have elevated PTP expression. For FACS experiments with normal and cancer cells, the cells were seeded in NuncTM (Cat. No. 153066) 3.5 cm dishes and grown till ~90% confluence. Next, compound **3-2** (80 μM) was added

and the cells were incubated further for 30 min. Upon UV-irradiation for 1 min, cells were further incubated for 2 h, then washed with PBS (2x), detached from dishes, washed (with PBS), fixed with 1% paraformaldehyde in PBS for 10 min, then analyzed by FACS. While UV irradiation itself didn't result in any apparent effect on the cells (Figure. 3.11), FACS analysis of 5 common mammalian cells, shown in Figure 3.13, indicated significantly higher fluorescence counts in cancer cells treated with 3-2 (MCF-7, HeLa & HepG2), possibly caused by their elevated endogenous PTP activities.

a)



HepG2



b)

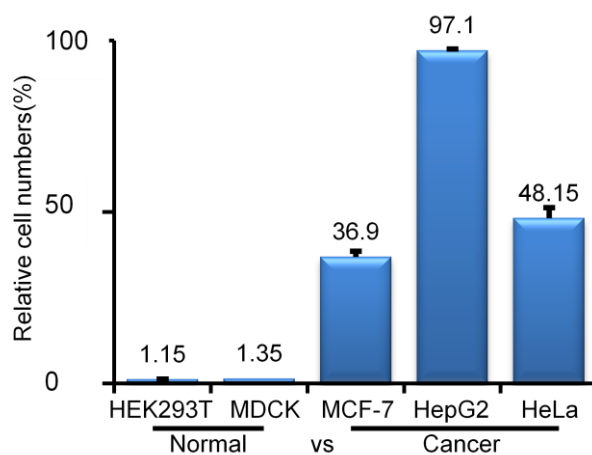


Figure 3.13 FACS analysis of five cell lines. a) Density-plot of FACS analysis of five cell lines treated with 80 μM of compound **3-2**. Blank: only cells (without any added compounds). The UV-irradiated samples were performed in duplicate (two right panels). The boundary between two colors corresponds to the maximum fluorescence in the blank sample. b) Summary of analysis of five cell lines. The data were extrapolated. Relative numbers of stained cells (by uncaged **3-2**; 80 μM) from the FACS results of different mammalian cells.

3.4.10 Microscopy experiments

Finally, we showed that **3-2**, upon incorporation into suitable peptide sequences, could be used for subcellular imaging of PTP activities in live cells. The cells were incubated with 20 μM of the peptide (either **pMito**, **pMem** or **pER**) in fresh Dulbecco's Modified Eagle Medium (DMEM). The cells were further incubated (1.5 h for **pER** and **pMito**, 30 min for **pMem**). Then the medium was refreshed. Upon UV-irradiation for 1.2 min, the cells were further incubated for 1.5 h. Next, 0.25 $\mu\text{g/mL}$ of a corresponding organelle tracker was added (ER-TrackerTM: Red

glibenclamide BODIPY[®] TR; CellMask[™] Plasma Membrane Stains C10045; MitoTracker[®]: Red CMXRos dye; all from Invitrogen). Upon further incubation according to the vendor's protocols, the cells were washed with PBS (3x), then fixed with 3.7% formaldehyde for 15 min at 37 °C. The stained cells were finally imaged with a Leica TCS SP5X confocal microscope system and processed with the Leica application suite advanced fluorescence (LAS AF).

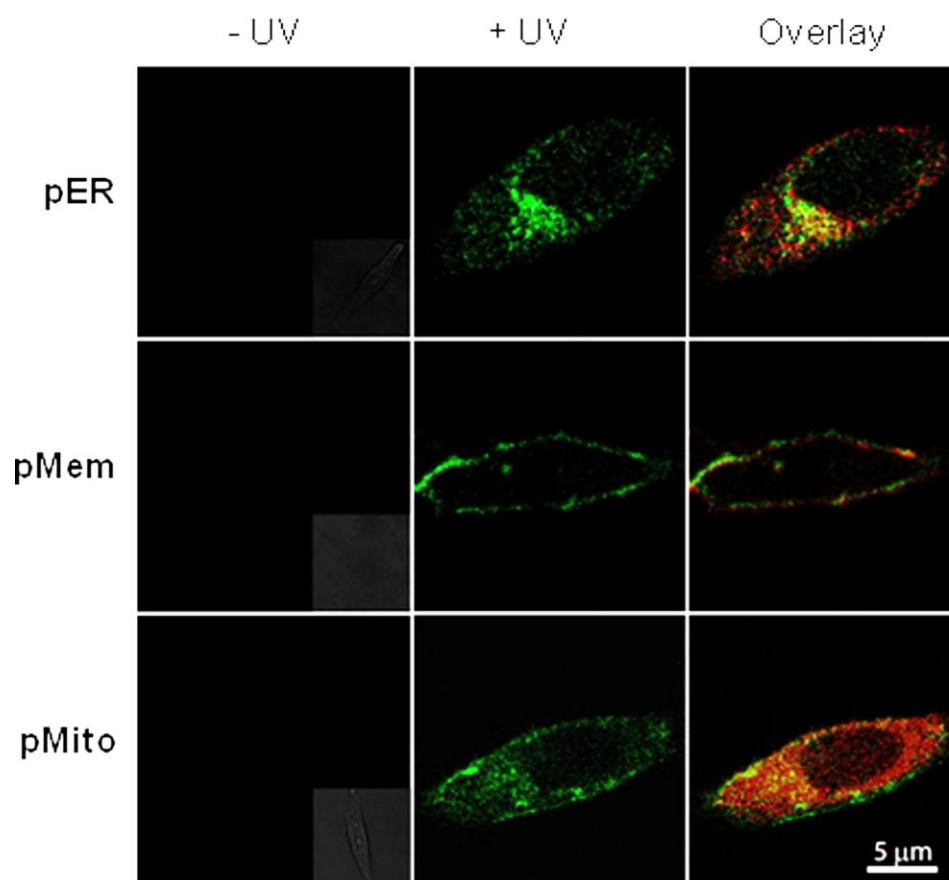


Figure 3.14 Fluorescence images of HeLa cells treated with different non-irradiated peptide probes (20 μM, 30 - 90 min incubation). Without UV-irradiation, the cells were directly imaged. No fluorescence was observed in any of the images, indicating the temporal control of our “caging” strategy. With UV-irradiated, further incubated (1.5 h) then imaged (in coumarin channel). “Overlay”: images from the coumarin channel overlaid with the tracker channel. Excitation channel: peptide channel (405 nm Diode laser); tracker channel (white laser, 590 nm for ER-Tracker, 543 nm for CellMask, 554 nm for Mito-Tracker, respectively).

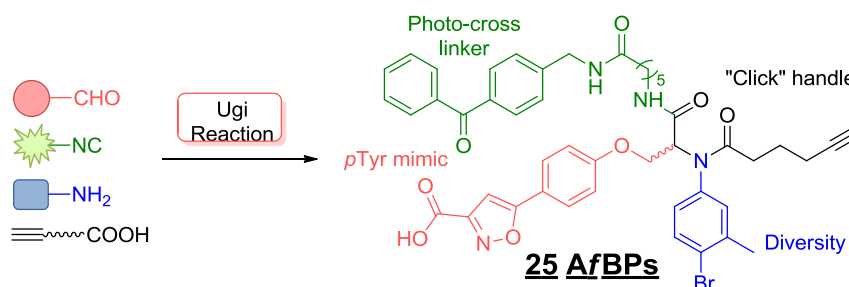
Figure 3.14 (- UV) is the fluorescence images HeLa cells treated with different non-irradiated peptide probes (negative controls). The result shows the fluorescence images of HeLa cells treated with or without the different uncaged peptide probes. **pMem**, **pER**, and **pMito**, each containing a CPP, served to deliver our *pTyr* mimic to different subcellular organelles (membrane, endoplasmic reticulum and mitochondria, respectively). Cells were subsequently imaged to detect subcellularly localized, endogenous PTP activities; all three peptides gave rise to strong fluorescence signals only in their intended organelles, indicating successful delivery and subcellular imaging of PTPs.

3.5 Conclusion

In conclusion, we have shown the utility of this newly developed, self-immobilizing and fluorogenic *pTyr* mimic in studying endogenous PTP activities using FACS and bioimaging experiments. Future work will focus on incorporation of this unnatural amino acid into peptides and proteins, to engineer the corresponding enzyme sensors capable of addressing the substrate specificity of individual PTPs.

Chapter 4 Ugi reaction-assisted rapid assembly of affinity-based probes against potential protein tyrosine phosphatases

Summary



This chapter summarizes the development of a new method, the multi-component Ugi reaction to assemble a small library of affinity-based probes (AfBPs) that target potential protein tyrosine phosphatases (PTPs). The probes showed good inhibition and labelling of PTP1B and MtpB, and were subsequently used to label endogenous PTP1B in MCF-7 cancer cells.

4.1 Introduction

Protein tyrosine phosphates (PTPs) are essential enzymes in the phosphoproteome networks. Dysregulation of these enzymes are implicated in a variety of human diseases. For example, protein tyrosine phosphatase 1B (PTP1B) is a well-known negative regulator of the insulin pathway and its malfunction can cause diabetes and obesity.⁹⁸ MtpB, another well-known protein tyrosine phosphatase encoded by *Mycobacterium tuberculosis*, is a promising target for new tuberculosis (TB) drugs (Figure 4.1).⁹⁹ To better understand both the *in vitro* and *in vivo* enzymatic activities of PTPs, new chemical and biological methods

have emerged in recent years. Among them, strategies based on activity-based protein profiling (ABPP) are particularly appealing, as they enable sensitive detection of active PTPs present in a crude proteome. (Chapter 1 section 1.4) Depending on the design and the probes used, it is possible to profile different PTPs and obtain their substrate specificity.⁵⁰ ABPP typically operates with small molecule activity-based probes (ABPs) derived either from suicide inhibitors or reversible inhibitors of the target enzyme, and in the later cases, these probes are more accurately referred to as affinity-based probes (A/BPs).^{51, 100} Compared to the traditional ABPs, A/BPs provide an easily accessible tool to most protein classes (including enzymes and non-enzymes) by conversion of reversible inhibitors into covalent protein-profiling probes with a suitable photo-cross linker and a reporter tag.^{61, 101} Many A/BPs have been developed so far against different enzymes, but none against PTPs.

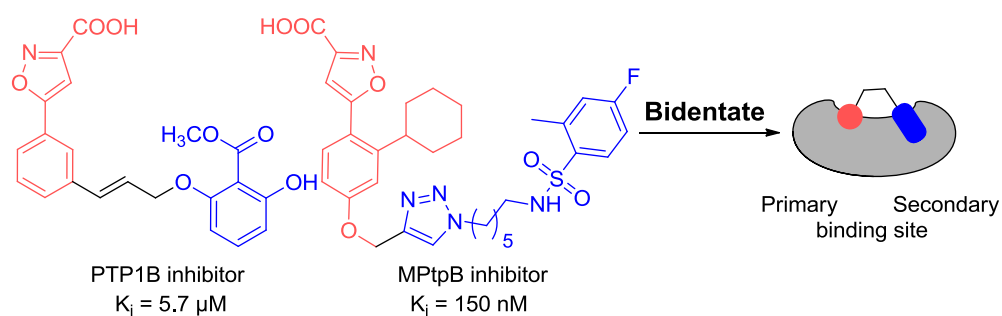


Figure 4.1 Structures of representative PTP1B/MptpB bidentate inhibitors with the core group (in red) binding to the primary binding site and the peripheral group (in blue) binding to the secondary binding site of PTP.

Herein, we report the design and synthesis of A/BPs of PTPs. In order to obtain a sizeable panel of A/BPs suitable for substrate specificity profiling of

PTPs, we have used the multi-component reaction (MCR) to rapidly assemble these probes in a single step (Figure 4.2). To the best of our knowledge, this is the first report in which MCR has been applied for the synthesis of Af/BPs.

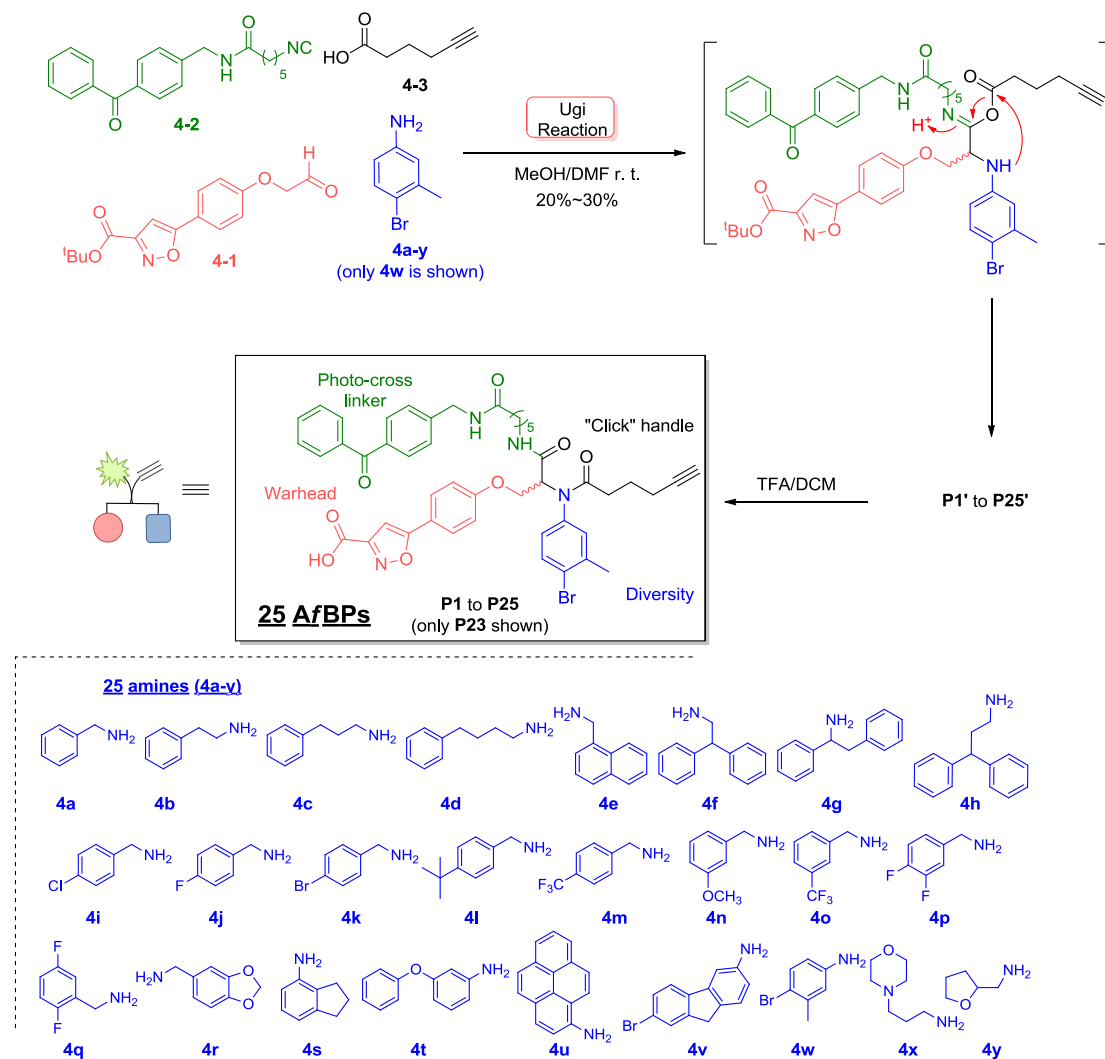


Figure 4.2 Chemical synthesis of the 25 Af/BPs using Ugi reaction. The four components – the aldehyde **4-1** (in red), the isonitrile **4-2** (in green), the “clickable” **4-3** (in black) and the 25 different diversity groups used **4a-y** (in blue) - were shown, and the corresponding Af/BPs were assembled in one-pot to give the 25 *t*Bu-protected Af/BPs, **P1'** to **P25'**, which upon TFA treatment, gave the 25 final Af/BPs, **P1** to **P25** (boxed).

Recently, it has been discovered that many PTPs, including PTP1B and

MptpB, possess a highly distinct secondary binding site in addition to their homologous primary active site where phosphotyrosine (*p*Tyr) binds (Figure 4.1).⁸⁸ This thus makes it possible to develop bidentate inhibitors which could impart both potency and specificity against the target PTP (Figure 4.1).^{76, 77, 102} Although MCR has been extensively used in drug discovery,¹⁰³ including the synthesis of PTP inhibitors,¹⁰⁴ its application for rapid assembly of AfBPs has not been explored. By carefully considering the structural property of previously developed bidentate PTP inhibitors (Figure 4.1), where a *p*Tyr mimic (shown in red) is connected to a diversity group (shown in blue) through a suitable linker, we conceived an inhibitor-to-AfBP conversion could be conveniently made through the Ugi reaction using the following four components (Figure 4.2): 1) an aldehyde-containing isoxazole carboxylic acid, **4-1**, that mimics *p*Tyr and anchors the AfBP to the PTP active site; 2) an amine-containing diversity component, **4a-y**, which is readily accessible from commercially available amines and may bind to the secondary binding site of the PTP; 3) an isonitrile-containing benzophenone, **4-2**, which upon UV irradiation would initiate covalent cross-linking between the PTP and the AfBP; 4) a terminal alkyne-containing carboxylic acid, **4-3**, which serves as a clickable tag for subsequent in-gel fluorescence scanning and pull-down experiments by conjugation to suitable reporters (rhodamine-N₃ or biotin-N₃ in Figure 4.3) via the bioorthogonal click chemistry.¹⁰⁵ Direct introduction of a fluorescent/biotin tag in the AfBP was not desirable as the bulky reporter might disrupt the binding between the PTP and the probe.⁵⁴ In this study, a total of 25 AfBPs targeting potential PTPs have been successfully assembled in a single step, and their ability for ABPP against recombinant PTPs and endogenous PTPs and other cellular off-targets investigated (Figure 4.2).

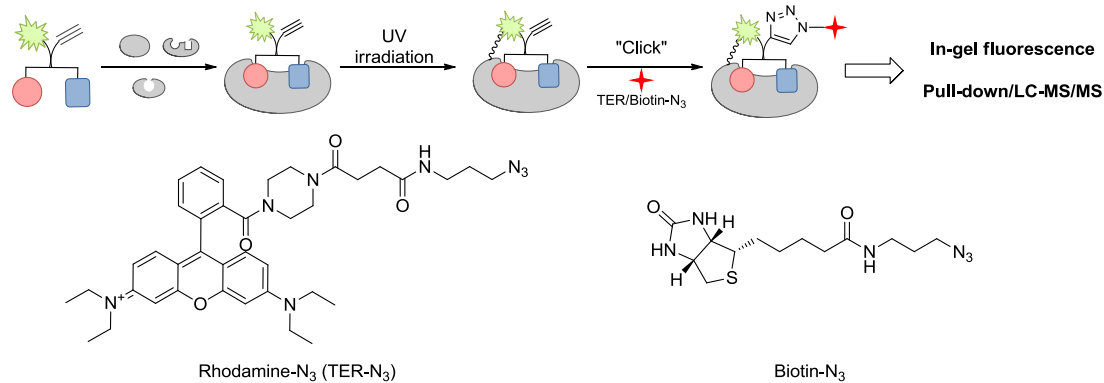


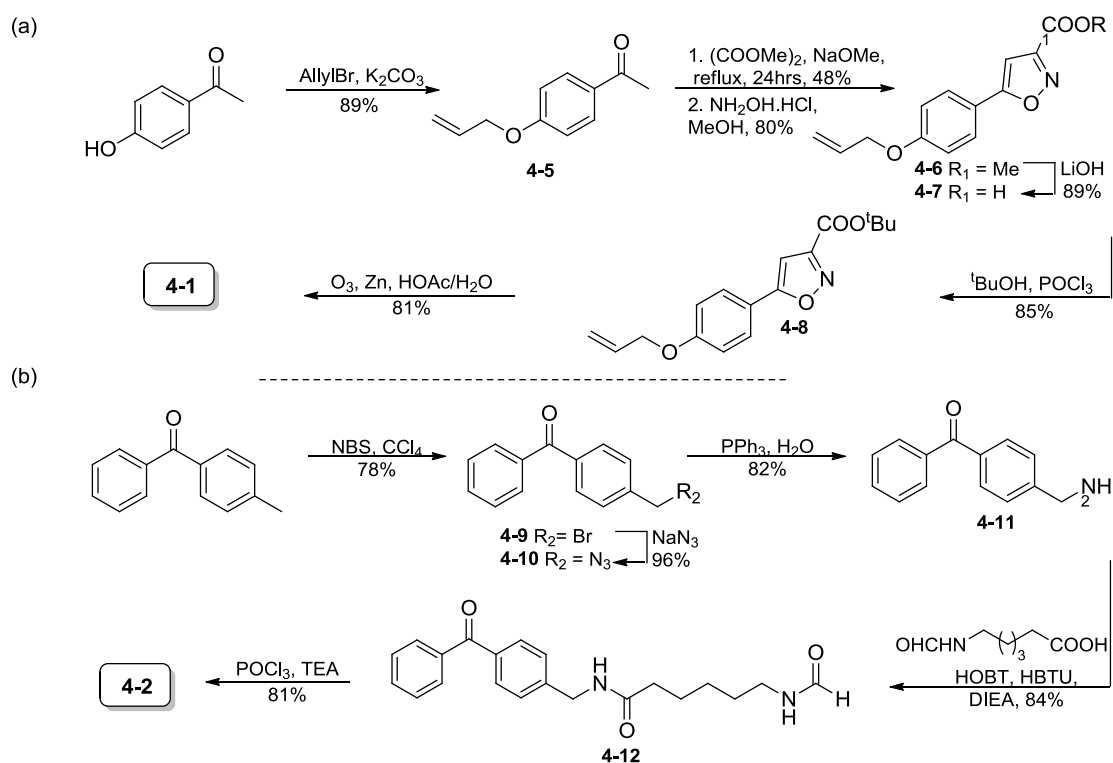
Figure 4.3 The AfBP work flow. First, the enzymes were incubated with the probe for 30 min. After photolysis under UV for 20 min in an ice bath, the mixtures were subjected to the [3+2] click reaction for 2 h with TER-N₃. Finally, the labeled reaction was separated by SDS-PAGE and visualized by in-gel fluorescence scanning. For pull-down/LC-MS/MS, Biotin-N₃ was used in place of TER-N₃.

4.2 Synthesis of two components-aldehyde and isonitrile

The warhead aldehyde part (Scheme 4.1a, up) was prepared from *p*-hydroxyacetophenone, which underwent allylation to get protected phenol **4-5**. Subsequently, Claisen condensation with dimethyl malonate was processed to get the diketo product in 48% yield and following by cyclization to yield the isoxazole carboxylic methyl ester in 80% yield. Next methyl group was removed by LiOH to give compound **4-7** and protected *t*-butyl group using POCl₃ and pyridine to obtain compound **4-8** in 75% yield of two steps. The final aldehyde **4-1** was obtained in 81% yield after ozonolysis. The synthesis of the photo-cross linker part (Scheme 4.1b) began from 4-methylbenzophenone. Bromination of the starting material was resulted bromomethyl benzophenone **4-9** in reasonable yield (78%). Substitution with NaN₃, following with reduction using PPh₃ was obtained the amine (**4-11**). Subsequently, **4-11** was coupled with the formamide linker, following by dehydration using POCl₃ and triethylamine to convert the final isonitrile **4-12** in

good yield (84%).¹⁰⁶ In addition, we used 25 amines (**4a-y**) including aromatic amine, aliphatic amine with aromatic ring and hetero ring (Figure 4.2).

Next, as shown in Figure 4.2, the four-compound Ugi reaction with **4-1**, **4-2**, **4-3** and one of the 25 amines (**4a-y**) was carried out in 1.5 mL tubes; upon optimizations, it was found the MCR reaction carried out with a co-solvent system of methanol and DMF (6:1) at room temperature for 9 h was sufficient to deliver high-quality products (**P1'** to **P25'**) with consistent yields (~ 20-30% after HPLC purification). The desired probes, **P1** to **P25**, in acceptable purity (> 85%) were obtained by further treatment with TFA/DCM (1:1) for 2 h and used directly for subsequent experiments. All the purification results were shown in Chapter 6.



Scheme 4.1 Chemical synthesis of **4-1** and **4-2**.

4.3 Biological experiments and discussion

4.3.1 Inhibition assay

Prior to protein labeling studies, the probes were evaluated for their inhibition against recombinant PTP1B and MptpB in a standard enzymatic assay. The inhibition of MCR products to against protein tyrosine phosphatases (PTPs) was assessed by measuring the rate of hydrolysis of the fluorogenic substrate, 6,8-difluoromethylumbellifery phosphate (DiFMUP, Invitrogen, USA) in 25 μL reaction volumes in black polypropylene flat-bottom 384-well microtiter plates (Greiner, Germany), using PTP1B and MptpB as model proteins. For IC_{50} studies, dose-dependent reactions were set up by varying the concentration of each inhibitor while maintaining a fixed enzyme and substrate concentration. Briefly, a two-fold dilution series of an inhibitor, from approximately 400 μM to 3.125 μM (final concentrations) was prepared. The reaction conditions are shown below:

PTP1B (2 $\mu\text{g}/\text{mL}$) = 10 μL

DiFMUP (10 μM) = 10 μL

Inhibitor (Varied) = 5 μL (in 40% DMSO)

Assay buffer (PTP1B) = 25 mM HEPES, 150 mM NaCl, 0.1 mg/mL BSA pH=7.5

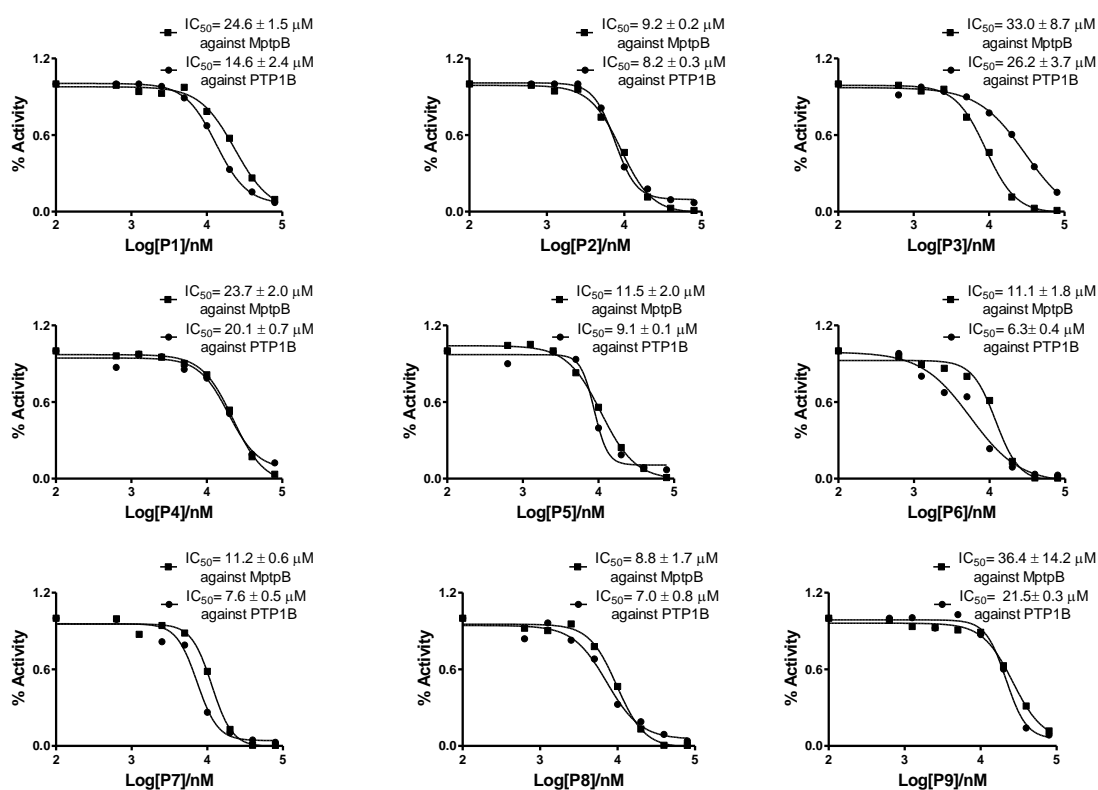
Assay buffer (MptpB) = 25 mM HEPES, 2.5 mM EDTA, 50 mM NaCl, 0.02%

Triton-100, 2 mM DTT, pH = 7.4

Negative controls were performed in the absence of enzyme and positive controls were carried out in the presence of enzyme with DMSO (i.e. without inhibitor). The reactions were allowed to incubate at room temperature for 30 min before being initiated by addition of DiFMUP. The enzymatic reactions were immediately monitored with a SynergyTM 4 Multi-Mode Microplate Reader (BioTek, USA), at λ_{ex} = 355 nm and λ_{em} = 460 nm for a period of 15 min. Each IC_{50} plot was

generated by averaging duplicates from two independent assays (Figure 4.4) and IC_{50} values were summarized in Table 4.1.

Most A_f BP s had IC_{50} values of $\sim 10 \mu M$ against PTP1B, which was similar to the $5.7 \mu M$ (in K_i value) obtained from the original bidentate inhibitor (Figure 4.1),^{77b} indicating that they likely bound to PTP1B at both the primary and secondary binding sites. Most probes possessed a greater inhibition against PTP1B than MptpB, indicating the isoxazole carboxylic acid moiety chosen fitted the active site of PTP1B better.^{77b, 87a} These results confirmed that most of our 25 A_f BP s had moderate affinity against both PTP1B and MptpB, and therefore may be used for subsequent substrate specificity profiling/proteomic pull-down experiments



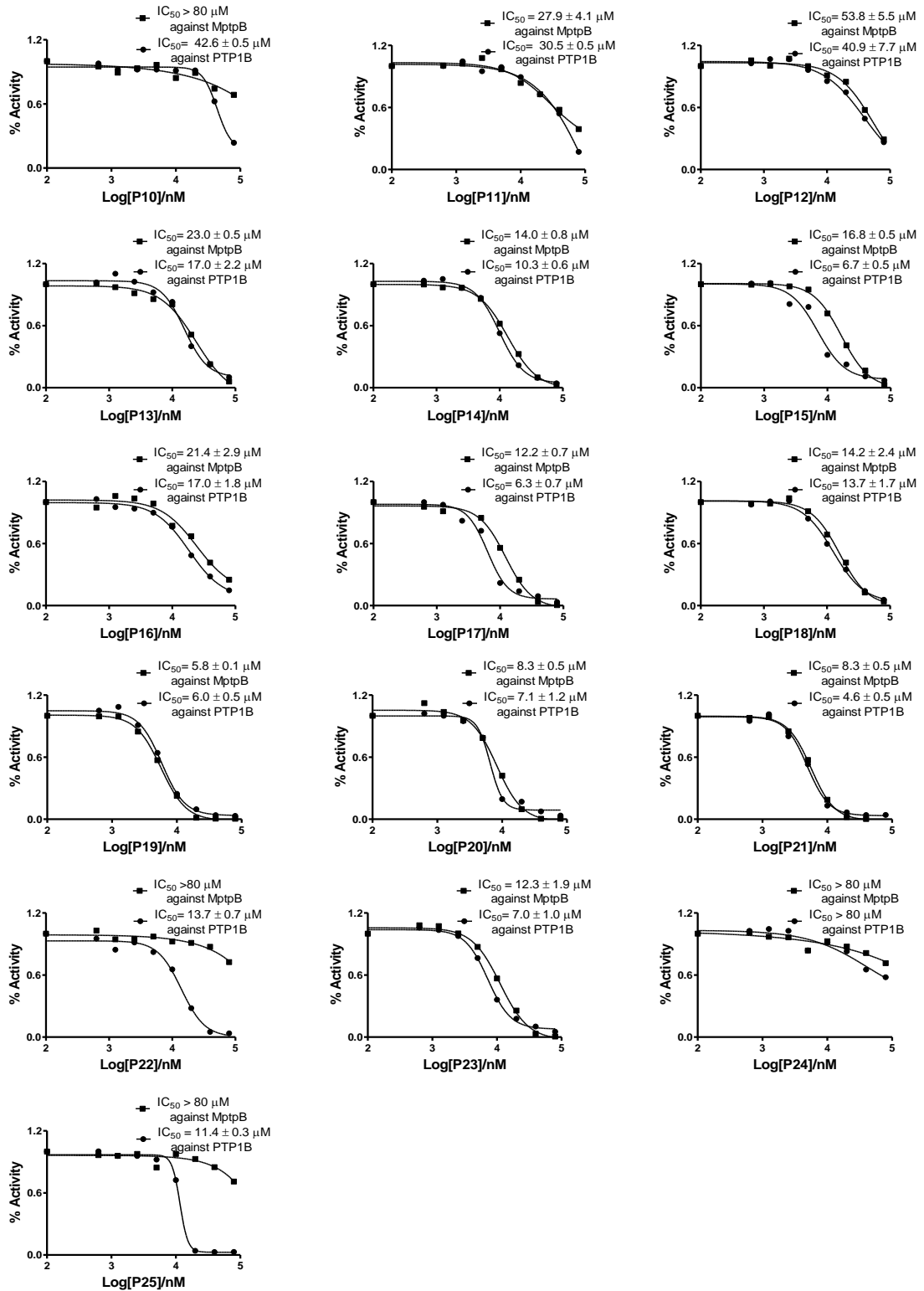


Figure 4.4 IC₅₀ graphs of all 25 A β /BPs against PTP1B and MptpB

Table 4.1 Summary of IC₅₀ values of 25 A/BPs against PTP1B/MptpB

Probe ID	IC ₅₀ (μM)		Probe ID	IC ₅₀ (μM)	
	PTP1B	MptpB		PTP1B	MptpB
P1	14.6 ± 2.4	24.6 ± 1.5	P14	10.3 ± 0.6	14.0 ± 0.8
P2	8.2 ± 0.3	9.2 ± 0.2	P15	6.7 ± 0.5	16.8 ± 0.5
P3	26.2 ± 3.7	33.0 ± 8.7	P16	17.0 ± 1.8	21.4 ± 2.9
P4	20.1 ± 0.7	23.7 ± 2.0	P17	6.3 ± 0.7	12.2 ± 0.7
P5	9.1 ± 0.1	11.5 ± 2.0	P18	13.7 ± 1.7	14.2 ± 2.4
P6	6.3 ± 0.4	11.1 ± 1.8	P19	6.0 ± 0.5	5.8 ± 0.1
P7	7.6 ± 0.5	11.2 ± 0.6	P20	7.1 ± 1.2	8.3 ± 0.5
P8	7.0 ± 0.8	8.8 ± 1.7	P21	4.6 ± 0.5	5.3 ± 0.5
P9	21.5 ± 0.3	36.4 ± 14.2	P22	13.7 ± 0.7	> 80
P10	42.6 ± 0.5	> 80	P23	7.0 ± 1.0	12.3 ± 1.9
P11	30.5 ± 0.5	27.9 ± 4.1	P24	> 80	> 80
P12	40.9 ± 7.7	53.8 ± 5.5	P25	11.4 ± 0.3	> 80
P13	17.0 ± 2.2	23.0 ± 0.5			

4.4.2 Fingerprint profiling of PTP1B and MptpB

4.4.2.1 Expression and purification of PTPs

PTP1B and MptpB containing His₆-tag were expressed in *E.coli* strain BL21-DE3, as previously described, and purified from the lysates on Nickel-nitrilotriacetic acid (Ni-NTA) metal-affinity chromatography matrices (Qiagen) according to the manufacturer's instructions. The purified proteins were then dialyzed against dialysis buffer and stored in 30% glycerol at -20 °C before use. (Dialysis buffer: PTP1B: 10 mM Tris·HCl, 25 mM NaCl, pH = 7.5; MptpB: 50 mM Tris·HCl, 100 mM NaCl, pH = 7.4). Stock solutions of enzymes were prepared in final concentrations of 2-5 mg/mL (in dialysis buffer). Stock solutions of the probes were prepared in DMSO and stored at before use. UV photolysis experiments were carried out using a B100A UV

lamp (UVP, USA). Fluorescence imaging was performed using a Typhoon 9410 fluorescence gel scanner at $\lambda = 533$ nm and analyzed with the ImageQuant Software.

4.3.2.2 General optimization for enzyme labeling

The enzyme stock solutions were diluted to 1 mg/mL. Generally 1 μ L enzyme solution was added into 14 μ L dialysis buffer solution with 1 μ L probe and shaken at room temperature in the dark for 30 min. The mixture was next irradiated under the long-rang wavelength UV channel for 20 min on the ice. Subsequently, 1 μ L of rhodamine-N₃ and the click reagents (total 3 μ L; see reference 107) were added into the solution (final solution contained CuSO₄ (500 μ M), TBTA liagnd (100 μ M) and the reducing reagent (PTP1B: sodium ascorbate 250 μ M; MptpB: DTT 500 μ M)). The whole solution (total 20 μ L) was shaken at room temperature for 2 h. The reaction was then quenched by addition of 4 μ L of 6x SDS loading dye followed by boiling at 95 °C for 10 min. The samples were analyzed on a 10% denaturing SDS-PAGE gel. The fluorecence was detected with the fluorecence gel scanner.

To determine the concentration-dependent labeling with varied amounts of the enzyme, 1 μ L of MptpB (2000 ng, 1000 ng, 500 ng, 250 ng) was added into different reaction vessels, each containing 5 μ M of the probe **P15**. The samples were then treated as mentioned above. To determine the concentration-dependent labeling with varied amounts of the probe **P15**, 1 μ g of MptpB was incubated with the probe **P15** (20, 10, 5, 2.5, 1.25, 0.6, 0.3, 0 μ M). The reaction was incubated for 30 min and treated as mentioned above. For UV- irradiation time-dependent experiments, identical reaction mixtures containing 1 μ L of MptpB and 5 μ M of the probe **P15** were similarly set up. The reactions were incubated at room temperature in the dark for 30 min. The mixtures were then irradiated for varied lengths of time (10, 15, 20,

30, 60 min) with long-wavelength UV light. Then the mixture was analyzed as mentioned above. As shown in Figure 4.5a and 4.5b, the labeling intensity increased with increasing probe or protein concentration. And with more UV-irradiation time which generates cross-link reaction, the intensity is also higher.

For labeling experiments with different click reagents and click reaction time, the reactions were prepared as mentioned above by using the probe **P15**. Varied amounts of CuSO₄ (100, 200, 500 μM) were used. Then the mixture was treated as described above. Probes **P12** and **P17** were used to test the optimized amount of the ligand (TBTA, tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine) in the labeling of MptpB. Probe **P15** was used to test the difference in the reducing agent (i.e. DTT vs sodium ascorbate) in the labeling reaction. Different amounts of sodium ascorbate (250, 500, 1000 μM) and DTT (125, 250, 500, 1000 μM) were used. All other conditions were similar as above mentioned. To optimize the click reaction time, after addition of the click reagent as mentioned above, the reaction was further incubated for 0.5, 1, 1.5, 2, and 3 h. After analyzing the results, we used higher copper concentration, with ligand and two different reducing reagents to label recombinant PTP1B and MptpB.

For heat-denaturing experiments, MptpB (in Tris buffer) was heated at 95 °C for 10 min and allowed to cool to room temperature. Each reaction was set up using 1000 ng denatured and normal MptpB. Different probes (**P13**, **P14**, **P19** and **P20**) were used. Other procedures were the same as above described. The labeling profiles of both the active and heat-denatured MptpB were compared (Figure 4.5h), confirming that successful labeling of PTPs by these AfBPs were indeed dependent upon the integrity of the enzyme active site.

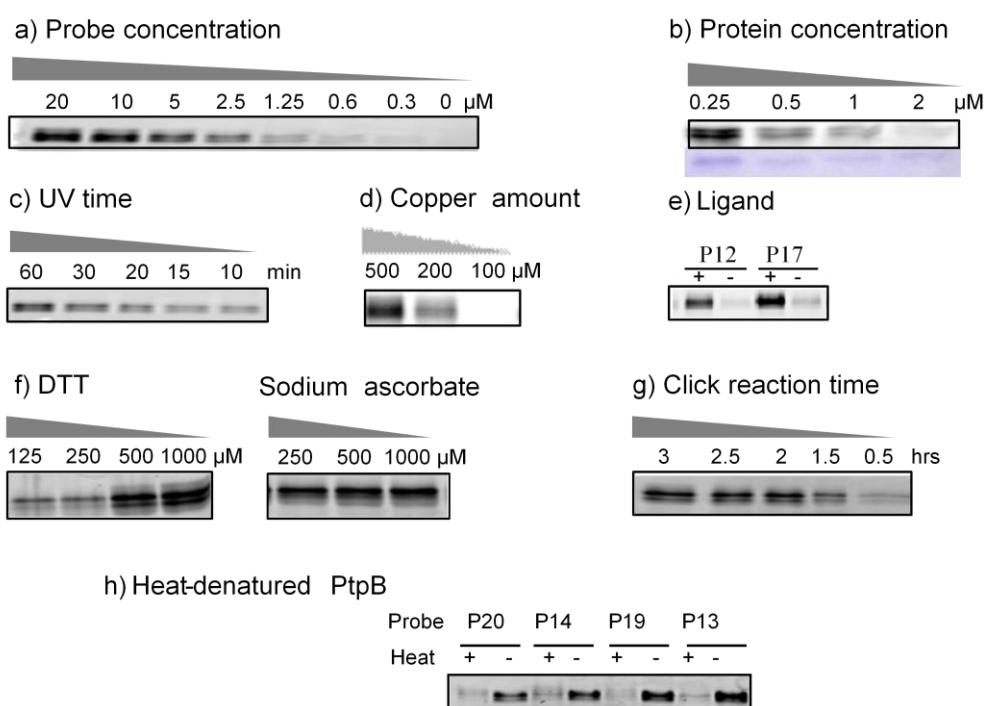


Figure 4.5 Optimization of labeling of pure PTPs. a) Effect of probe **P15** concentration on labeling intensity: MptpB was incubated with decreasing concentrations of **P15** (final concentration: 20, 10, 5, 2.5, 1.25, 0.6, 0.3, and 0 μM , respectively); b) Effect of protein concentrations on labeling intensity: different MptpB concentrations (final concentration: 2, 1, 0.5, and 0.25 μg , respectively); c) Effect of UV irradiation time on labeling intensity: MptpB was exposed to long-wavelength UV light for varying amounts of time (60, 30, 20, 15, 10 min, respectively); d) Effect of copper amount on labeling intensity: click reaction was carried out with increasing amounts of CuSO_4 (final concentration: 100, 200, and 500 μM); e) Effect of TBTA ligand on labeling intensity: two different probes were incubated with or without the TBTA ligand (final concentration: 125 μM); f) Effect of reducing reagents on labeling intensity: MptpB was incubated with different concentrations of DTT (final concentration: 125, 250, 500, and 1000 μM , respectively); PTP1B was incubated with different concentrations of sodium ascorbate (final concentration: 250, 500, and 1000 μM); g) Effect of click time on the labeling intensity: click reactions were carried with increasing time (3, 2.5, 2, 1.5 and 0.5 h, respectively); h) Different probes (**P13**, **P14**, **P19**, and **P20**; final probe concentration: 5 μM) against heat-denatured and native MptpB.

4.3.2.3 Affinity-based profiling of MptpB/PTP1B by using 25 AfBPs

After carrying out ABPP using recombinant PTP1B and MptpB by optimizing the protein labeling conditions (see above); it was found that 1 μg of the protein, incubated with 5 μM of the probe for 30 min, followed by 20 min of UV irradiation then 2 h of click chemistry with the reporter tag, gave the best labeling results. The labeling procedure was the same as mentioned above. One of the key advantages of our current approach is the rapid access to a large number of potential AfBPs using Ugi reaction, and in the present study, the 25 AfBPs obtained were immediately suitable for substrate specificity profiling against PTPs. As shown in Figure 4.6, the “fingerprint” generated for PTP1B and MptpB against the panel of AfBPs were unique, and may be used to further delineate different binding pockets present in PTPs. In the case of PTP1B, it appeared that probes having a small, hydrophobic diversity group (**P1**, **P16**, **P17**, **P23**, **P24** and **P25**) gave the strongest fluorescence labeling. This is in good agreement with previously established structural information of PTP1B binding pockets.⁸⁸



Figure 4.6 Affinity-based profiling of recombinant PTP1B and MptpB against the 25 A/BPs. The fluorescently labeled band was pseudo-colored in grey. PTP1B: protein 1 μg ; probe 5 μM ; dye 40 μM ; Copper sulfate 500 μM ; ligand 100 μM ; sodium ascorbate 500 μM ; UV time 20 min; click time 2 h. MptpB: protein 1 μg ; probe 5 μM ; dye 40 μM ; Copper sulfate 500 μM ; ligand 100 μM ; DTT 500 μM ; UV time 20 min; click time 2 h.

4.3.4 *In vitro* labeling mammalian proteomes and pull-down/LC-MS/MS

analysis

We next performed proteome profiling of MCF-7 mammalian cell lysates and large-scale pull-down/LC-MS/MS target identification experiments using the **P23** probe.¹⁰⁷ **P23** was chosen because of its good inhibition and strong labeling against PTP1B. Briefly, **P23** (20 μ M) was incubated with MCF-7 cell lysates (10 mg/10 mL in a chilled buffer of 50 mM Tris·HCl, 150 mM NaCl, pH 8.0, 1.0% NP-40, 100 μ M PMSF) for 20 min at room temperature. Then the samples were irradiated for 45 min under UV light. Subsequently, a small amount of the irradiated sample was treated with rhodamine-N₃ (100 μ M) and click reagent (1 mM CuSO₄, 1 mM TCEP, 100 μ M TBTA) for 3 h. Following acetone precipitation, washing (2x with methanol), resolubilization (in 1x SDS loading buffer) and heating (10 min at 95 °C), the sample was separated by SDS-PAGE (10% gel) and visualized by in-gel fluorescence scanning on a typhoon 9410 variable mode imager scanner. The fluorescent gel was shown in Figure 4.7. As shown in Figure 4.7a, in-gel fluorescence scanning of the labeled lysates indicated multiple cellular targets, one of which was subsequently confirmed to be endogenous PTP1B by western blot (labeled with “*” in Figure 4.7b) and LC-MS/MS analysis (Figure 4.7c)

The remaining labeled lysates (prior to click chemistry) were used for large-scale pull-down/LC-MS analysis to identify other cellular targets of the probe. Briefly, the sample was treated with Biotin-N₃ (100 μ M) and click reagent (1 mM CuSO₄, 1 mM TCEP, 100 μ M TBTA) for 3 h. Subsequently, it was acetone-precipitated, washed (2x methanol) and resolubilized in PBS (containing 0.1% SDS) with sonication. The sample was then incubated with avidin-agarose beads (Thermo Scientific) for 4 h at room temperature. After centrifugation, the supernatant was

removed and the beads were washed with washing buffers (3x with 1 M NaCl, 20 mM Tris·HCl, 5 mM EDTA, 0.1% NP-40, pH = 7.3; 6x with 2 mM Tris·HCl, 0.5 mM EDTA, 0.1% NP-40, pH = 7.3; 6x with 4 M urea, 10 mM Tris·HCl, 1 mM EDTA, 0.1% NP-40, pH = 7.3; 2x with 2 mM Tris·HCl, 0.5 mM EDTA, pH = 7.3). The beads were then boiled in 1x SDS loading buffer for 10 min. Samples were separated by SDS-PAGE gels and stained by colloidal blue solution. Gel lanes corresponding to both DMSO- and probe-treated samples were then cut into small slices, respectively. Next, trypsin digestion (In-Gel Trypsin Digestion Kit, Pierce Co., USA) was performed for the whole pull-down proteins. Digested peptides were then extracted from the gel with 50% acetonitrile and 1% formic acid in H₂O. Two samples was then dried *in vacuo* and stored at -20 °C for further LCMS analysis. Briefly, samples were analyzed as previously described. The samples were resuspended in 0.1% formic acid in H₂O. The peptides were separated and analyzed on a Shimadzu UFLC system (Shimadzu, Kyoto, Japan) coupled to an LTQ-FT Ultra (Thermo Electron, Germany). Peptides were then analyzed on LTQ-FT with an ADVANCE™ CaptiveSpray™ Source (Michrom BioResources, USA) and the raw data were analyzed using an in-house Mascot Server (version 2.2.07, Matrix Science, UK) with MS tolerance of 10 ppm and MS/MS tolerance of 0.8 Da. Two missed cleavage sites of trypsin were allowed. Carbamidomethylation (C) was set as a fixed modification. And oxidation (M) was set as variable modifications. After data analysis, a large number of proteins were identified from each LCMS experiments, many of which are non-specific binders and need to be removed. For those proteins with cores of < 50, they were automatically removed from the list. For those proteins that appeared in the negative run (pull-down/LCMS experiments with DMSO only), they also were excluded from the list as well. Highly abundant proteins that commonly appeared in irrelevant pull-down

experiments were further removed. The final list was shown in Chapter 6, with selected hits shown in Figure 4.7c.

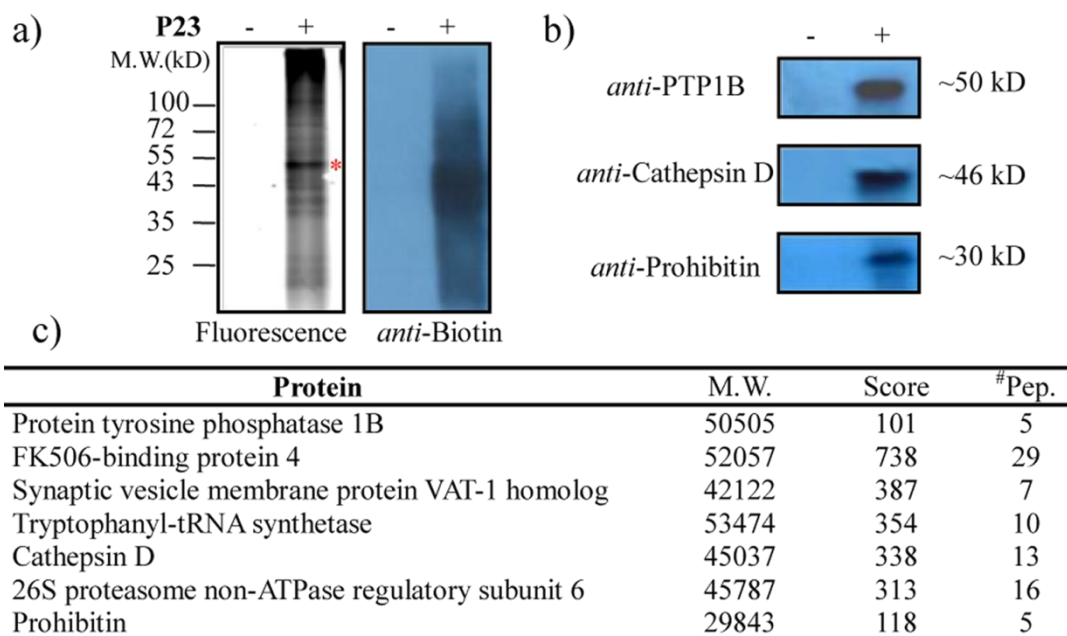


Figure 4.7 Results of proteome profiling of **P23** against MCF-7 lysate. (a) Proteome profiling of **P23** against MCF-7 cell lysates. (Left) In-gel fluorescence scanning. (Right) Western blot of the pulled-down sample after **P23** labeling and clicked with Biotin-N₃. (-) Netegative controls with DMSO. (*) labeled PTP1B which was subsequently confirmed by WB (in Figure 4.6b). (b) Target validation by pull-down/WB experiments. (c) A list of high-scoring proteins identified from *in vitro* pull-down/LC-MS experiments. For complete list, see Chapter 6, section 6.11

In addition to PTP1B, several other high-scoring proteins were listed (for complete list, see Chapter 6.11). Some of them (Cathepsin D and Prohibitin) were further validated to be true targets of **P23** by pull-down/WB analysis (Figure 4.6b). Given the fact that most of our probes were moderate binders of PTP1B, it was not surprising our A/BPs were also successful in labeling these “off-target” proteins from

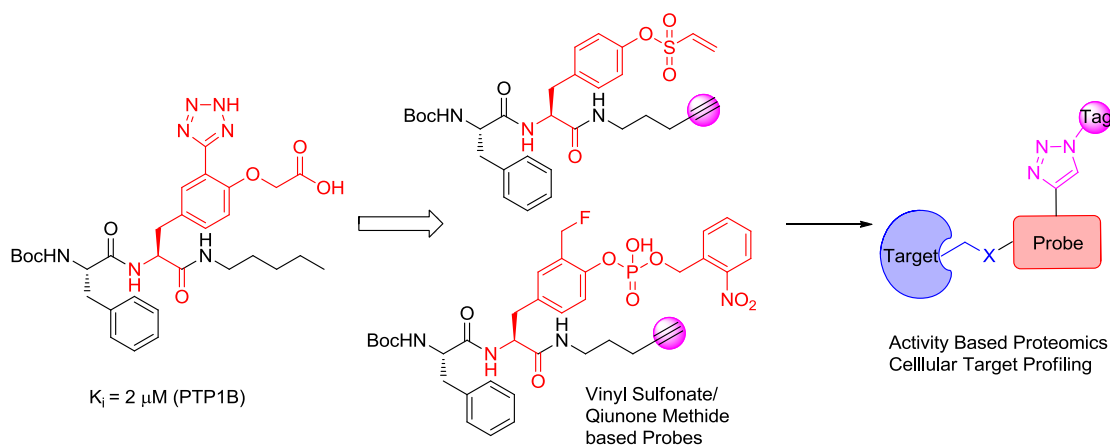
MCF lysates.^{61g} Information obtained from our current study would nevertheless be helpful in the design of more potent and specific AfBPs against PTPs in future.

4.4 Conclusion

In summary, we have successfully used the multi-component Ugi reaction for rapid assembly of a panel of affinity-based probes targeting potential PTPs. The probes obtained, though moderate binders of PTP1B, were able to successfully label endogenous PTP1B in mammalian lysates. From this study, we have also identified some “off-targets” of these probes which will help in future design of better PTP probes. The marriage between MCR and ABPP will add another useful chemical tool to the emerging field of Catalomics.¹⁰⁸

Chapter 5 Unexpected phosphotyrosine mimicking probes target protein disulfide isomerase in different cell lines

Summary



In this chapter, four clickable, phosphotyrosine mimics-, activity-based probes, which were derived from a known PTP1B inhibitor, have been synthesized and used *in situ* for profiling and identification of cellular targets in cell lines. Unexpectedly, protein disulfide isomerase was found as a major and selective target.

5.1 Introduction

Protein tyrosine phosphates (PTPs), together with protein tyrosine kinases (PTKs) plays an essential role in maintaining the *pTyr* level in cellular events and regulating various fundamental cellular processes.² Malfunction in PTP activity is related with a variety of human diseases, including cancer, diabetes and obesity, and autoimmune disorders.⁴ Many powerful chemical methods have been developed to integrate the analysis of PTPs in whole proteome.¹⁰ In particular, activity based probe

profiling (ABPP) is a very useful tool and have been applied in drug discovery, target identification and biomarker discovery.^{51, 109} ABPs for PTPs have already been developed based on PTP catalytic mechanism.¹¹⁰ However, due to the high conserved active site of PTPs, targeting an individual member of PTPs in crude proteome still needs to be addressed.

Herein, we designed and synthesized four activity-based probes based on a known PTP1B inhibitor (**1-19**, Figure 5.1)⁴¹ by replacing the phosphotyrosine (*p*Tyr) mimetic head to some reactive groups, including phenyl vinyl sulfonates (PVS, **5-1** and **5-2**) and 2-fluoromethyl phenyl phosphate (FMPP, **5-3** and **5-4**). FMPP moiety generates a highly reactive quinone methide intermediate upon PTP hydrolysis and subsequently immobilizes itself covalently to the nearby proteins.^{50, 90, 111} PVS, developed by Zhang group⁶⁵, as a new *p*Tyr mimic with better cell permeability, is an active site-directed and irreversible inactivation compound toward PTPs through 1,4-Michael addition of cysteine in active site. Furthermore, the probe scaffold contains an alkyne handle, which serves a clickable tag for subsequent visualization, enrichment and identification of targets by conjugation to suitable reporters (trifunctional tag^{54c, 112}) *via* the bioorthogonal click chemistry. Under consideration, since the aliphatic chain of **1-19** is out of the binding pocket and there is no hydrogen bond between the chain and protein, the minimal structural modification was established by introducing a C-C triple bond through replacing the terminal C-C single bond of this aliphatic chain. Furthermore, a photolabile protecting group was introduced into FMPP based probe **5-3** and **5-4** to improve the cell permeability and also provide temporal control by photolysis. Besides, compared with probes **5-1** and **5-3**, probes **5-2** and **5-4** were designed with the coumarin-based aromatic ring to further mimic *p*Tyr.

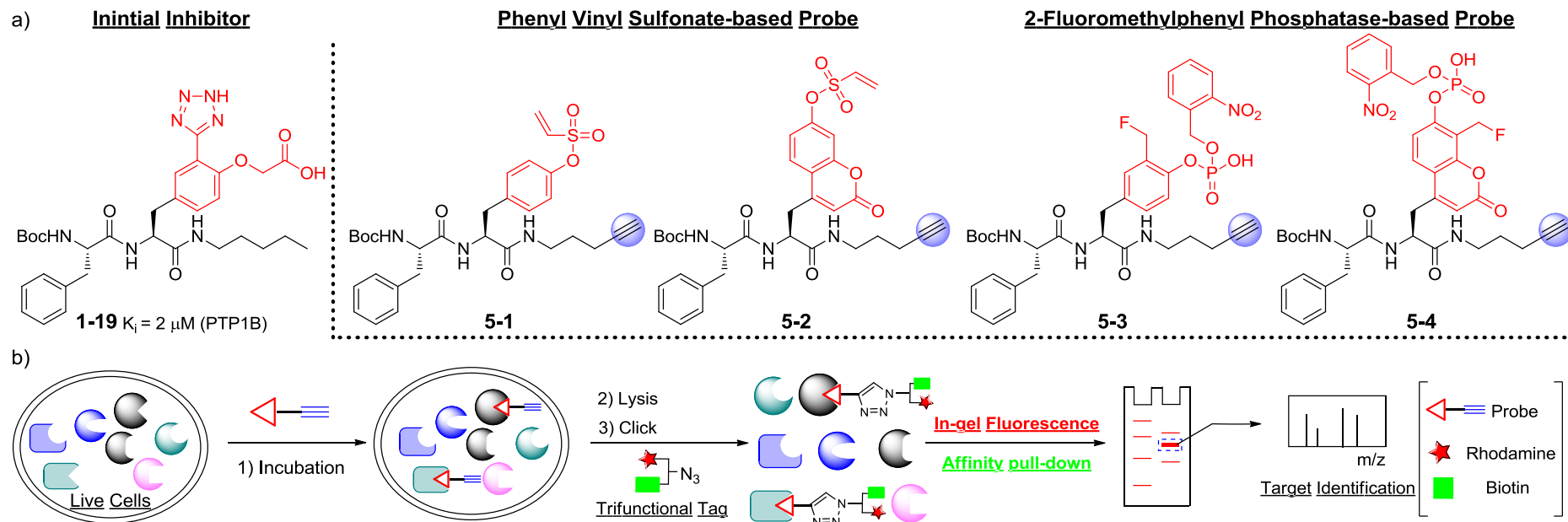


Figure 5.1 Overall strategy of probe design. a) Structures of PTP1B inhibitor **1-19** and four **1-19**-derived clickable probes (PVS-based probes **5-1** and **5-2**, as well as 2-FMPP based probes **5-3** and **5-4**). b) Workflow for *in situ* targeting profiling using probes. Protein visualization and identification is achieved by click chemistry using trifunctional tag.

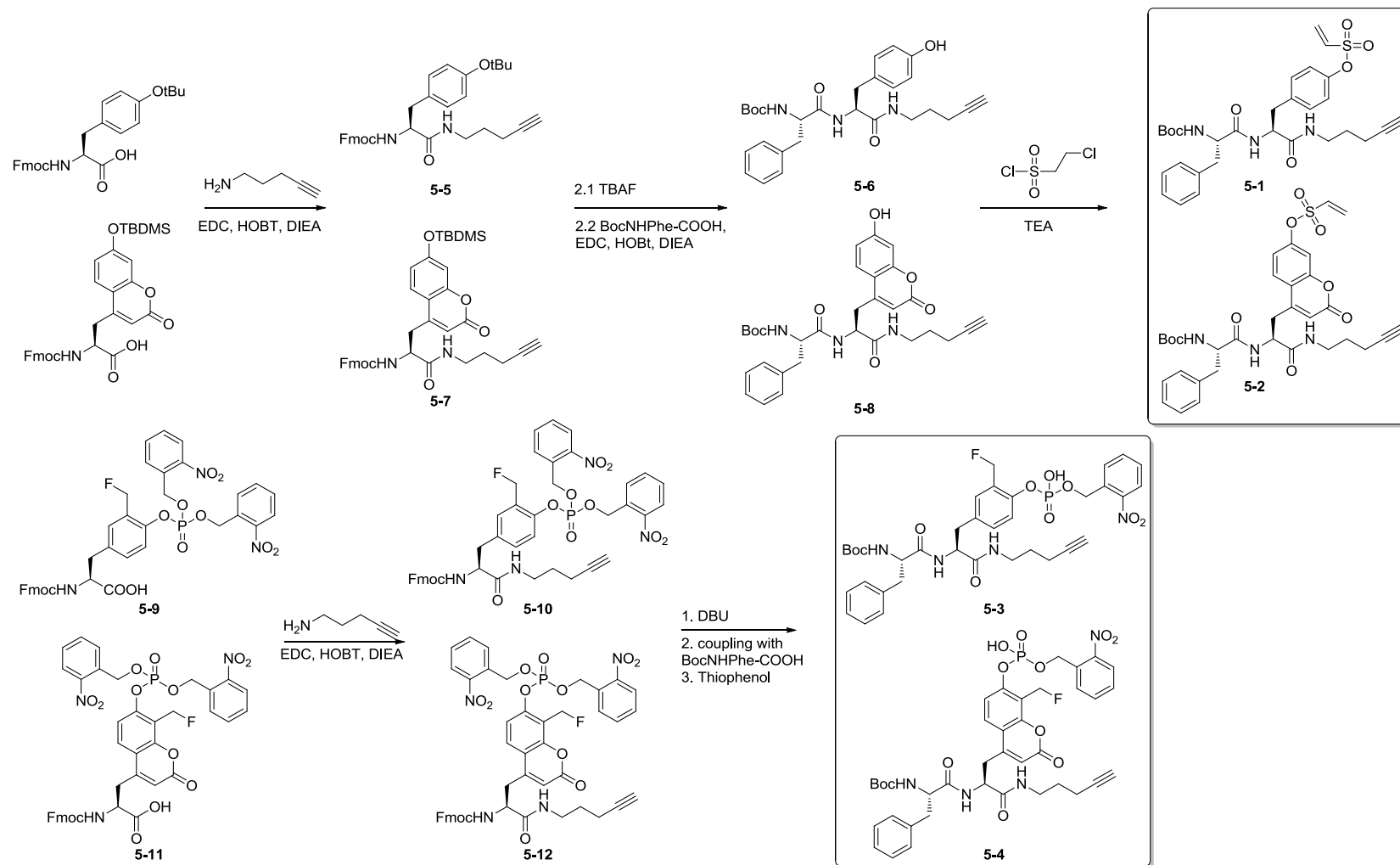


Figure 5.2 Synthetic schemes of four probes.

5.2 Synthesis of four probes

The synthetic scheme is shown in Figure 5.2. In brief, the aliphatic chain with the alkyne modification part was attached by amine and carboxylic acid coupling reaction using EDC, HOBT activation. The deprotection of Fmoc group was achieved using either piperidine or DBU. Subsequently, the addition of Boc-protected phenylaniline was also obtained by coupling reaction using EDC, HOBT activation. For probe **5-1** and **5-2**, the vinyl sulfonate modification on phenol was obtained by nucleophilic substitution and elimination using 2-Chloro-1-ethanesulfonylchloride and triethylamine as a base.⁶⁵ Probes **5-3** and **5-4** which are smaller than **5-10** and **5-12** but also have better cell bioavailability were obtained by removing one of 2-nitrobenzyl groups in **5-10** and **5-12** under TEA/thiophenol condition.

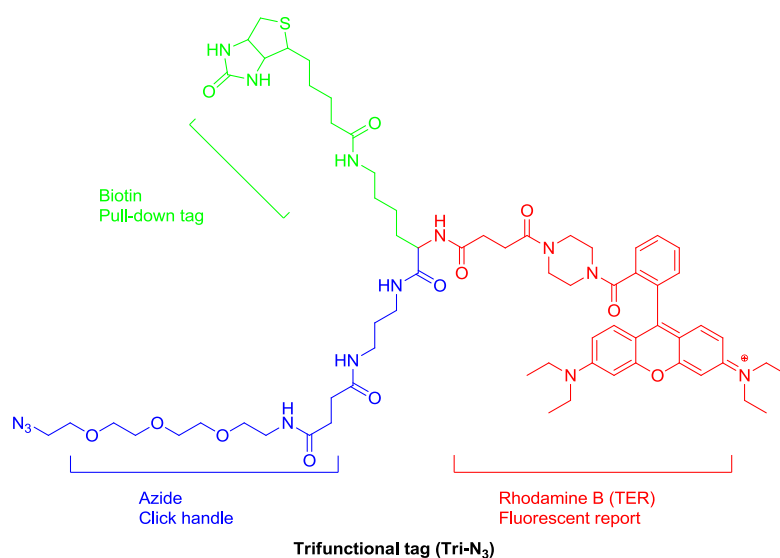


Figure 5.3 The structure of Tri- N_3 .

The trifunctional tag (Figure 5.3) was also synthesized through routine acid/amine coupling reactions. The tag was containing three elements: 1) PEG azide, which is used for click reaction by reacting with the alkyne; 2) Rhodamine B dye,

which could be used for in-gel visualization; 2) Biotin, which is used for isolation/pull-down experiments. Combining with Rhodamine B dye and Biotin together, tracking the pull-down experiment could be done using fluorescent scanning which is faster and easier than Western blotting analysis.

5.3 Biological experiments and discussion

5.3.1 *In situ* proteomic profiling and in-gel fluorescence scanning

With the four probes in hand, we next assessed whether probes could be an effective probe in live cells. To analyze the cellular target profiling, we incubated probes with living MCF-7 cancer cells following the common procedure (see the flow in Figure 5.1).

Briefly, probes (**5-1** and **5-2** in 5 μ M, **5-3** and **5-4** in 20 μ M, **VS-1**¹¹⁴ as a comparison) were directly treated with cell medium. Final DMSO concentration in labeling is less than 0.1%. The probes were stocked in DMSO, -20 $^{\circ}$ C. For **5-1** and **5-2**, cells were incubated for 2 h incubation. For **5-3** and **5-4**, cells were first incubated with probes for 2 h, allowing sufficient time for them to enter the cells. Subsequently these cells were UV-irradiated for uncaging of the 2-nitrobenzyl group then further incubated for 2 h.

After incubation, the cells were washed with PBS three times and detached from the plates by scrapping or trypsin. Afterward, the cells were lysed in 300 μ L buffer (50 mM Tris-HCl, 150 mM NaCl, 1% NP-40, 100 μ M PMSF, pH = 8.0). To initial the click reaction, 4 equivalents of TER-azide (structure shown in Figure 4.3), 4 equivalents of TBTA, 40 equivalents of TCEP and 40 equivalents of CuSO₄ was added. The click reagent mixture was freshly prepared. The reaction was incubated at room temperature for 2 h with gentle shaking before termination by the addition of 5-

fold volume prechilled acetone (incubation for at least 1 h at -20 °C). Precipitated proteins were subsequently collected by centrifugation (13,000 rpm x 10 min at 4 °C). The supernatant was discarded and the residue pellet was washed with prechilled methanol, air-dried until the pellet started to shrink. Then the residue was resuspended in 1x standard SDS-loading buffer, sonicated 10 min and heated for 10 min at 95 °C with gentle mixing. Finally, the protein sample was loaded into 10% or 12% SDS-PAGE gel, separated following by in-gel fluorescence scanning with a Typhoon 9410 variable mode imaging scanner (GE Amersham).

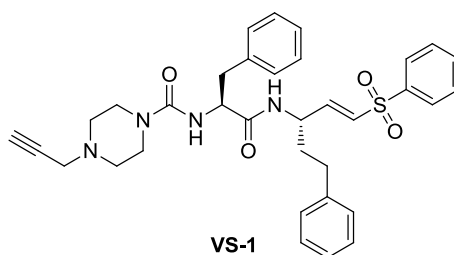
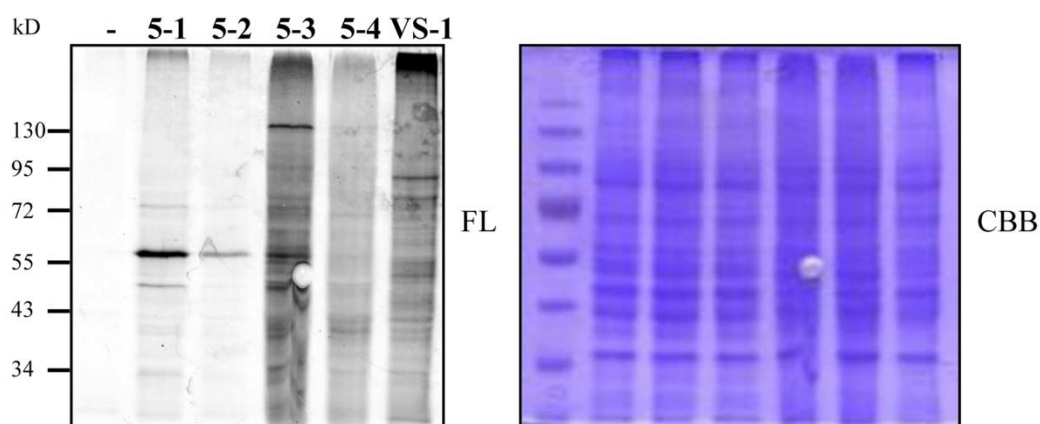


Figure 5.4 *In situ* proteome-profiling of different probes against MCF-7 (**5-1** and **5-2**: 5 μ M; **5-3** and **5-4**: 20 μ M; **VS-1**: 10 μ M; 2 h incubation). FL = fluorescent gel. CBB = coomassie brilliant blue-stained gel. “-” means negative control with only DMSO. Below is the structure of **VS-1**.

In-gel fluorescent SDS-PAGE analysis of the labeled proteomes in Figure 5.4 showed all the probes have some distinct fluorescently labeled bands compared with blank labeling (only DMSO, without any probe). Especially, **5-1** and **5-2** gave a very strong specific protein bands at a molecular weight of higher than 55 kD. Interestingly, it appears not a highest abundant protein.

Subsequently, dose- and time- dependent *in situ* labelling were carried on by varying the concentrations of **5-1** and incubation time. The results showed the specific fluorescent intensity to ~57 kD protein and other non-specific labeling increased with increasing concentrations of the probe (Figure 5.5 left). Furthermore, the labeling did not vary much with increasing incubation time (Figure 5.5 right).

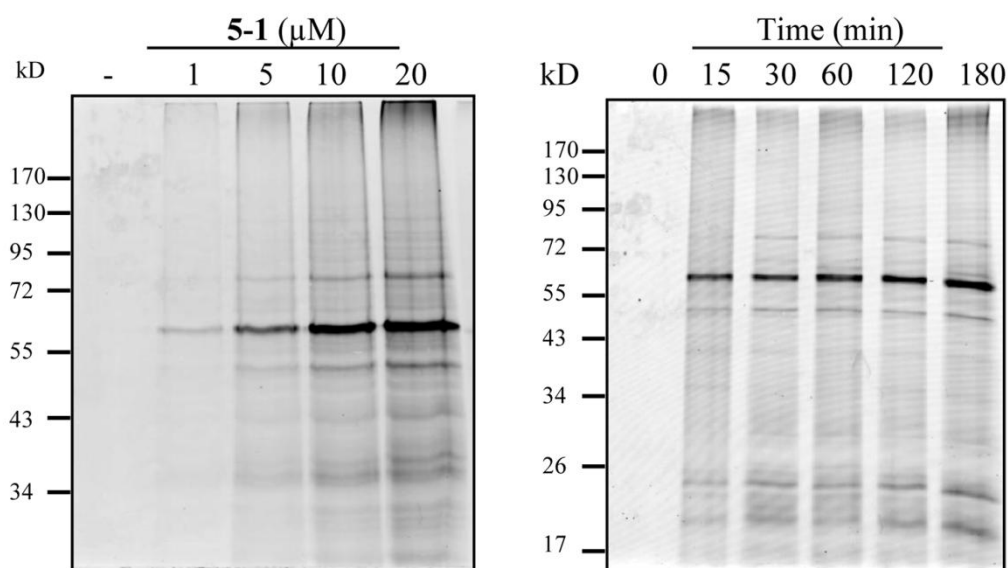


Figure 5.5 Dose and time-dependent *in situ* proteome profilings. a) Dose-dependent *in situ* proteome profiling of MCF-7 with **5-1** (different concentrations, 2 h incubation) resolved in 10% SDS gel; b) Time-dependent *in situ* proteome profiling of MCF-7 with **5-1** (10 μM, different time incubations) resolved in 12% SDS gel.

5.3.2 Co-localization bioimaging

Then we also performed the fluorescence microscopy to visualize the localization of targets of probe. Briefly, MCF-7 cells were seeded in glass bottom dishes (CELLview™, Cat. No. 627861) and grown till ~ 60% confluence. The cells were incubated with 20 μ M **5-1** in fresh growth medium (200 μ L). The cells were further incubated for 1 h. Then the cells were washed with PBS three times. Subsequently, cells were fixed with 3.7% formaldehyde in PBS for 20 min at 37 $^{\circ}$ C/CO₂, washed twice again and permeabilized with 0.1% Triton X-100 in PBS for 15 min at 37 $^{\circ}$ C/CO₂, washed twice again. Cell were then blocked with 2% BSA, 0.05% Tween-20 in PBS for 30 min at room temperature, washed twice with PBS and then subsequently treated with a freshly premixed click chemistry reaction in a 200 μ L (10 μ M TER-PEG-azide, 200 μ M TBTA 2 mM TCEP, 2 mM CuSO₄) for 1 h at room temperature with gentle shaking. Cells were washed with 2x PBS, several times PBS containing 1% Tween-20 and 0.5 mM EDTA (until there was no small crystal under microscopy), 2x PBS, 2x methanol, and 2x PBS. For ER-localization experiments, cells were further incubated with ER tracker™ Green (glibenclamide BODIPY® FL, Invitrogen, 0.3 μ M in final concentration) for 2 h, and washed with PBS twice. Continually, cells were incubated with nucleus stain (Hoechst, 0.2 μ g/mL in final concentration) for 20 min and washed twice with PBS. Finally, the cells were washed ,imaged with a Leica TCS SP5X confocal microscope system and processed with the Leica application suite advanced fluorescence (LAS AF).

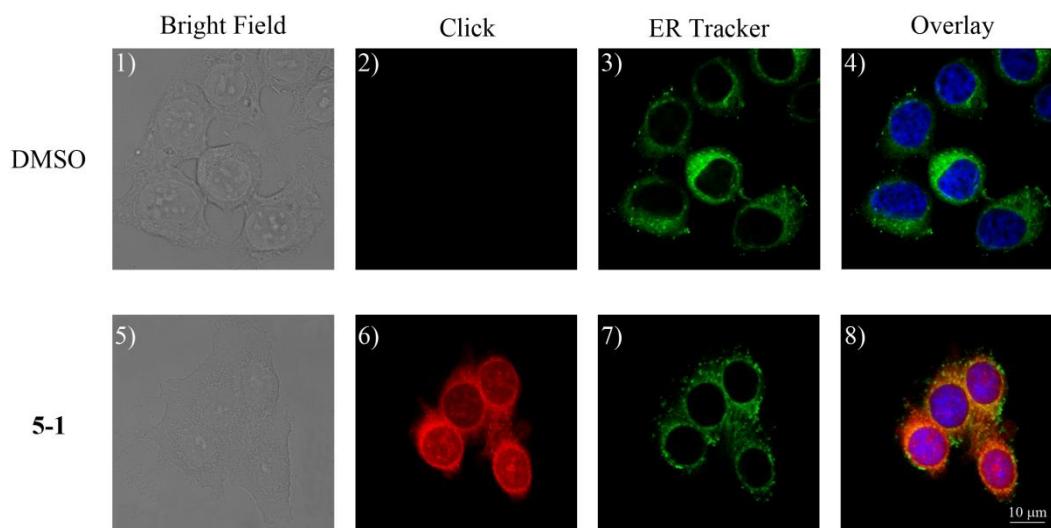


Figure 5.6 Organelle co-localization bioimaging experiments. Cellular imaging results of MCF-7 cells treated with 20 μM **5-1** or DMSO. Co-localization experiments were performed with ER and Nucleus tracker. Panel 1) and 5): Bright field images of the corresponding cells; Panel 2) and 6) Probe clicked with TER-PEG-azide 543 nm Channel (pseudocolored in red); Panel 3) and 7) ER tracker (pseudocolored in green); Panel 4) and 8) Overlay images of 1), 2), 3) and 5), 6), 7) together with nucleus stain (pseudocolored in blue). All the images were acquired under the same settings. Excitation channel: Probe channel (white laser, 543 nm), ER-tracker (white laser, 488 nm), Hoechst (405 nm diode laser).

In Figure 5.6, minimal fluorescence was observed in samples treated with only DMSO. Whereas, in **5-1** treated cells, fluorescence signals were observed throughout the whole cell excluding the membrane. Most signals were co-localized with ER which indicated our target is probably localized in ER.

5.3.3 Pull-down/LCMS/MS experiment

Subsequently, we performed large-scale proteomic analysis using **5-1** to identify potential cellular targets in MCF-7 cells by affinity pull-down/LC-MS/MS experiments, especially the specific labeled protein at ~ 57 kD. The targets were enriched and visualized using a trifunctional tag. The target-tag complexes were

isolated by using avidin beads and separated by SDS-PAGE. Through the fluorescent scanning and silver stain of the gel, 72kD, 57 kD and 48 kD bands “*” (Figure 5.7) were also precisely excised and subjected to in-gel tryptic digestion and mass spectrometry analysis.

LCMS results were obtained from above experiments with **5-1** as well as with DMSO as a negative control. For those proteins that appeared in the negative run (pull-down/LCMS experiments with DMSO only), they were excluded from the list. The result of ~ 48kD, 57 kD and 73 kD bands were shown in Table 5.1. From the data, the highest score of each bands (Figure 5.7) are protein disulfide isomerase A6 (PDIA6, 48 kD), protein disulfide isomerase A1 (PDIA1, 57 kD), protein disulfide isomerase A4 (PDIA4, 73 kD), respectively. To confirm the results of the MS, we did the western blotting experiments with pull-down sample using PDIA1 antibody (Figure 5.6, right). The results again confirmed that the protein was indeed positively and specifically labeled by **5-1**.

Table 5.1 Protein identified by mass spectrometry of three specific band (**5-1**-tagged band, 10 μ M)

Name (gene protein)	M.W.	Score	Peptide	emPAI
~ 48 kD band				
PDIA6 Protein disulfide-isomerase A6	48490	1854	72	1.46
TUBB Tubulin beta polypeptide	48135	881	38	1.72
ATP5B ATP synthase subunit beta	48083	383	27	0.91
EEF1AL3 Putative elongation factor 1-alpha-like 3	50495	277	16	0.48
PLIN3 Perilipin-3	47189	246	8	0.27
~ 57 kD band				
P4HB Protein disulfide-isomerase	57081	4780	316	22.11
PDIA3 Protein disulfide-isomerase A3 precursor	56747	287	16	1.08
ATP5A1 ATP synthase subunit alpha, mitochondrial precursor	59714	243	8	0.38

PHGDH D-3-phosphoglycerate dehydrogenase	56614	175	6	0.33
CCT8 Uncharacterized protein CCT8	59741	88	2	0.06
VISA Isoform 1 of Mitochondrial antiviral-signaling protein	56493	62	3	0.12
ABAT 4-aminobutyrate aminotransferase, mitochondrial precursor	56403	59	1	0.06
DYNC1LI1 Cytoplasmic dynein 1 light intermediate chain 1	56544	52	3	0.12
~ 73 kD band				
PDIA4 Protein disulfide-isomerase A4	73229	4700	210	4.78
HSPA5 glucose-regulated protein	72402	2072	91	2.39
HSPA9 Stress-70 protein, mitochondrial	72641	289	12	0.22
KHSRP Far upstream element-binding protein 2	73443	285	11	0.21

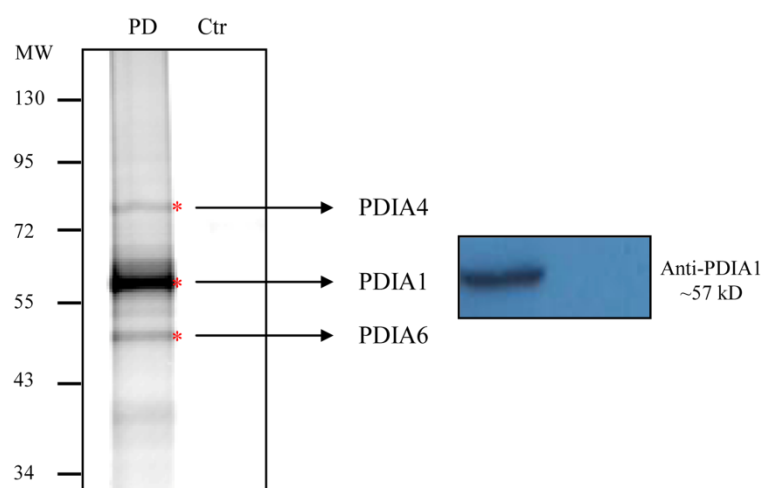


Figure 5.7 Fluorescence image of MCF-7 pull-down experiments after **5-1** (10 μ M) labeling and clicking with trifunctional probe. “*” labeled bands were precisely cut and analyzed by mass spectroscopy. PDIA1 was subsequently confirmed by WB (right). Ctr = negative pull-down assay with DMSO.

5.3.4 Immunofluorescence

Immunofluorescence (IF) was also carried out to visualize the localization of PDIA1 and the overlapping with the probe. For immunofluorescent experiments, cells were still treated with **5-1**, fixed, permeabilized and treated with TER-PEG-N3

following previously protocols (Chapter 5.3.2). Further cells were incubated with PDIA1 primary antibody (1: 200 in 2% BSA, Santa Cruz, sc-166474) for 1 h at room temperature, washed once with 2% BSA and twice with PBS, and then incubated with FITC-conjugated anti-mouse IgG secondary antibody (1:100 in 2% BSA, Santa Cruz, sc-2010), washed once with 2% BSA and twice with PBS. Finally, the cells were washed imaged with a Leica TCS SP5X confocal microscope system and processed with the Leica application suite advanced fluorescence (LAS AF).

The results are shown in Figure 5.8. DMSO treated cells has very faint signals. From Figure 5.8(8), IF signal of PDIA1 appears to mostly overlap with **5-1**. Also, the colocalization indicated, beside the presence of PDIA1 as a target of **5-1**, there are also other side targets. Together results of organelle colocalization and IF, most targets of **5-1** are distributed in ER where most of PDI family is localized. Furthermore, these results indicate that **5-1** was a suitable imaging probe to detect its cellular targets.

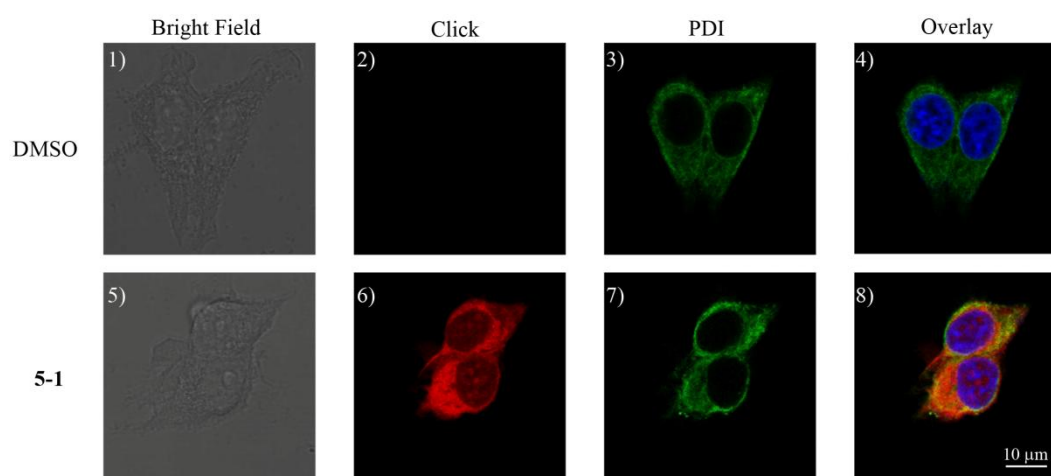


Figure 5.8 Immunofluorescence experiments were performed with anti-PDIA1 Cellular imaging results of MCF-7 cells treated with 20 μ M **5-1** or DMSO. Panel 1) and 5): Bright filed images of the corresponding cells; Panel 2) and 6) Probe clicked with TER-PEG-azide 543 nm Channel (pseudocolored in red); Panel 3) and 7) anti-PDI immunofluorescent staining

(pseudocolored in green); Panel 4) and 8) Overlay images of 1), 2), 3) and 5), 6), 7) together with nucleus stain (pseudocolored in blue). All the images were acquired under the same setting. Excitation channel: Probe channel (white laser, 543 nm), FITC-tracker (white laser, 488 nm), Hoechst (405 nm diode laser).

5.3.5 Protein disulfide isomerase

Protein disulfide isomerase (PDIA1) is one of members in PDIs family.¹¹⁴ It is a characterized 57 kD oxidoreductase that localize in the ER as a chaperone protein. It facilitates the formation, breakage and rearrangement of disulfide bonds *via* catalysis of thiol-disulfide exchange. PDI has four structural thioredoxin-like domains (a, b, b' and a') with the catalytic CGHC motif present in a and a' domains. The pK_a of cysteine in active site is very low, around 4.6, which makes it as a good nucleophile and likely reacts with our probes. PDI could also reach the cell surface from ER through a regulatory pathway, with HIV virus infection.¹¹⁵ PDIs are involved in the recognition and repairment of abnormal proteins and their expression level is increased in patients with neurological disorders, including Parkinson's disease^{116a} and amyotrophic lateral sclerosis (ALS)^{116b}. Recently, Stockwell group found that an inhibitor 16F16A could target PDIs (PDIA1 and PDIA3) and inhibit PDI in rat brain cells. And they also showed that inhibiting PDI activity in normal rat brain cells suppresses the toxicity of misfolded huntingtin and of A β peptides processed from the amyloid precursor protein.^{116c}

5.4.5.1 IC₅₀-determination of **5-1** against protein disulfide isomerase (PDI)

In order to confirm the specific reactivity of PDI with **5-1**, we did the dose-dependency of **5-1** on PDI reductase activity using the insulin turbidometry assay. The inhibitor assay was carried out according to literature¹¹⁷ in 384-well plates. In

each well was contained 100 mM sodium phosphate and 0.2 mM EDTA, pH 7.0. 100 ng PDI was first incubated with 10 μ L buffer (with or without 200 μ M DTT) for 30 min. And then different concentrations of probes in buffer (40, 20, 10, 5, 2.5, 1.25, 0.625, 0.3125 μ M, 1% DMSO in final concentration, 10 μ L) were then added. After 2 h incubation, insulin (0.16 mM in final concentration) and dithiothreitol (1 mM in final concentration) was added into each well. The enzyme reaction was monitored at $\lambda_{\text{abs}} = 650$ nm on a Biotek microplate (Germany). IC₅₀ plot was shown in Figure 5.9. Addition of DTT before assay is known to increase PDI activity.¹¹⁸ The IC₅₀ of **5-1** without DTT (4.3 ± 1.1 μ M) is slight higher than **5-1** with DTT incubation (3.3 ± 1.1 μ M). It appears that increasing PDI activity with DTT is more convenient for **5-1** to labeling with PDI and then cause higher inhibition. The hypothesis may also be approved by PDI labeling experiments.

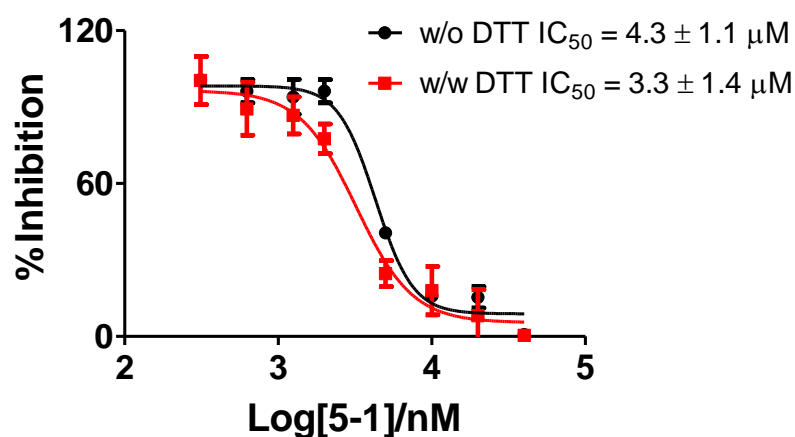


Figure 5.9 IC₅₀ curves for **5-1** against PDI with or without DTT pretreatment.

5.4.5.2 Labeling experiment of PDI protein

Next, we did the dose-dependency by varying the concentration of **5-1** in the pure PDI labeling reaction. Different concentrations of **5-1** were incubated with 25 ng PDI for 2 h. After incubation, click cocktail reagents were then added. The mixtures were gently shaken for 2 h. SDS-loading buffer was then added and the solution was heated for 10 min at 95 °C with gentle mixing. The proteins were resolved in a 10% SDS-PAGE gel and the labeled bands were visualized by in-gel fluorescence scanning. From Figure 5.10, as expected, continued increases in the fluorescence intensity of labeled bands were observed with increasing concentration of **5-1**.

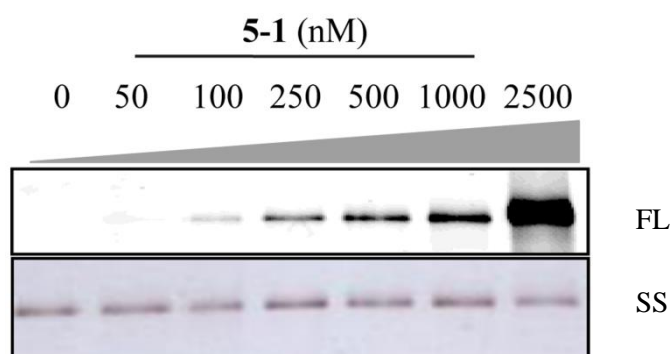


Figure 5.10 Concentration-dependent labeling of pure PDI protein by **5-1** (2 h incubation). FL = fluorescent gel. SS = silver stained gel.

Next, we went to test whether DTT, H₂O₂ and cystamine could affect the labeling efficiency. Cystamine, an organic disulfide, has been shown to decrease the PDI activity¹¹⁹ and could compete with a known PDI inhibitor (16F16A) for binding to PDI.^{116c} H₂O₂ is known to oxidize the cysteine.¹²⁰ Three different concentrations of DTT (1000, 200, 50 μM in final concentration), H₂O₂ (10, 2, 0.5 mM in final concentration) and cystamine (1000, 200, 50 μM in final concentration) was incubated with PDI (100 ng) for 1 h first in previous assay buffer. **5-1** (1 μM in final

concentration) was then added. After 2 h incubation, click cocktail was added to initial click reaction. The mixtures were incubated for 2 h. All the labeling reactions were quenched by addition of SDS-loading buffer followed by heating at 95 °C. The labeled proteins were resolved on a 10% SDS-PAGE gel and the labeled bands were visualized under fluorescence scanner.

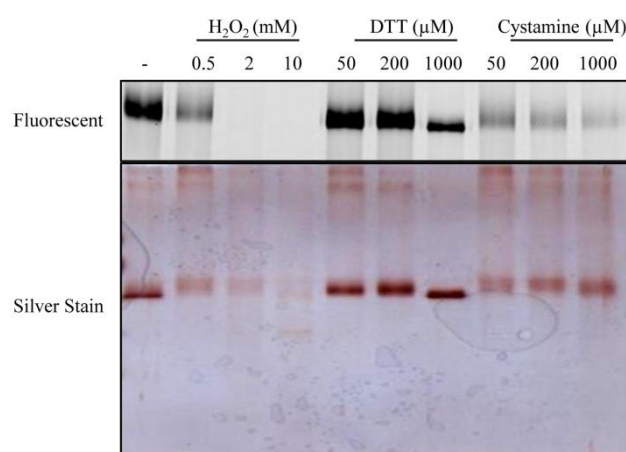


Figure 5.11 Labeling of pretreatment of PDI with different concentrations of H₂O₂, DTT, cystamine by **5-1** (up: fluorescent gel; down: silver stain).

Subsequent fluorescent SDS-PAGE analysis also clearly shows an increase intensity of labeling PDI in the case of adding DTT, while pretreatment with either cystamine or H₂O₂, the labeling efficiency is decreasing. From the silver stain, PDI could be decomposed under cystamine and H₂O₂, which needs to be addressed. H₂O₂ is oxidized cysteine of PDI which could nucleophilic attack with **5-1**. Cystamine may react with cysteine of PDI and form a disulfide bond and this cysteine subsequently is lost its nucleophilic property. All these data indicated that our probe is possibly reacting with cysteine of PDI active site.

5.4.5 *In situ* proteomic profiling and pull-down experiments of different cell lines

We investigated *in situ* proteome profiling with different cell lines. We chose four cell lines (HepG2, HeLa, MCF-7 and HEK293T), containing normal and cancer cell lines. 10 μ M **5-1** was incubated with different cell lines in DMEM medium for 2 h. Then cells were lysed and subject to CC reaction condition. After precipitation and washing, in-gel fluorescence analysis of the reaction products was obtained (Figure 5.12).

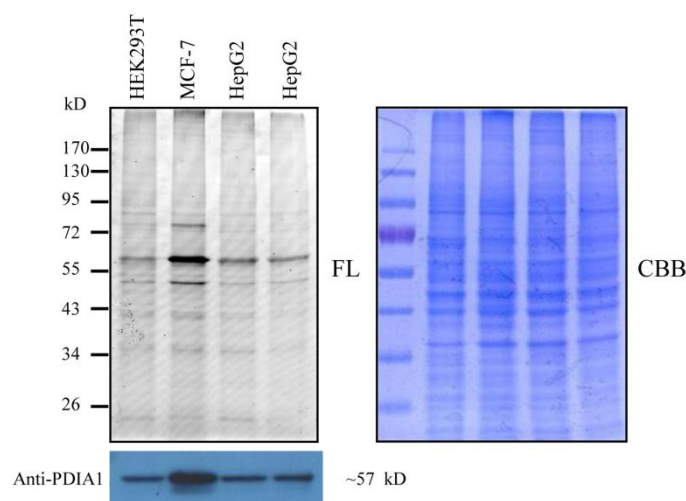


Figure 5.12 a) *in situ* labeling of four cell lines (HEK293T, MCF-7, HeLa, HepG2) by **5-1** (10 μ M, 2 h incubation). b) Western blotting analysis of different cell lines with anti-PDIA1 antibody.

Then pull-down experiments was done to analyze whether the labeling band is PDIA1. Different cell lines' pull-down is also the same as the pull-down experiment of MCF-7 cell line using trifunctional tag. Subsequently, SDS-PAGE gel was analyzed by fluorescent scanning (Figure 5.13, left) and then transferred to a PVDF membrane to obtain anti-PDIA1 western results (Figure 5.13, right).

From the results, ~ 57 kD band is still the specific labeling in different cell lines. It appears cancer cell lines have higher labeling intensity. Then western blotting for pull-down samples of four cell lines confirmed PDIA1 is the major target of **5-1** against four cell lines. Subsequently, western blotting analyzed the amount of PDIA1 in these labeled proteome samples. MCF-7 has more PDI in four cell lines and the labeling intensity is also higher.

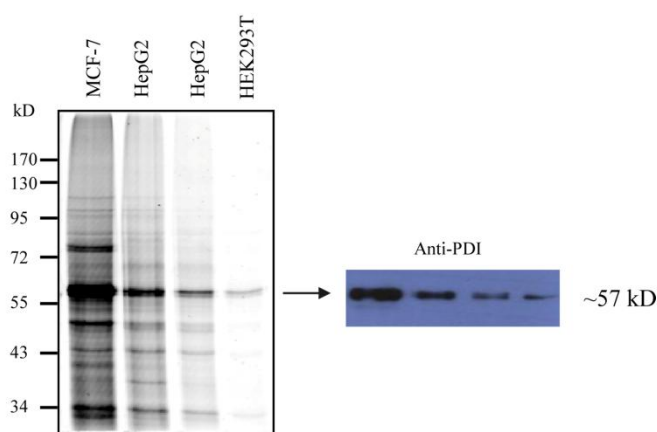


Figure 5.13 *In situ* pull-down results of four cell lines. Left: Fluorescent gel of *in situ* pull-down samples of four cell lines treated with **5-1** (5 μ M, 2 h incubation) and Right: western blotting analysis of these samples with anti-PDI antibody.

5.4.5 Cell proliferation-XTT assay

We measured the cytotoxic effect of **5-1** on the viability of four cell lines using XTT cell. HeLa, HepG2, MCF-7 and HEK293T cell lines were seeded in 96-well plates at a concentration of 4000 cells per well. The cells were grown for 24 h before treatment with **5-1**. After removing the growth medium by suction, different concentrations of **5-1** (40, 20, 10, 5, 2.5, 1, 0.5, 0.2, 0.1 μ M) in growth medium (100 μ L, 1% DMSO) were added and the cells were incubated for 24 h at 37 $^{\circ}$ C/5% CO₂. Subsequently, 50 μ L (1 mg/mL XTT (Invitrogen), 0.025 mM PMS (Sigma)) solution

were added and the mixture were incubated at 37 °C /5% CO₂ for 6 h. The absorbance was measured at 450 nm and background absorbance was measured at 650 nm using a Bioteck™ plate reader. Cells incubated with 1% DMSO served as positive control. Duplicates were conducted. IC₅₀ plot was shown in Figure 5.14. Especially, more PDIA1 protein in the cells (Figure 5.11) showed stronger labeling intensity and higher cytotoxicity. It indicated that PDIA1 may play an important role in the viability of cells.

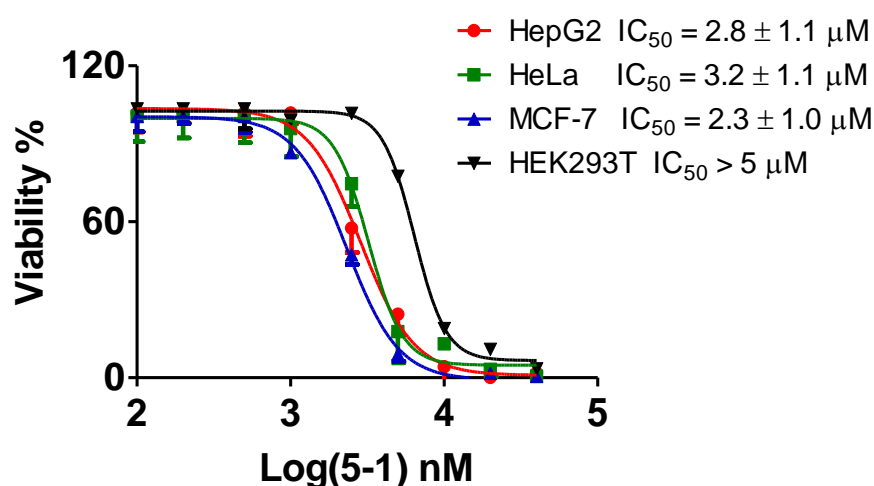


Figure 5.14 Effect of 5-1 on the cell proliferation of four cell lines.

5.5 Conclusion

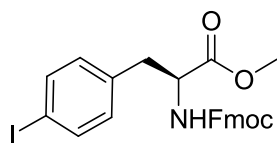
In conclusion, we have successfully synthesized and evaluated inhibitor-like probes with different phosphotyrosine mimetics. While subsequently *in situ* labelling results were surprising that our major target is PDI not PTPs. To my knowledge, it is the first time reported that a vinyl sulfonate based small molecular is specific labelled with PDI family *in vivo*. We believe that our probes give the new insight in developing new PDI inhibitors and also will be useful in understanding the functions of PDI family in cellular process.

Chapter 6 Experimental section

6.1 General Information

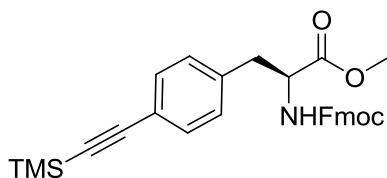
All chemicals were purchased from commercial vendors and used without further purifications, unless otherwise noted. ^1H NMR spectra and ^{13}C NMR were recorded on Bruker 300 MHz, 500 MHz or DPX-300 NMR spectrometers. Chemical shifts are reported in parts per million referenced with respect to residual solvent ($\text{CDCl}_3 = 7.26$ ppm and $d_6\text{-DMSO} = 2.5$ ppm) or from internal standard tetramethylsilane (TMS = 0.00 ppm). The following abbreviation was used in reporting spectra: s = singlet, d = doublet, t = triplet, q = quarter, m = multiplet, dd = doublet of doublets. All solvents used were of HPLC grade, unless otherwise indicated. All reactions requiring anhydrous conditions were conducted under a nitrogen or argon atmosphere in flame-dried glasswares. Dimethylformamide was dried over calcium hydride and distilled under reduced pressure. Analytical thin layer chromatography was performed using Merck silica gel plates (0.25 mm thickness) with fluorescent indicator UV254. Spots were visualized by ultraviolet illumination, iodine staining, KMnO_4 staining (for alkenes, aldehydes), Nindhydrin staining (for primary amines), or cerium molybdate staining. All LC profiles and mass spectra were recorded on Shimadzu LC-IT-TOF and LC-ESI systems equipped with an autosampler, using reverse-phase Phenomenex Luna 5 C_{18} (2) 100 Å 50 × 3.0 mm column. 0.1% formic acid/ H_2O and 0.1% formic acid/acetonitrile were used as eluents for the LC-ESI/MS system and 0.1% TFA/ H_2O and 0.1% TFA/acetonitrile for the LC-IT-TOF/MS. The flow rate for both was 0.6 ml/min. Solid-phase reactions were performed on an IRORI AccuTagTM Combinatorial Chemistry System.

6.2 Synthesis of unnatural amino acid **2-6'** and relatives



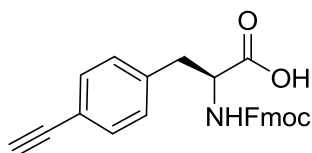
Methyl 2-(((9*H*-fluoren-9-yl)methoxy)carbonylamino)-3-(4-iodophenyl) propanoate (**2-7**)

Thionyl Chloride (15 mL) was added dropwise to 4-Iodo-*L*-phenylalanine (12 g, 41.3 mmol, synthesized following reference 78) in a methanol (200 mL) solution at 0 °C. The yellow solution was refluxed for 2 h. The solution was evaporated and dissolved in a minimal amount of hot MeOH. Et₂O was added to obtain the white product precipitation. The crude product was dissolved in NaHCO₃ (sat.) and dichloromethane (DCM) (20 mL each). To this solution was added Fmoc-OSu (15.2 g, 45.4 mmol), followed by stirring for 5 h. DCM was evaporated *in vacuo*. The residue was extracted with EtOAc three times. The organic layer was combined, washed with brine and dried over Na₂SO₄. Flash column chromatography (9/1 to 3/1 Hexane/EtOAc gradients) afforded the protected unnatural amino acid **2-7** as a white solid (15.2 g, 70 % in two steps). ¹H-NMR (300 MHz, CDCl₃) δ 7.78 (d, *J* = 7.38 Hz, 2H), 7.55-7.61 (m, 4H), 7.41 (t, *J* = 7.40 Hz, 2H), 7.32 (t, *J* = 7.39 Hz, 2H), 6.81 (d, *J* = 7.89 Hz, 2H), 5.23 (d, *J* = 8.22 Hz, NH), 4.63-4.65 (1H, m), 4.34-4.50 (2H, m), 4.20 (t, *J* = 6.42 Hz, 1H), 3.73 (s, 3H), 2.98-3.08 (m, 2H). ¹³C-NMR (75 MHz, CDCl₃) δ 171.6, 155.5, 143.6, 141.3, 137.6, 135.4, 131.2, 127.7, 127.1, 125.0, 120.0, 92.6, 66.8, 54.5, 52.4, 47.14, 37.7. IT-TOF: *m/z* [M+H]⁺ calcd: 528.06, found: 528.03.



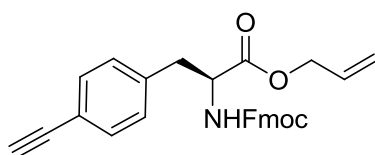
Methyl 2-(((9*H*-fluoren-9-yl)methoxy)carbonylamino)-3-(4-((trimethylsilyl) ethynyl) phenyl)propanoate (**2-8**)

DMF (25 mL) was degassed by bubbling with argon in a dry two-neck round bottle flask. Subsequently, compound **2-7** (1.2 g, 2.28 mmol) was added and the solution was stirred until it became homogenous. Pd(PPh₃)₂Cl₂ (16 mg, 0.02 mmol) and CuI (18.9 mg, 0.1 mmol), DIEA (1.2 mL, 6.5 mmol) were subsequently added under an N₂ atmosphere. Trimethylsilyl acetylene (0.63 mL, 4.56 mmol) was then added by a needled syringe. The reaction was heated at 45 °C for 3 h before being quenched by addition of an ammonium chloride solution and EtOAc. Then the aqueous phase was extracted with EtOAc. The combined organic phases were washed with brine and dried with Na₂SO₄. After removal of the solvent under reduced pressure, column chromatography on silica gel (1/9 to 1/6 EtOAc/Hexane gradients) afforded **2-8** as a light yellow solid (84 % yield). ¹H-NMR (300 MHz, CDCl₃) δ 7.75 (d, *J* = 7.38 Hz, 2H), 7.53-7.55 (broad, 2H), 7.27-7.41 (m, 6H), 7.01 (d, *J* = 7.71 Hz, 2H), 5.27 (d, *J* = 8.04 Hz, NH), 4.64 (m, 1H), 4.35-4.46 (m, 2H), 4.19 (t, *J* = 6.66 Hz, 1H), 3.69 (s, 3H), 3.05-3.10 (m, 2H), 0.25 (s, 9H). ¹³C-NMR (75 MHz, CDCl₃) δ 171.6, 155.4, 143.7, 141.3, 136.3, 132.1, 129.1, 127.7, 127.0, 125.0, 121.9, 120.0, 104.7, 94.4, 66.9, 54.6, 52.3, 47.1, 38.1, -0.1. IT-TOF: *m/z* [M+Na]⁺ calcd: 520.20, found: 520.17.



2-(((9*H*-fluoren-9-yl)methoxy)carbonylamino)-3-(4-ethynylphenyl)propanoic acid (**2-9**)

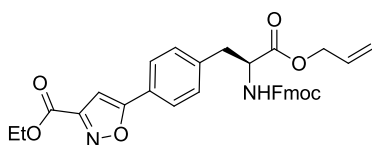
The compound **2-8** (2.0 g, 4 mmol) was dissolved in MeOH (10 mL) and cooled to 0 °C. NaOH (10 mL, 10 M) was added dropwise. After 5 h, the *pH* of the solution was adjusted to 8 using HCl and NaHCO₃ (sat). Then to the mixture was added Fmoc-OSu (1.6 g, 4.8 mmol) followed by stirring overnight. MeOH was removed *in vacuo* and the aqueous layer was extracted with EtOAc. The combined organic phases were washed with brine and dried with Na₂SO₄. After removal of the solvent under reduced pressure, column chromatography on silica gel (9/1 Hexane/EtOAc to 9/1 EtOAc/MeOH with 0.1 % acetic acid gradients) afforded **2-9** (1.07 g, 65 %) as a white solid. ¹H-NMR (300 MHz, CD₃OD) δ 7.78 (d, *J* = 7.38 Hz, 2H), 7.57 (d, *J* = 7.41 Hz, 2H), 7.38 (t, *J* = 7.56 Hz, 4H), 7.26-7.31 (m, 2H), 7.11 (d, *J* = 8.04 Hz, 2H), 4.13-4.41 (m, 4H), 3.42 (s, 1H), 2.90-3.24 (m, 2H). ¹³C-NMR (75 MHz, DMSO-*d*₆) δ 174.8, 158.3, 145.2, 142.5, 139.6, 133.0, 128.7, 128.1, 126.1, 122.1, 120.9, 84.3, 78.5, 67.9, 56.5, 48.3, 38.3, 26.3. IT-TOF: *m/z* [M+Na]⁺ calcd: 434.15, found: 434.12.



Allyl 2-(((9*H*-fluoren-9-yl)methoxy)carbonylamino)-3-(4-ethynylphenyl)propanoate (**2-10**)

The compound **2-9** (0.75 g, 1.8 mmol) was dissolved in MeOH (10 mL). Cs₂CO₃ (0.28 g, 0.85 mmol) was added over a period of 10 min. Subsequently, MeOH was

removed *in vacuo*. The residue was redissolved in DMF (20 mL). Allyl bromide (0.68 g, 5.6 mmol) was added. After 1 h, DMF was distilled off under reduced pressure. Water was added and the solution was extracted with EtOAc. The combined organic phase was washed with brine, dried over Na₂SO₄ and concentrated. The crude product was purified by silica gel chromatography (EtOAc/Hexane gradients). The white product **2-10** was obtained in 90 % yield. ¹H-NMR (300 MHz, CDCl₃) δ 7.78 (d, *J* = 7.41 Hz, 2H), 7.57 (d, *J* = 7.23 Hz, 2H), 7.39-7.43 (m, 4H), 7.32-7.34 (t, *J* = 7.40 Hz, 2H), 7.06 (d, *J* = 7.86 Hz, 2H), 5.83-5.88 (m, 1H), 5.26-5.35 (m, 3H), 4.61-4.72 (m, 3H), 4.35-4.50 (m, 2H), 4.21 (t, *J* = 6.82 Hz, 1H), 3.08-3.15 (m, 3H). ¹³C-NMR (75 MHz, CDCl₃) δ 170.9, 155.5, 143.8, 141.3, 136.6, 132.3, 131.2, 129.4, 127.7, 127.0, 125.0, 121.0, 120.0, 119.3, 77.3, 83.3, 66.9, 66.2, 54.6, 47.2, 38.1. IT-TOF: *m/z* [M+Na]⁺ calcd: 474.18, found: 474.15.



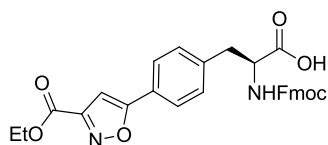
Ethyl 5-(4-(2-(((9H-fluoren-9-yl)methoxy)carbonylamino)-3-(allyloxy)-3-oxopropyl)phenyl)isoxazole-3-carboxylate (**2-11**)^{79, 80}

Method 1: Compound **2-10** (0.1 g, 0.22 mmol) and ethyl chlorooximidoacetate (33 mg, 0.22 mmol) were dissolved in *t*-BuOH (5 mL). To the resulting solution, H₂O was added, followed by sodium ascorbate (4.35 mg, 0.02 mmol), copper sulfate (2.75 mg, 0.01 mmol) and finally KHCO₃ (90 mg, 0.9 mmol). The reaction mixture was stirred at room temperature for 24 h. Upon evaporation of *t*-BuOH, the reaction was extract several times with EtOAc. The organic layers were then collected, dried over anhydrous Na₂SO₄ and filtered, and then the solvent was removed under reduced

pressure to yield a crude product. The product was purified by flash chromatography to yield **2-11** (21 %) as a white solid.

Method 2: Compound **2-10** (0.5 g 1.1 mmol) and ethyl chlorooximidoacetate (0.25 g, 1.6 mmol) were dissolved in THF (20 mL). Triethylamine (0.22 g, 2.2 mmol) was slowly added into solution. The mixture was stirred for 1 day. THF and TEA was evaporated and the residue redissolved in EtOAc. After washing with brine and drying over Na₂SO₄, the organic layer was evaporated. The crude product was purified by flash chromatography to obtain **2-11** (15%).

¹H-NMR (500 MHz, CDCl₃) δ 7.77 (d, *J* = 7.60 Hz, 2H), 7.70 (d, *J* = 7.60 Hz, 2H), 7.55-7.57 (m, 2H), 7.40 (t, *J* = 7.40 Hz, 2H), 7.30 (t, *J* = 7.55 Hz, 2H), 7.21 (d, *J* = 8.20 Hz, 2H), 5.85-5.91 (m, 1H), 5.26-5.38 (m, 3H), 4.70-4.74 (m, 1H), 4.63 (d, *J* = 5.7 Hz, 1H), 4.47 (q, *J* = 7.55 Hz, 2H), 4.35-4.39 (m, 1H), 4.19 (t, *J* = 6.95 Hz, 1H), 3.18 (dd, *J* = 13.87, 5.65 Hz, 2H), 1.45 (t, *J* = 6.96 Hz, 3H). ¹³C-NMR (125 MHz, CDCl₃) δ 171.1, 170.8, 159.9, 156.9, 155.5, 143.7, 141.3, 138.9, 131.1, 130.1, 127.7, 127.0, 126.0, 125.4, 125.0, 120.0, 119.4, 99.8, 66.8, 66.2, 62.2, 54.6, 47.1, 38.1, 14.1. IT-TOF: *m/z* [M+H]⁺ calcd: 567.21, found: 567.20.



2-(((9H-fluoren-9-yl)methoxy)carbonylamino)-3-(4-(3-(ethoxycarbonyl)isoxazol-5-yl)phenyl)propanoic acid (**2-6'**)⁸¹

Compound **2-11** (20 mg, 0.03 mmol) was dissolved in DCM under an N₂ atmosphere. Phenylsilane (8 mg, 0.07 mmol) and Pd(PPh₃)₄ (1.7 mg, 1.5 μmol) were added. After 30 min, the solution became slightly dark and TLC was showed the reaction was finished. The solvent was removed under reduced pressure and the residue was

directly purified by flash chromatography to obtain compound **2-6'** (89 %). $^1\text{H-NMR}$ (300 MHz, $\text{DMSO-}d_6$) δ 13.1 (br, 1H), 7.86 (d, $J = 7.23$ Hz, 4H), 7.61-7.68 (m, 3H), 7.37-7.45 (m, 5H), 7.25-7.31 (m, 2H), 4.39 (q, $J = 6.9$ Hz, 2H), 4.17-4.19 (m, 4H), 2.91-3.19 (m, 2H), 1.35 (t, $J = 7.05$ Hz, 3H). $^{13}\text{C-NMR}$ (75 MHz, $\text{DMSO-}d_6$) δ 173.1, 171.1, 159.4, 156.8, 155.9, 143.7, 141.4, 140.7, 130.1, 127.6, 127.0, 125.7, 125.2, 124.3, 120.1, 100.4, 65.6, 61.2, 55.2, 46.5, 36.3, 14.0. HRMS calcd for $\text{C}_{30}\text{H}_{25}\text{N}_2\text{O}_7$: 525.1662, found: 525.1647.

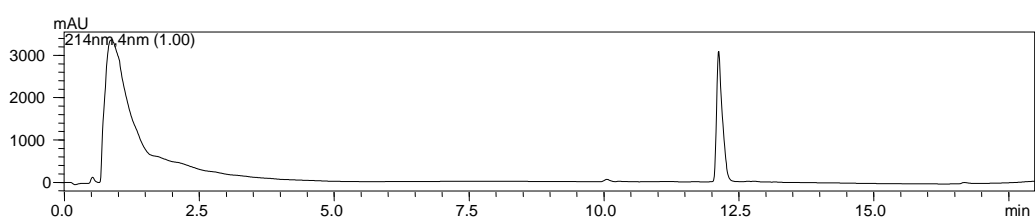
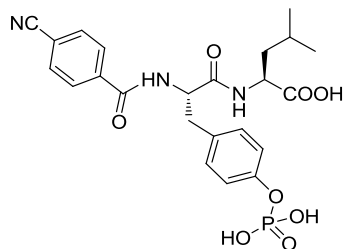


Figure 6.1 The analytical HPLC results of compound **2-6'**

6.3 Solid-phase synthesis of **I1** and **I2**



Inhibitor **I1**. Wang resin was used to synthesize the inhibitors. First, Wang resin (100 mg, 0.05 mmol, and loading ~ 0.5 mmol/g) was previously swollen in DMF. Fmoc-Leu-OH (4 equiv) was pre-activated using HOBT (4 equiv), HBTU (4 equiv), DIEA (8 equiv) in DMF. And the resins were added into the solution. The reaction was shaken for 4 h at room temperature. After that, the resin was washed with DMF (3x), DCM (3x) and DMF (3x). Next, Fmoc-protected amino functionalized resin was treated with 20% piperidine in DMF for 1 h at room temperature. The resin was

washed with DMF (3x), DCM (3x) and DMF (3x). Fmoc-Tyr (HPO₃Bzl)-OH were pre-activated by HOBT (4 equiv), HBTU (4 equiv), DIEA (8 equiv) in DMF. The resin was added into the solution and shaken at room temperature overnight. Subsequently, the resin was washed with DMF (3x), DCM (3x) and DMF (3x). Deprotection of the Fmoc group was carried out with 2% DBU in DMF for 2 h. And the resin was washed again. 4-Cyanobenzoic acid was coupled with the resin using HOBT (4 equiv), HBTU (4 equiv), and DIEA (8 equiv) in DMF overnight. The resin was then washed with DCM and MeOH before being dried *in vacuo*. The cocktail cleavage solution contains TFA/TIS/DCM (95:2.5:2.5). The resin was treated with the cocktail for 2.5 h, after which the resin was filtered off and the solvent was removed under reduced pressure. The resulting inhibitor was further purified by preparative HPLC.

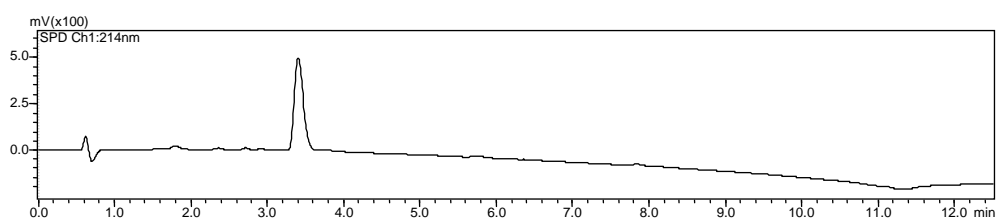
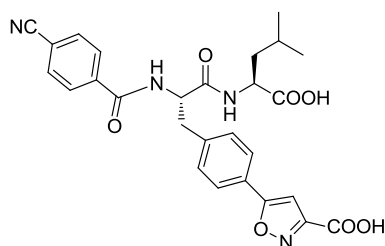


Figure 6.2 The LCMS results of compound **I1**



Inhibitor **I2**. The inhibitor **I2** was synthesized similarly as **I1** following procedures as shown above. The unnatural amino acid, compound **2-6'**, was coupled to the resin as follows. First, **2-6'** was pre-activated by HOBT (4 equiv), HBTU (4 equiv), DIEA (8 equiv) in DMF. Then resin was added into the solution and the resulting mixture was

shaken at room temperature overnight. Subsequently, the resin was washed with DMF (3x), DCM (3x) and DMF (3x). The deprotection of Fmoc group was carried on 2 % DBU in DMF for 2 h. And the resin was washed again. 4-Cyanobenzoic acid was coupled with the resin using HOBt (4 equiv), HBTU (4 equiv), and DIEA (8 equiv) in DMF overnight. The resin was washed, after coupling, with DMF (3x), DCM (3x) and DMF (3x). At the end of solid phase synthesis, the resin was washed with DCM and MeOH before being dried *in vacuo*. The inhibitor was cleaved off the resin by treatment with the cocktail mixture, TFA/TIS/DCM (95:2.5:2.5), for 2.5 h. After that, the resin was filtered off and the solvent was removed under reduced pressure to give the protected form of the final product (precursor). The final product was subsequently obtained by treating the precursor with LiOH (2 M) in THF for 2 h followed by purification with preparative HPLC. ¹H-NMR (500 MHz, DMSO-*d*₆) δ 8.90 (d, *J* = 8.85 Hz, 1H), 8.42 (d, *J* = 8.20 Hz, 1H), 7.90-7.92 (m, 4H), 7.84 (d, *J* = 8.20 Hz, 1H), 7.54 (d, *J* = 8.85 Hz, 2H), 7.34 (s, 1H), 4.81-4.84 (m, 1H), 4.25-4.29 (m, 1H), 3.01-3.25 (m, 2H), 1.60-1.69 (m, 1H), 1.53-1.59(m, 2H), 0.89 (dd, *J* = 14.80, 6.30 Hz, 6H) ¹³C-NMR (125 MHz, DMSO-*d*₆) δ 173.9, 171.0, 164.9, 160.8, 157.7, 141.4, 139.5, 137.9, 132.3, 130.0, 128.2, 125.6, 124.4, 118.3, 113.7, 100.4, 53.4, 50.4, 36.9, 24.3, 22.8, 21.3. HRMS calcd C₂₇H₂₅N₄O₇: m/z [M+H]⁺ calcd: 517.1724, found: 517.1721.

Based on the ¹H NMR result, we found no sign of racemization in the final product **I2**, indicating that during the synthesis of the new amino acid **2-6'** and the final inhibitor **I2**, the two steps where base treatments were involved (**2-8** to **2-9** in Scheme 2.1 and the LiOH treatment of the cleaved peptide giving **I2** in Scheme 2.2) did not cause noticeable racemization.

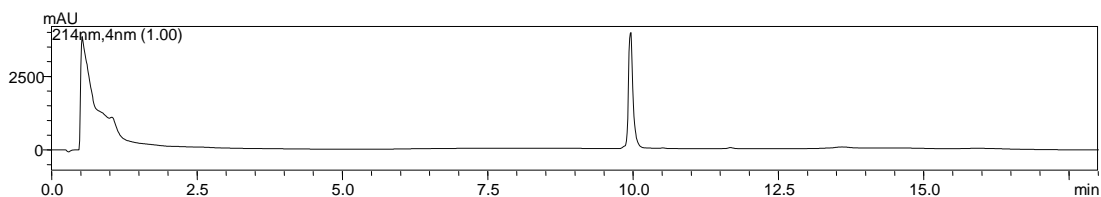
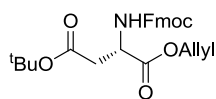


Figure 6.3 The analytical HPLC results of compound **12**

6.4 Solid-phase synthesis of SH2-binding peptide Fluorescein-GY*LPQTV-NH₂

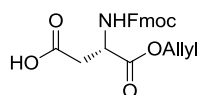
Rink amide resin (100 mg, loading ~ 0.5 mmol/g) was first swollen in DMF and washed with DMF (2x). Before coupling, 20% piperidine in DMF was used to deprotect the Fmoc group for 1 h at RT. Subsequently, the resin was washed with several cycles of DMF and DCM. Fmoc-Val-OH (4 equiv) was pre-activated using HOBt (4 equiv), HBTU (4 equiv), DIEA (8 equiv) in DMF. The resin was put into the solution and shaken for 4 h. Next, the resin was washed with several cycles of DMF and DCM and treated with 20% piperidine to remove the Fmoc group prior to the next coupling. The cycle of coupling was repeated with Fmoc-Thr(OtBu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Pro-OH, Fmoc-Leu-OH, Fmoc-Tyr(HPO₃Bzl)-OH, Fmoc-Gly-OH. Final, the resin was coupled with 3', 6'-diacetyl fluorescein (4 equiv). After washing, the acetyl protecting group was removed by treatment with 20% piperidine in DMF. Finally, the resin was washed with DMF (3x), DCM (3x) and MeOH (3x) and dried thoroughly under reduced pressure. A solution of cleavage solution TFA/TIS/DCM (95:2.5:2.5) was added to the resin for 2 h. The resin was filtered off. The solvent was removed *in vacuo*. The peptide was purified by preparative HPLC.

6.5 Synthesis of unnatural amino acid **3-2** and relatives



(*S*)-1-allyl 4-tert-butyl 2-(((9H-fluoren-9-yl)methoxy)carbonylamino)succinate (**3-3**)

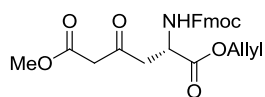
The compound FmocNH-Asp(O^tBu)-COOH (20.0 g, 48.6 mmol) was dissolved in MeOH (150 mL). Cs₂CO₃ (7.9 g, 24.3 mmol) was added into the solution and the reaction was stirred for 30 min. Subsequently, MeOH was removed *in vacuo*. The residue was redissolved in DMF (200 mL). Allyl bromide (8.4 mL, 72.9 mmol) was added dropwise and the reaction mixture was stirred overnight. DMF was removed under reduced pressure. Water was added and the solution was extracted with EtOAc. The combined organic phase was washed with water, 10% NaHCO₃ and brine. The organic layer was dried over Na₂SO₄ and concentrated. The product **3-3** was obtained as a white solid (16.6 g, 95%). ¹H NMR (500 MHz, CDCl₃) δ 7.75 (d, *J* = 7.5 Hz, 2H), 7.62 (t, *J* = 7.1 Hz, 2H), 7.39 (t, *J* = 7.4 Hz, 2H), 7.31 (t, *J* = 7.4 Hz, 2H), 5.97 (d, *J* = 8.5 Hz, 1H), 5.87 - 5.96 (m, 1H), 5.34 (d, *J* = 17.2 Hz, 1H), 5.24 (d, *J* = 10.4 Hz, 1H), 4.63 - 4.73 (m, 3H), 4.33 - 4.47 (m, 2H), 4.25 (t, *J* = 7.2 Hz, 1H), 2.90 (dd, *J* = 16.8, 4.5 Hz, 2H), 1.47 (s, 9H). ¹³C NMR (125 MHz, CDCl₃) δ 170.39, 169.71, 155.77, 143.70, 141.06, 131.35, 127.48, 126.84, 124.95, 119.75, 118.46, 81.52, 67.02, 65.99, 50.45, 46.88, 37.56, 27.79. IT-TOF: *m/z* [M+H]⁺ calcd: 452.207, found: 452.216.



(*S*)-3-(((9H-fluoren-9-yl)methoxy)carbonylamino)-4-(allyloxy)-4-oxobutanoic acid

(**3-4**)

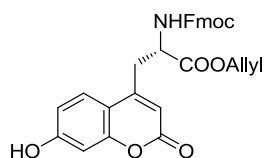
The compound **3-3** (10 g, 22.2 mmol) was dissolved in DCM : TFA (1:1, 80 mL) and the mixture was stirred at room temperature for 5 h. The solvent was then evaporated under reduced pressure and the residue was precipitated by addition of cold water (400 mL). The solution was stirring overnight. The resulting precipitate was filtered, dried to give the desired product as a white solid (8.4 g, 96%). ¹H NMR (500 MHz, *d*₆-DMSO) δ 7.87 (t, *J* = 7.8 Hz, 2H), 7.71 (d, *J* = 7.4 Hz, 2H), 7.42 (t, *J* = 7.4 Hz, 2H), 7.33 (t, *J* = 7.3 Hz, 2H), 5.84 - 5.92 (m, 1H), 5.30 (d, *J* = 17.3 Hz, 1H), 5.19 (d, *J* = 10.5 Hz, 1H), 4.59 (d, *J* = 5.1 Hz, 2H), 4.49 (dd, *J* = 13.7, 7.9 Hz, 1H), 4.29 - 4.36 (m, 2H), 4.23 (t, *J* = 6.9 Hz, 1H), 2.73 (dd, *J* = 16.6, 6.8 Hz, 2H). ¹³C NMR (125 MHz, *d*₆-DMSO) δ 171.44, 170.85, 155.88, 143.78, 140.75, 132.24, 127.65, 127.09, 125.20, 120.11, 117.56, 65.80, 65.07, 50.60, 46.63, 35.90. IT-TOF: *m/z* [M+H]⁺ calcd: 396.145, found: 396.121. *m/z* [M+Na]⁺ calcd: 418.127, found: 418.109.



(*S*)-1-allyl 6-methyl 2-(((9H-fluoren-9-yl)methoxy)carbonylamino)-4-oxohexanedioate (**3-5**)

Compound **3-4** (10 g, 25.3 mmol), Meldrum's acid (4.7 g, 32.6 mmol) and DMAP (3.1 g, 25.3 mmol) in 100 mL DCM was cooled in an ice bath. To this reaction mixture, *N,N'*-diisopropylcarbodiimide (DIC) was added very slowly. The reaction mixture was allowed to stir around 5 h, or till the disappearance of the starting material was detected by TLC. The resulting solution was washed with 10% aqueous solution of KHSO₄ (75 mL, 3x) and brine. The organic layer was dried over Na₂SO₄ and evaporated *in vacuo*. The residue was dissolved in methanol/toluene (10:1). The solution was refluxed overnight. After evaporation of the solvent, the desired product

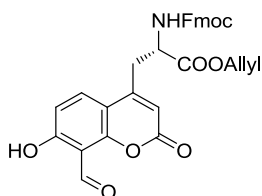
was obtained by recrystallization in hexane/EtOAc (10.1 g, 89%).^[1] ¹H NMR (500 MHz, CDCl₃) δ 7.74 (d, *J* = 7.5 Hz, 2H), 7.55 - 7.62 (m, 2H), 7.38 (t, *J* = 7.4 Hz, 2H), 7.29 (t, *J* = 7.4 Hz, 2H), 5.96 (d, *J* = 7.1 Hz, 1H), 5.82 - 5.90 (m, 1H), 5.30 (d, *J* = 17.2 Hz, 1H), 5.23 (d, *J* = 10.4 Hz, 1H), 4.61 - 4.67 (m, 3H), 4.29 - 4.42 (m, 2H), 4.21 (t, *J* = 7.0 Hz, 1H), 3.68 (s, 3H), 3.46 (s, 2H), 3.22 (dd, *J* = 18.4, 4.4 Hz, 2H). ¹³C NMR (125 MHz, CDCl₃) δ 200.48, 170.19, 166.74, 155.80, 143.61, 141.02, 131.27, 127.47, 126.83, 124.92, 119.73, 118.48, 66.96, 66.08, 52.14, 49.68, 48.52, 46.83, 44.26. IT-TOF: *m/z* [M+H]⁺ calcd: 452.171, found: 452.143.



(*S*)-allyl-2-(((9H-fluoren-9-yl)methoxy)carbonylamino)-3-(7-hydroxy-2-oxo-2H-chromen-4-yl)propanoate (**3-6**)

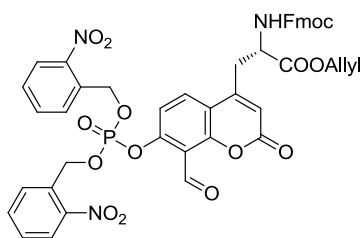
Compound **3-5** (8 g, 17.7 mmol) was dissolved in methanesulfonic acid (100 mL). The solution was cooled down in an ice bath. Resorcinol (1.95 g, 17.7 mmol) was added into the solution. The mixture was stirred in 4 °C for 4 h, when TLC showed the starting material was completely consumed. The reaction mixture was then slowly poured into ice water, and the resulting white precipitate was collected by vacuum filtration. Upon recrystallization in ethanol/water, the desired product was obtained (8.1 g, 90%). ¹H NMR (500 MHz, *d*₆-DMSO) δ 10.58 (s, 1H), 8.01 (d, *J* = 8.2 Hz, 1H), 7.87 (d, *J* = 7.5 Hz, 2H), 7.60 - 7.65 (m, 3H), 7.40 (t, *J* = 7.3 Hz, 2H), 7.29 (q, *J* = 7.3 Hz, 2H), 6.83 (d, *J* = 8.7 Hz, 1H), 6.74 (d, *J* = 2.0 Hz, 1H), 6.17 (s, 1H), 5.83 - 5.91 (m, 1H), 5.29 (d, *J* = 17.2 Hz, 1H), 5.20 (d, *J* = 10.5 Hz, 1H), 4.60 (d, *J* = 4.0 Hz, 2H), 4.42 - 4.45 (m, 1H), 4.17 - 4.20 (m, 2H), 4.19 (t, *J* = 6.9 Hz, 1H), 3.06 - 3.30 (m, 2H). ¹³C NMR (125 MHz, *d*₆-DMSO) δ 170.71, 161.22, 159.98, 155.88, 155.09, 152.16,

143.64, 140.66, 132.08, 127.59, 127.01, 125.92, 125.07, 120.06, 117.83, 113.11, 111.23, 110.88, 102.54, 65.78, 65.25, 52.80, 46.50, 32.36. IT-TOF: m/z $[M+H]^+$ calcd: 512.171, found: 512.135.



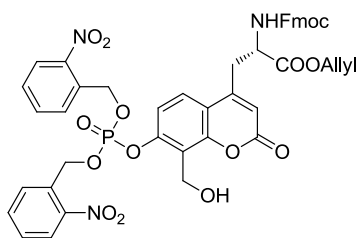
(*S*)-allyl-2-(((9H-fluoren-9-yl)methoxy)carbonylamino)-3-(8-formyl-7-hydroxy-2-oxo-2H-chromen-4-yl)propanoate (**3-7**)

Compound **3-6** (5 g, 9.78 mmol) was added into acetic acid (50 mL). The mixture was heated to around 100 °C, until compound **3-6** was dissolved. Hexamine (4.1 g, 29.4 mmol) was added slowly. Then the reaction was heated to 150 °C and stirred for 3 h. The solvent was removed *in vacuo*. 1 M HCl was added to the residue and the mixture was further stirred for 1 h. The solution was extracted with EtOAc (3x), washed with water and brine (2x). The organic layer was dried over Na₂SO₄ and concentrated. The product was obtained by purification with flash column chromatography (1.32 g, 25%).⁹⁴ ¹H NMR (500 MHz, CDCl₃) δ 12.21 (s, 1H), 10.56 (s, 1H), 7.90 (d, *J* = 9.0 Hz, 1H), 7.74 (d, *J* = 7.5 Hz, 2H), 7.54 (d, *J* = 7.3 Hz, 2H), 7.38 (t, *J* = 7.4 Hz, 2H), 7.29 (t, *J* = 7.4 Hz, 2H), 6.88 (d, *J* = 9.0 Hz, 1H), 6.13 (s, 1H), 5.80 - 5.91 (m, 1H), 5.66 (d, *J* = 7.6 Hz, 1H), 5.27 - 5.38 (m, 2H), 4.69 (dd, *J* = 13.9, 6.9 Hz, 1H), 4.63 (d, *J* = 5.8 Hz, 2H), 4.41 (d, *J* = 6.8 Hz, 2H), 4.17 (t, *J* = 6.7 Hz, 1H), 3.19 (d, *J* = 6.5 Hz, 2H). ¹³C NMR (75 MHz, CDCl₃) δ 193.10, 170.00, 165.22, 158.53, 156.29, 155.63, 150.85, 143.41, 141.13, 132.58, 130.64, 127.69, 126.98, 124.80, 119.92, 114.51, 113.11, 110.73, 108.69, 67.17, 66.67, 53.11, 46.85, 34.99. IT-TOF: m/z $[M+H]^+$ calcd: 540.167, found: 540.136.



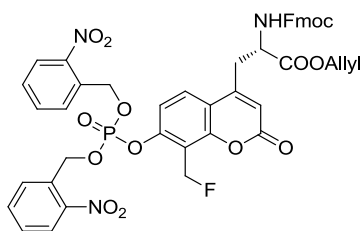
(*S*)-allyl 2-(((9H-fluoren-9-yl)methoxy)carbonylamino)-3-(7-(bis(2-nitrobenzyloxy)-phosphoryloxy)-8-formyl-2-oxo-2H-chromen-4-yl)propanoate (**3-8**)

To a suspension of di(2-nitrobenzyl) phosphate (1.22 g, 3.3 mmol) in distilled DCM (20 mL) was added two drops of DMF as a catalyst. Oxalyl chloride (2.12 g, 16.5 mmol) was then added dropwise at room temperature to the mixture. The reaction was stirred for 1 h. Upon solvent removal *in vacuo*, the product obtained was directly used in the next step without further purification. Compound **3-7** (900 mg, 1.67 mmol; dissolved in 20 mL DCM) was added the above product (dissolved in DCM with ~1 mg of DMAP) in an ice bath. Subsequently, triethylamine (0.86 g, 4.18 mmol) was added dropwise. The reaction mixture was stirred for 5 h until the starting material was completely consumed. Upon solvent removal *in vacuo*, the resulting residue was purified by flash column chromatography (1:1 hexane/EtOAc) to give **3-8** as a slightly yellow solid (1.34 g, 46%). ¹H NMR (500 MHz, CDCl₃) δ 10.61 (s, 1H), 8.11 (d, *J* = 8.2 Hz, 2H), 8.02 (d, *J* = 8.5 Hz, 1H), 7.73 - 7.77 (m, 4H), 7.63 - 7.69 (m, 2H), 7.55 (d, *J* = 7.4 Hz, 2H), 7.47 - 7.50 (m, 3H), 7.36 - 7.39 (m, 2H), 7.30 (td, *J* = 7.3, 3.1 Hz, 2H), 6.27 (s, 1H), 5.80 - 5.90 (m, 1H), 5.74 (t, *J* = 7.1 Hz, 4H), 5.62 (d, *J* = 6.8 Hz, 1H), 5.25 - 5.35 (m, 2H), 4.63 - 4.66 (m, 3H), 4.42 (d, *J* = 5.8 Hz, 2H), 4.18 (t, *J* = 6.6 Hz, 1H), 3.21 - 3.22 (m, 2H). ¹³C NMR (125 MHz, CDCl₃) δ 186.48, 170.62, 158.61, 156.74, 156.42, 152.06, 150.62, 147.55, 144.21, 142.00, 134.82, 132.13, 131.32, 131.07, 129.90, 129.51, 128.50, 127.80, 125.76, 125.57, 121.05, 120.72, 118.14, 117.40, 116.79, 116.62, 68.11, 68.01, 67.57, 53.78, 47.76, 36.21. ³¹P NMR (121 MHz, CDCl₃) δ -7.84. IT-TOF: *m/z* [M+CF₃COO]⁻ calcd: 1002.173, found: 1002.143.



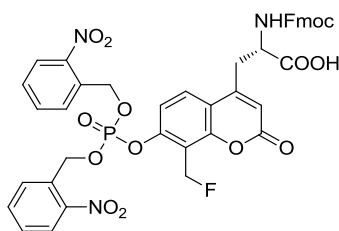
(*S*)-allyl 2-(((9H-fluoren-9-yl)methoxy)carbonylamino)-3-(7-(bis(2-nitrobenzyloxy) phosphoryloxy)-8-(hydroxymethyl)-2-oxo-2H-chromen-4-yl)propanoate (**3-9**)

Compound **3-8** (0.93 g, 1.0 mmol) was dissolved in DCM/Ethanol (1:1, 30 mL) and cooled to 0 °C. NaBH₄ (9.8 mg, 0.25 mmol) in 1 mL water was then added dropwise into the reaction. The reaction mixture was stirred at 0 °C for 5 min. The reaction was quenched with 0.25 M HCl. The organic solvents were removed *in vacuo*. The residue was purified by flash column chromatography (1:1 hexane/EtOAc) to obtain **3-9** as a white solid (757 mg, 85%). ¹H NMR (500 MHz, CDCl₃) δ 8.11 (d, *J* = 8.2 Hz, 2H), 7.68 - 7.80 (m, 4H), 7.65 (t, *J* = 7.5 Hz, 2H), 7.55 (d, *J* = 7.0 Hz, 2H), 7.49 (t, *J* = 7.7 Hz, 2H), 7.36 - 7.39(m, 2H), 7.27 - 7.34 (m, 3H), 6.23 (s, 1H), 5.78 - 5.89 (m, 1H), 5.63 - 5.66 (m, 5H), 5.23 - 5.34 (m, 2H), 4.89 (s, 2H), 4.65 - 4.47 (m, 1H), 4.61 (d, *J* = 5.7 Hz, 2H), 4.39 (d, *J* = 6.7 Hz, 2H), 4.17 (t, *J* = 6.7 Hz, 1H), 3.22 (d, *J* = 6.4 Hz, 2H). ¹³C NMR (125 MHz, CDCl₃) δ 170.15, 159.28, 155.71, 153.26, 150.93, 150.23, 146.89, 143.53, 141.28, 134.17, 131.20, 131.14, 130.73, 129.40, 128.86, 127.79, 127.11, 125.15, 124.92, 121.49, 120.12, 120.01, 116.83, 115.85, 67.46, 67.32, 66.76, 53.18, 47.07, 35.24. ³¹P NMR (121 MHz, CDCl₃) δ -6.11. IT-TOF: *m/z* [M+H]⁺ calcd: 892.212, found: 892.171. *m/z* [M+CF₃COO]⁻ calcd: 1004.189, found: 1004.135.



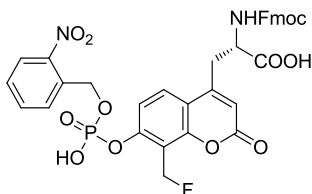
(*S*)-allyl 2-(((9H-fluoren-9-yl)methoxy)carbonylamino)-3-(7-(bis(2-nitrobenzyloxy)phosphoryloxy)-8-(fluoromethyl)-2-oxo-2H-chromen-4-yl)propanoate (**3-10**)

Compound **3-9** (602 mg, 0.67 mmol) was dissolved into distilled DCM (15 mL) at 0 °C. Diethylaminosulfur trifluoride (DAST, 218 mg, 1.35 mmol) was added slowly under an Ar atmosphere. The reaction was stirred for 10 min. The solvent was evaporated and the residue was purified by flash column chromatography to obtain compound **3-10** as a white foam solid (436 mg, 74%). ¹H NMR (500 MHz, CDCl₃) δ 8.11 (d, *J* = 8.2 Hz, 2H), 7.86 (d, *J* = 8.5 Hz, 1H), 7.69 - 7.75 (m, 4H), 7.65 (td, *J* = 7.8, 3.5 Hz, 2H), 7.55 (d, *J* = 7.0 Hz, 2H), 7.46 - 7.52 (m, 3H), 7.36 - 7.41 (m, 2H), 7.29 - 7.32 (m, 2H), 6.24 (s, 1H), 5.78 - 5.89 (m, 1H), 5.71/5.61(d, *J* = 47.4 Hz, 2H, CH₂F), 5.67 (dd, *J* = 7.7, 2.5 Hz, 4H), 5.58 (d, *J* = 7.4 Hz, 1H), 5.26 - 5.35 (m, 2H), 4.68 (d, *J* = 6.9 Hz, 1H), 4.61 (d, *J* = 5.2 Hz, 2H), 4.41 (s, 2H), 4.18 (t, *J* = 6.7 Hz, 1H), 3.22 (d, *J* = 6.2 Hz, 2H). ¹³C NMR (125 MHz, CDCl₃) δ 170.07, 158.85, 155.68, 153.35, 152.23, 150.08, 146.82, 143.51, 141.30, 134.16, 131.29, 131.23, 130.65, 129.33, 128.69, 127.80, 127.12, 126.91, 125.14, 124.98, 120.20, 120.02, 116.45, 116.08, 115.86, 115.74, 72.48 (d, *J* = 166.3 Hz, 1C, C-F), 67.32, 66.79, 53.15, 47.08, 35.33. ¹⁹F NMR (282 MHz, CDCl₃) δ -135.37 (t, *J* = 47.6 Hz, 1F). ³¹P NMR (121 MHz, CDCl₃) δ -7.03. IT-TOF: *m/z* [M+H]⁺ calcd: 894.207, found: 894.131. *m/z* [M+CF₃COO]⁻ calcd: 1006.185, found: 1006.100.



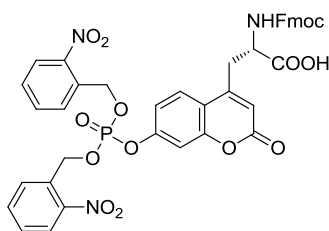
(*S*)-2-(((9H-fluoren-9-yl)methoxy)carbonylamino)-3-(7-(bis(2-nitrobenzyloxy)phosphoryloxy)-8-(fluoromethyl)-2-oxo-2H-chromen-4-yl)propanoic acid (**3-11**)

Compound **3-10** (220 mg, 0.25 mmol) was dissolved in DCM under an Ar atmosphere. Phenylsilane (107 mg, 1 mmol) and Pd(PPh₃)₄ (14 mg, 0.01 mmol) were added. After 5 h, the solution became brown dark, with TLC showing the completion of the reaction. The solvent was removed *in vacuo*, and the resulting residue was purified by flash chromatography to give the desired product **3-11** (190 mg, 90%). ¹H NMR (500 MHz, *d*₆-DMSO) δ 8.11 (dd, *J* = 8.1, 2.2 Hz, 2H), 7.96 (d, *J* = 7.2 Hz, 1H), 7.86 (dd, *J* = 8.0, 3.3 Hz, 2H), 7.74 - 7.78 (m, 2H), 7.68 - 7.70 (m, 2H), 7.58 - 7.64 (m, 4H), 7.43 (d, *J* = 8.9 Hz, 1H), 7.38 (t, *J* = 6.5 Hz, 2H), 7.27 (q, *J* = 7.3 Hz, 2H), 6.45 (s, 1H), 5.65 (d, *J* = 7.6 Hz, 4H), 5.57 (d, *J* = 47.8 Hz, 2H), 4.31 - 4.37 (m, 1H), 4.25 (d, *J* = 7.1 Hz, 2H), 4.14 - 4.21 (m, 2H), 3.24 (ddd, *J* = 25.4, 14.7, 7.2 Hz, 2H). ¹³C NMR (75 MHz, *d*₆-DMSO) δ 173.41, 159.69, 157.02, 153.89, 153.17, 152.27, 147.93, 144.74, 141.72, 135.33, 131.90, 131.80, 130.75, 130.08, 128.65, 128.05, 126.18, 125.97, 121.12, 117.34, 116.83, 115.83, 115.70, 73.62 (d, *J* = 162.2 Hz, 1C, C-F), 67.92, 66.80, 53.55, 47.61, 33.61. ¹⁹F NMR (282 MHz, *d*₆-DMSO) δ -133.34 (t, *J* = 47.9 Hz, 1F). ³¹P NMR (121 MHz, *d*₆-DMSO) δ -6.93. IT-TOF: *m/z* [M-H]⁻ calcd: 852.161, found: 852.165.



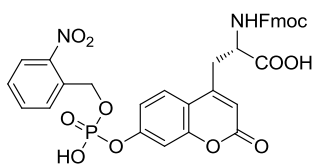
(2*S*)-2-(((9*H*-fluoren-9-yl)methoxy)carbonylamino)-3-(8-(fluoromethyl)-7-(hydroxyl (2-nitrobenzyloxy)phosphoryloxy)-2-oxo-2*H*-chromen-4-yl)propanoic acid (**3-2**)

Compound **3-11** (20 mg, 23 μ mol) was dissolved in 1 mL DCM. 500 μ L of 1 M thiophenol and triethylamine were subsequently added. After 1 h, the solvent was removed under reduced pressure. The residue was purified by preparative HPLC, giving the desired product **3-2** (10 mg, 60%). ^1H NMR (500 MHz, d_6 -DMSO) δ 8.09 (d, J = 8.1 Hz, 1H), 7.92 (d, J = 8.9 Hz, 1H), 7.86 (d, J = 7.8 Hz, 2H), 7.70 - 7.78 (m, 2H), 7.62 (dd, J = 12.1, 7.6 Hz, 2H), 7.55 (t, J = 8.3 Hz, 1H), 7.47 (d, J = 8.9 Hz, 1H), 7.39 (dd, J = 7.3, 5.2 Hz, 2H), 7.28 (q, J = 7.7 Hz, 2H), 6.38 (s, 1H), 5.55 (d, J = 47.9 Hz, 2H), 5.42 (d, J = 7.9 Hz, 2H), 4.32 - 4.39 (m, 1H), 4.25 (d, J = 7.2 Hz, 2H), 4.18 (t, J = 7.0 Hz, 1H), 3.23 (ddd, J = 25.2, 14.6, 7.3 Hz, 2H). ^{13}C NMR (125 MHz, d_6 -DMSO) δ 172.35, 158.89, 158.06, 155.90, 153.20, 152.93, 152.28, 146.69, 143.65, 140.61, 134.06, 132.57, 132.51, 128.94, 128.59, 127.54, 127.10, 126.96, 125.10, 124.64, 120.00, 116.00, 114.88, 114.31, 113.72, 73.27, 71.98, 65.69, 64.94, 52.47, 46.51, 32.53. ^{19}F NMR (282 MHz, d_6 -DMSO) δ -132.67 (t, J = 47.9 Hz, 1F). ^{31}P NMR (121 MHz, d_6 -DMSO) δ -6.67. IT-TOF: m/z $[\text{M}+\text{H}]^+$ calcd: 719.144, found: 719.118. m/z $[\text{M}-\text{H}]^-$ calcd: 717.129, found: 717.092.



(*S*)-2-(((9H-fluoren-9-yl)methoxy)carbonylamino)-3-(7-(bis(2-nitrobenzyloxy)phosphoryloxy)-2-oxo-2H-chromen-4-yl)propanoic acid (**3-1-2**)

Prepared according to procedures for preparation of compound **3-8** using di(2-nitrobenzyl) chlorophosphate [prepared from di(2-nitrobenzyl) phosphate (0.55 g, 1.5 mmol)], compound **3-6** (0.38 g, 0.75 mmol), triethylamine (0.38 g, 1.88 mmol). The allyl group of the phosphorylated product was deprotected using Pd(PPh₃)₄ and PhSiH₃, and purified by flash chromatography to give the desired product **3-1-2** (276 mg, two steps in 45% yield). ¹H NMR (500 MHz, *d*₆-DMSO) δ 8.11 (d, *J* = 8.1 Hz, 2H), 7.84 - 7.89 (m, 3H), 7.76 (t, *J* = 7.4 Hz, 2H), 7.69 (d, *J* = 7.6 Hz, 2H), 7.57 - 7.65 (m, 4H), 7.39 (t, *J* = 7.3 Hz, 2H), 7.26 - 7.30 (m, 4H), 6.39 (s, 1H), 5.63 (d, *J* = 7.7 Hz, 4H), 4.31 - 4.37 (m, 1H), 4.25 (d, *J* = 7.2 Hz, 2H), 4.17 (t, *J* = 7.1 Hz, 1H), 3.22 (ddd, *J* = 25.3, 14.6, 7.3 Hz, 2H). ¹³C NMR (125 MHz, *d*₆-DMSO) δ 172.88, 159.70, 156.45, 154.42, 152.39, 147.40, 144.18, 141.16, 134.72, 131.38, 131.32, 130.15, 129.58, 128.08, 127.48, 126.94, 125.62, 125.38, 120.55, 117.05, 116.74, 115.09, 108.94, 67.14, 66.23, 53.00, 47.04, 33.01. ³¹P NMR (121 MHz, *d*₆-DMSO) δ -6.70. IT-TOF: *m/z* [M-H]⁻ calcd: 820.154, found: 820.055.



(*2S*)-2-(((9H-fluoren-9-yl)methoxy)carbonylamino)-3-(7-(hydroxy(2-nitrobenzyloxy)phosphoryloxy)-2-oxo-2H-chromen-4-yl)propanoic acid (**3-1**)

Prepared according to procedures for the preparation of compound **3-2** using compound **3-1-2** (18 mg, 22 μmol), triethylamine and thiophenol (500 μmol each). The product was purified by preparative HPLC (9.8 mg, 65%). ^1H NMR (500 MHz, d_6 -DMSO) δ 8.08 (d, $J = 8.2$ Hz, 1H), 7.86 (t, $J = 6.8$ Hz, 2H), 7.79 - 7.81 (m, 1H), 7.70 - 7.77 (m, 2H), 7.60 - 7.64 (m, 2H), 7.52 - 7.59 (m, 1H), 7.39 (t, $J = 6.9$ Hz, 2H), 7.28 (dd, $J = 15.0, 7.5$ Hz, 2H), 7.20 - 7.21 (m, 2H), 6.33 (s, 1H), 5.41 (d, $J = 7.8$ Hz, 2H), 4.30 - 4.36 (m, 1H), 4.23 - 4.26 (m, 2H), 4.15 - 4.20 (m, 1H), 3.21 (ddd, $J = 25.3, 14.6, 7.3$ Hz, 2H). ^{13}C NMR (75 MHz, d_6 -DMSO) δ 172.45, 159.47, 155.93, 153.98, 152.08, 146.73, 143.61, 140.63, 134.09, 132.63, 132.52, 128.97, 128.65, 127.58, 126.99, 125.94, 125.06, 124.68, 120.05, 116.62, 114.90, 113.75, 107.91, 65.71, 64.87, 52.51, 46.51, 32.52. ^{31}P NMR (121 MHz, d_6 -DMSO) δ -6.59. IT-TOF: m/z $[\text{M}+\text{H}]^+$ calcd: 687.138, found: 687.106. m/z $[\text{M}-\text{H}]^-$ calcd: 685.122, found: 685.088.

6.6 Solid phase synthesis of three localization peptides

6.6.1 General information

All peptide synthesis described herein are carried out using standard Fmoc chemistry. HBTU, HOBt, and Fmoc-protected amino acids were purchased from GL Biochem (Shanghai, China). Fmoc-protecting amino acids with side chain protecting groups which normally could be deprotected by TFA are used. IRORI system and MacroKansTM were used for the synthesis of individual peptide sequences.

6.6.2 General procedure for solid phase synthesis

Next step is following Scheme 3.3.

Procedures for Fmoc deprotection a

The Fmoc-protected amino-functionalized resin was treated with 20% piperidine in DMF for 20 min at room temperature. After that, the resin was washed with DMF (3x), DCM (3x), DMF (3x) and DCM (3x).

Procedures for condition **b**

The Fmoc-protected amino acid or palmitic acid (4 equiv) was pre-activated using HOBT (4 equiv), HBTU (4 equiv), DIEA (8 equiv) in DMF for 10 min. Then the resin was put into the solution and shaken for 4 h or overnight. Next, the resin was washed with DMF (3x), DCM (3x) and DMF (3x).

Procedures for condition **c**

The Fmoc-protected unnatural amino acid, compound **3-11** (4 equiv) was pre-activated using HOAT (4 equiv), HATU (4 equiv), DIEA (8 equiv) in DMF for 10 min. Then the resin was put into the solution and shaken for 4 h or overnight. Next, the resin was washed with DMF (3x), DCM (3x) and DMF (3x).

Procedures for condition **d**

The Fmoc-protected amino-functionalized resin was treated with 2% DBU in DMF for 20 min at room temperature. After that, the resin was washed with DMF (3x), DCM (3x) and DMF (3x). If next step is using DCM as solvent, then the resin was washed with DCM (3x) again.

Procedures for condition **e**

The resin was swelled in DCM and washed with DCM. Then DIEA (10 equiv) was added to the reactor, followed by addition of acetic anhydride (5 equiv). The reaction

was shaken for 2 h. Then the resin was washed with DMF (3x), DCM (3x), DMF (3x), MeOH (3x), and dried *in vacuo*.

Procedures for condition f

The resin was swelled in DCM and washed with DCM. Then 1% TFA in DCM was added to the reactor. The reaction was shaken for 30 min, then washed with DMF (3x), DCM (3x) and DMF (3x).

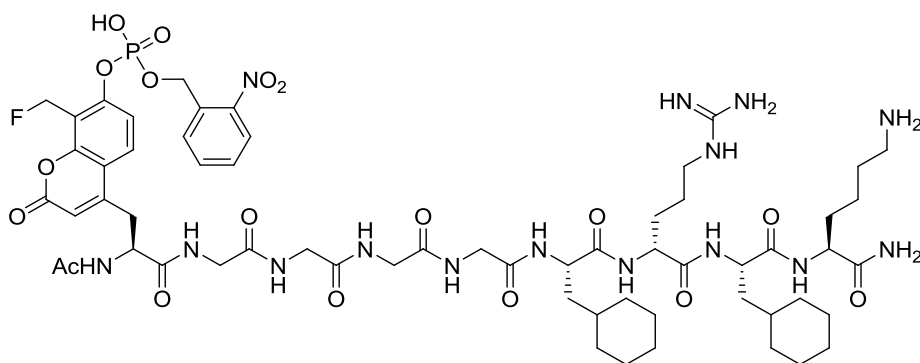
Procedures for condition g

A cleavage cocktail TFA/TIS/DCM (95:2.5:2.5) was added to the resin and the reaction was stirred for 2 h. Subsequently, the resin was filtered off, followed by removal of the solvent *in vacuo*. The resulting residue was precipitated with cold diethyl ether and purified by preparative HPLC.

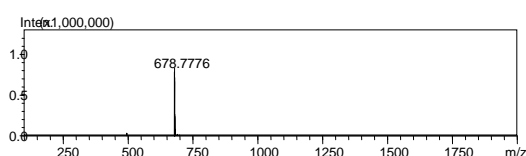
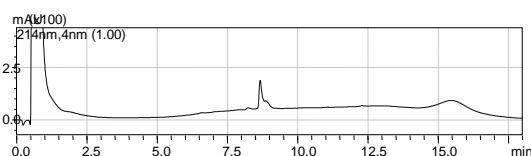
Mitochondrion probe (pMito)

Ac-X-GGGGF_xrF_xK-NH₂

IT-TOF-MS: Calcd: $[M+2H]/2 = 678.804$, found: 678.778



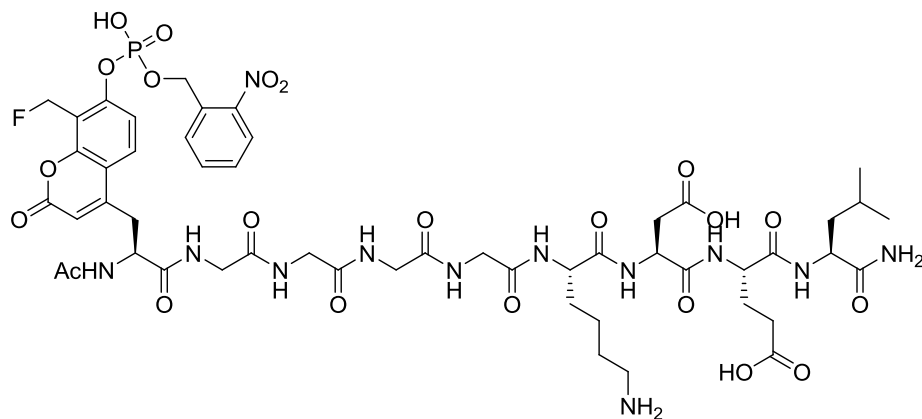
Exact Mass: 1355.6075



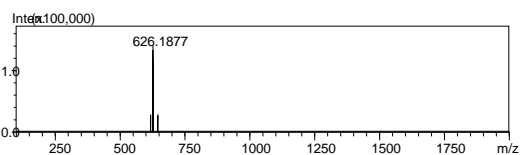
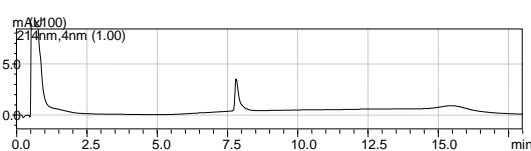
Endoplasmic Reticulum (pER)

Ac-X-GGGGKDEL-NH₂

IT-TOF-MS: Calcd: $[M+2H]/2 = 626.215$, found: 626.187



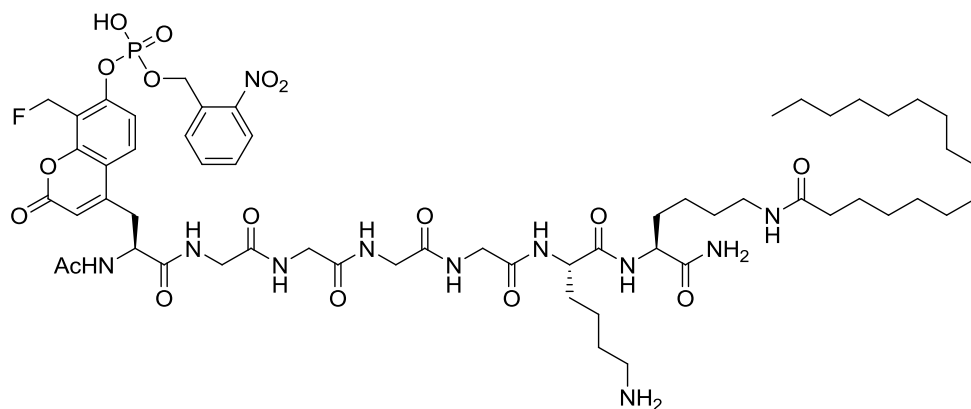
Exact Mass: 1250.4293



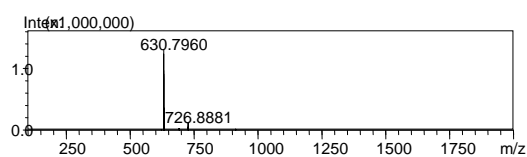
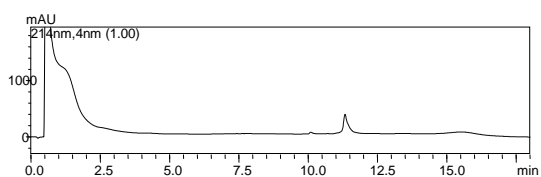
Membrane (pMem)

Ac-X-GGGGKK(palmitoyl)-NH₂

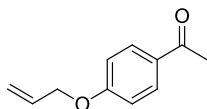
IT-TOF-MS: Calcd: $[M+2H]/2 = 630.800$, found: 630.796



Exact Mass: 1259.6003

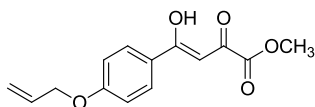


6.7 Synthesis of aldehyde and isonitrile for MCR reaction



1-(4-(allyloxy)phenyl)ethanone (**4-5**)

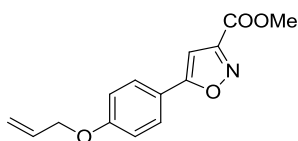
p-Hydroxyacetophenone (2.0 g, 14.7 mmol) was dissolved in acetonitrile (30 mL). K_2CO_3 (4.06 g, 29.7 mmol) and allyl bromide (2.13 g, 17.6 mmol) were added into the solution. The reaction was heated the reaction for 1 h at 50 °C, after which the organic solvent was removed under reduced pressure. The residue was dissolved in EtOAc and washed with water, brine. The organic layer was dried over Na_2SO_4 and the solvent was removed *in vacuo* to afford compound **4-5** (89%). 1H -NMR (300 MHz, $CDCl_3$) δ 7.89 (d, $J = 8.88$ Hz, 2H), 6.90 (d, $J = 8.88$ Hz, 2H), 5.95 - 6.08 (m, 1H), 5.39 (dd, $J_1 = 16.42$, $J_2 = 1.47$ Hz, 1H), 5.29 (dd, $J_1 = 10.51$ Hz, $J_2 = 1.32$ Hz, 1H), 5.56 (td, $J_1 = 5.28$ Hz, $J_2 = 1.47$ Hz, 2H), 2.51 (s, 3H). ^{13}C -NMR (75 MHz, $CDCl_3$) δ 196.5, 162.3, 132.4, 130.4, 118.0, 114.2, 68.7, 26.2. IT-TOF: m/z $[M+1]^+$ calcd: 177.08, found: 177.09



Methyl 4-(4-(allyloxy)phenyl)-4-hydroxy-2-oxobut-3-enoate (**4-6-1**)

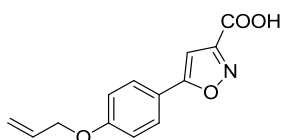
To a solution of compound **4-5** (2.0 g, 11.4 mmol) and dimethyl oxalate (1.47 g, 12.4 mmol) in MeOH (in an ice bath) under a nitrogen atmosphere was added a freshly prepared NaOMe (12.5 mmol, 0.5 M in MeOH) in small portions. The reaction was subsequently refluxed for 24 h before being cooled down to room temperature. Upon filtration of the white precipitate formed, the solution was collected, concentrated *in*

vacuo to provide the desired product **4-6-1** (48%). ¹H-NMR (300 MHz, CDCl₃) δ 7.97 (d, *J* = 8.85 Hz, 2H), 6.96 - 7.02 (m, 3H), 5.98 - 6.11 (m, 1H), 5.42 (dd, *J*₁ = 17.26 Hz, *J*₂ = 1.29 Hz, 1H), 5.32 (dd, *J*₁ = 10.44 Hz, *J*₂ = 1.23 Hz, 1H), 4.61 (d, *J* = 5.28 Hz, 2H), 3.92 (s, 3H). ¹³C-NMR (75 MHz, CDCl₃) δ 190.3, 167.8, 163.3, 162.9, 132.2, 130.3, 127.7, 118.3, 114.9, 97.8, 69.0, 53.1



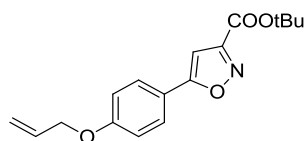
Methyl 5-(4-(allyloxy)phenyl)isoxazole-3-carboxylate (**4-6**)

Compound **4-6-1** (0.64 g, 2.4 mmol) and NH₂OH·H₂O (0.17 g, 2.9 mmol) dissolved in MeOH was added a catalytic amount of TsOH·H₂O (0.01 g, 0.05 mmol). The reaction mixture was refluxed for 24 h. After being cooled down, the resulting white precipitate was collected by suction filtration and washed with a mixture of ice-cold MeOH and deionized water to provide the desired product **4-6** as a solid (80%). ¹H-NMR (300 MHz, CDCl₃) δ 7.70 (d, *J* = 9.06 Hz, 2H), 6.97 (d, *J* = 9.06 Hz, 2H), 6.77 (s, 1H), 6.00-6.09 (m, 1H), 5.41 (dd, *J*₁ = 17.28 Hz, *J*₂ = 1.47 Hz, 1H), 5.30 (dd, *J*₁ = 10.53 Hz, *J*₂ = 1.17 Hz, 1H), 4.56 (d, *J* = 5.25 Hz, 2H), 3.97 (s, 3H) ¹³C-NMR (75 MHz, CDCl₃) δ 171.6, 160.5, 160.4, 156.5, 132.5, 127.4, 119.3, 118.0, 115.1, 98.4, 68.7, 52.7. IT-TOF: m/z [M+1]⁺ calcd: 260.08, found: 260.06



5-(4-(Allyloxy) phenyl) isoxazole-3-carboxylic acid (**4-7**)

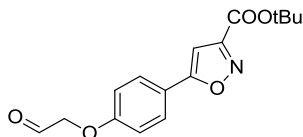
Compound **4-6** (1 g, 3.9 mmol) was dissolved in MeOH and water (1:1; 10 ml) in an ice bath. LiOH (0.47 g, 19.5 mmol) in 0.5 ml water was added slowly into the solution. After 2 h when TLC indicated the starting material was completely consumed, the solution was neutralized with 1 M HCl. The resulting white precipitate was collected, washed with water and chilled methanol to provide the desired compound **4-7** (89%). ¹H-NMR (300 MHz, *d*₆-DMSO) δ 7.87 (d, *J* = 8.70 Hz, 2H), 7.24 (s, 1H), 7.10 (d, *J* = 8.55 Hz, 2H), 5.99 - 6.11 (m, 1H), 5.41 (dd, *J*₁ = 17.26 Hz, *J*₂ = 1.15 Hz, 1H), 5.28 (dd, *J*₁ = 10.53 Hz, *J*₂ = 0.99 Hz, 1H), 4.65 (d, *J* = 5.10 Hz, 2H). ¹³C-NMR (75 MHz, *d*₆-DMSO) δ 170.8, 160.9, 160.1, 157.8, 133.2, 127.5, 119.0, 117.8, 115.4, 99.3, 68.4. IT-TOF: *m/z* [M+1]⁺ calcd: 246.07, found: 246.06



Tert-butyl 5-(4-(allyloxy)phenyl)isoxazole-3-carboxylate(**4-8**)

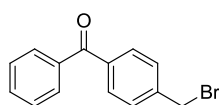
To a solution of compound **4-7** (500 mg, 2.0 mmol) in CHCl₃ was added *t*-BuOH (2 mL, 20 mmol) and pyridine (806 mg, 10.0 mmol) in an ice bath. After 5 min, POCl₃ (406 mg, 2.6 mmol) was slowly added into the solution. The reaction was stirred for 1 h. Dichloromethane and 10 mL dilute HCl were poured into the solution. The organic layer was separated, washed with dilute HCl (10 mL x 2), water (10 mL x 2), and brine (10 mL x 2), dried over Na₂SO₄. The solvent was removed and the residue was purified by silica gel chromatography (10% - 35% EtOAc in hexane) to afford compound **4-8** as a white solid (85%). ¹H-NMR (300 MHz, CDCl₃) δ 7.49 (d, *J* = 8.85 Hz, 2H), 6.99 (d, *J* = 9.02 Hz, 2H), 6.72 (s, 2H), 6.00 - 6.12 (m, 1H), 5.43 (dd, *J*₁ = 17.26, *J*₂ = 1.5 Hz, 1H), 5.32 (dd, *J*₁ = 10.53, *J*₂ = 1.32 Hz, 1H), 4.59 (d, *J* = 5.25 Hz, 2H), 1.63 (s, 9H). ¹³C-NMR (75 MHz, CDCl₃) δ 171.4, 160.5, 159.2, 158.1, 132.6,

127.5, 119.8, 118.1, 115.3, 98.6, 83.5, 68.9, 28.1. IT-TOF: m/z $[M+1]^+$ calcd: 302.13, found: 302.10



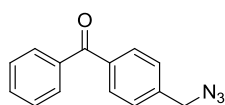
Tert-butyl 5-(4-(2-oxoethoxy)phenyl)isoxazole-3-carboxylate(**4-1**)

O₃ gas was bubbled into a solution of compound **4-8** (450 mg, 1.5 mmol) in DCM (10 mL) at -78 °C until the solution became blue. Excess O₃ was purged out using argon. Then, zinc dust (340 mg, 5.2 mmol) and a mixture of glacial AcOH and water (17:3; 360 μL) were added to the solution in small portions with vigorous stirring. Subsequently, the reaction mixture was allowed to warm to room temperature, followed by stirring for a further 1 h. The reaction mixture was then neutralized and filter through Celite to remove any residual solids. The filtrate was extracted with DCM and the organic layer was washed with saturated NaHCO₃ and brine, dried over Na₂SO₄ and concentrated *in vacuo*. The resulting crud product was purification by flash column chromatography (10% - 40% EtOAc in hexane) to afford the pure product **4-1** as a white solid (81%) ¹H-NMR (300 MHz, CDCl₃) δ 9.81 (s, 1H), 7.68 (d, *J* = 8.88 Hz, 2H), 6.94 (d, *J* = 8.88 Hz, 2H), 6.70 (s, 1H), 4.63 (d, *J* = 0.81 Hz, 2H), 1.99 (s, 9H) δ ¹³C-NMR (75 MHz, CDCl₃) δ 197.8, 170.8, 159.2, 158.9, 157.9, 127.5, 120.5, 115.0, 98.8, 83.5, 72.4, 27.9. IT-TOF: m/z $[M+1]^+$ calcd: 304.11, found: 304.09



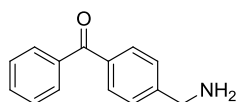
(4-(bromomethyl)phenyl)(phenyl)methanone (**4-9**)

4-Methylbenzophenone (3 g, 15.3 mmol) and *N*-bromosuccinimide (2.72 g, 15.3 mmol) were dissolved in CCl₄ (30 mL) solution. The reaction was placed under an IR lamp and the solution was refluxed for ~3 h when monitoring of the reaction by TLC indicated the complete consumption of the starting material. The resulting solution was cooled down and the solid filtered off. Upon concentration *in vacuo* to remove half of the solvent, the resulting solution was added hexane, and left standing for recrystallization to occur. The resulting solid was collected to give pure compound **4-9** (78%). ¹H-NMR (300 MHz, CDCl₃) δ 7.77 - 7.80 (m, 4H), 7.60 (t, *J* = 7.30 Hz, 1H), 7.46 - 7.51 (m, 4H), 4.53 (s, 2H) ¹³C-NMR (75 MHz, CDCl₃) δ 195.9, 142.0, 137.4, 137.3, 132.5, 130.5, 130.0, 128.9, 128.3, 32.2. IT-TOF: *m/z* [M+1]⁺ calcd: 275.00, 277.00 found: 274.99, 276.99.



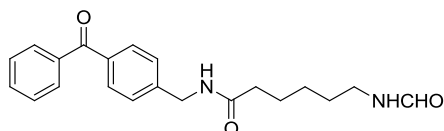
4-(Azidomethyl)phenyl(phenyl)methanone (**4-10**)

To a stirred solution of **4-9** (1.5 g, 5.5 mmol) in 16 mL mixture of acetone: H₂O (3:1) in an ice bath was added NaN₃ (0.7 g, 11 mmol; dissolved in water). After 1 h, some precipitates formed. Water (50 mL) was then added, and the resulting solution was filtered. The solid collected was washed with water and then dried *in vacuo* to afford compound **4-10** (96%). ¹H-NMR (300 MHz, CDCl₃) δ 7.80 (t, *J* = 8.49 Hz, 4H), 7.59 (t, *J* = 7.30 Hz, 1H), 7.41 - 7.50 (m, 4H), 4.43 (s, 2H). ¹³C-NMR (75 MHz, CDCl₃) δ 196.0, 139.8, 137.3, 132.5, 130.5, 130.0, 128.2, 127.7, 54.2. IT-TOF: *m/z* [M+1]⁺ calcd: 238.09 found: 238.07.



(4-(aminomethyl)phenyl)(phenyl)methanone (**4-11**)

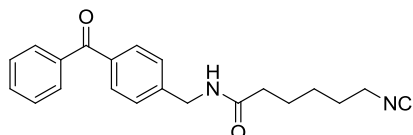
4-10 (0.5 g, 2.1 mmol) dissolved in THF (10 mL) was added H₂O (0.16 g) at room temperature. PPh₃ was then added and the resulting solution was heated to 60 °C for 1 h before being cooled down to 25 °C. Upon solvent removal under reduced pressure, the crude residue collected was subsequently purified by silica gel chromatography (30% EtOAc in hexane to 10% MeOH in DCM with 0.1% triethylamine) to give compound **4-11** (82%). ¹H-NMR (300 MHz, CDCl₃) δ 7.72 - 7.74 (d, *J* = 8.20 Hz, 4H), 7.53 (t, *J* = 7.25 Hz, 1H), 7.38 - 7.42 (m, 4H), 3.91 (s, 2H), 1.20 (s, 2H) ¹³C-NMR (75 MHz, CDCl₃) δ 196.2, 147.5, 137.5, 135.9, 132.1, 130.2, 129.7, 128.1, 126.7, 45.8.



N-(4-benzoylbenzyl)-6-formamido-hexanamide (**4-12**)

6-formamido-hexanoic acid (0.40 g, 2.5 mmol) was pre-activated with HOBT (0.37 g, 2.75 mmol), HBTU (1.0 g, 2.75 mmol) and DIEA (0.38 g, 3 mmol) in 15 mL DMF. After 10 min, compound **4-11** (0.53g, 2.5 mmol) was added into the solution. The reaction was stirred further for 2 h. Upon DMF removal *in vacuo*, the resulting residue was purified by silica gel chromatography (10% to 40% EtOAc in hexane) to yield compound **4-12** (84%). ¹H-NMR (300 MHz, CDCl₃) δ 8.06 (s, 1H), 7.15 (t, *J* = 8.05 Hz, 4H), 7.57 (t, *J* = 7.40 Hz, 1H), 7.45 (t, *J* = 7.24 Hz, 2H), 7.34 (d, *J* = 8.22 Hz, 2H), 6.77 (br, 1H), 6.34 (br, 1H), 4.47 (d, *J* = 5.91 Hz, 2H), 3.25 (q, *J* = 6.66 Hz, 2H), 2.24 (t, *J* = 7.39 Hz, 2H), 1.64 (qn, *J* = 7.45 Hz, 2H), 1.49 (qn, *J* = 7.11 Hz, 2H), 1.31

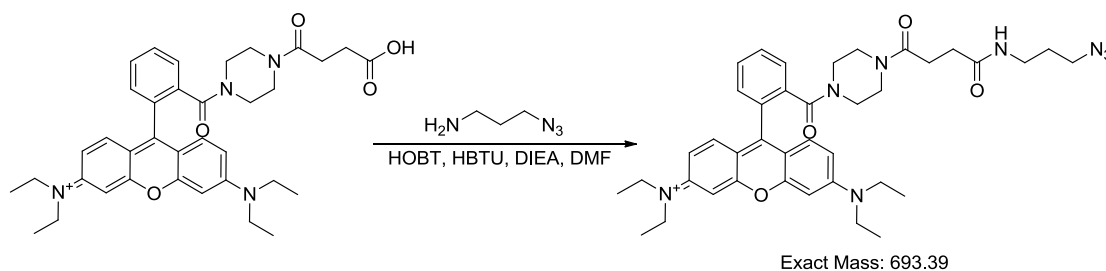
(qn, $J = 7.27$ Hz, 2H) ^{13}C -NMR (75 MHz, CDCl_3) δ 196.5, 173.3, 161.6, 143.3, 137.3, 136.4, 132.5, 130.3, 129.3, 128.3, 127.3, 43.1, 37.7, 36.1, 28.9, 26.2, 25.0. IT-TOF: m/z $[\text{M}+1]^+$ calcd: 354.18, found: 353.17



N-(4-benzoylbenzyl)-6-isocyano-hexanamide (**4-2**)

Compound **4-12** (1.0 g, 2.9 mmol) dissolved in anhydrous DCM in an ice bath was added triethylamine (1.86 g, 18.4 mmol) and POCl_3 (0.91 g, 5.9 mmol). The solution was stirred for 30 min. Subsequently, 5% NaHCO_3 was added and the reaction was stirred further for another 10 min. The resulting solution was extracted by DCM. The combine organic layer was washed with brine and dried over Na_2SO_4 . The crude residue was purified by silica gel chromatography (5% to 30% EtOAc in hexane) to provide compound **4-2** (81%). ^1H -NMR (300 MHz, CDCl_3) δ 7.71 - 7.76 (m, 4H), 7.58 (t, $J = 7.38$ Hz, 1H), 7.46 (t, $J = 7.20$ Hz, 2H), 7.35 (d, $J = 8.04$ Hz, 2H), 6.23 (br, 1H), 4.49 (d, $J = 5.91$ Hz, 2H), 3.34 - 3.40 (m, 2H), 2.25 (t, $J = 7.38$ Hz, 2H), 1.65 - 1.75 (m, 2H), 1.48 - 1.50 (qn, $J = 7.11$ Hz, 2H). ^{13}C -NMR (75 MHz, CDCl_3) δ 196.3, 172.6, 155.9, 143.2, 137.4, 136.6, 132.5, 130.4, 129.9, 128.3, 127.4, 43.1, 41.3, 36.0, 28.7, 25.9, 24.6. IT-TOF: m/z $[\text{M}+1]^+$ calcd: 335.17, found: 335.15

6.8 Synthesis of TER-azide

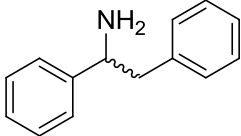
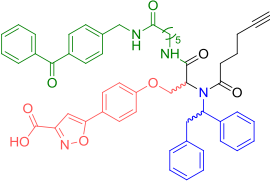
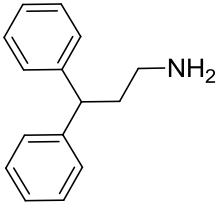
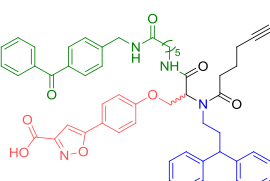
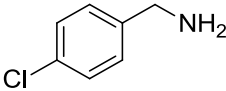
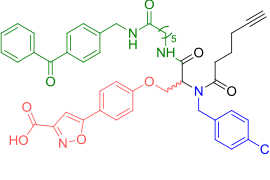
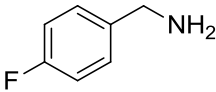
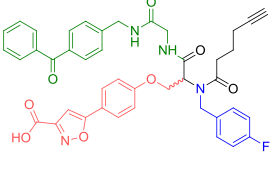
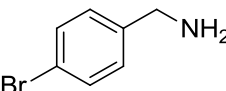
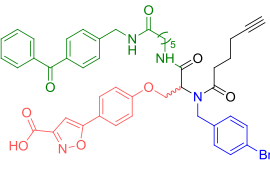
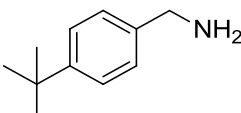
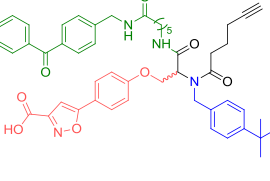
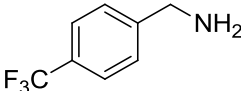
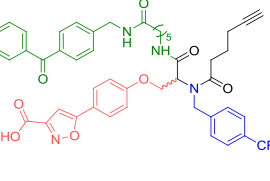


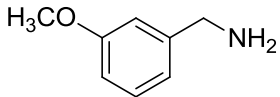
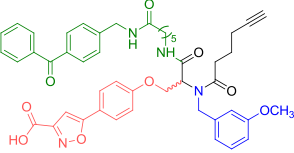
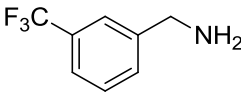
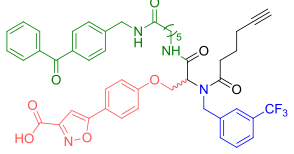
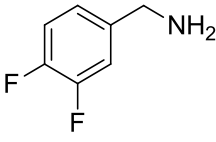
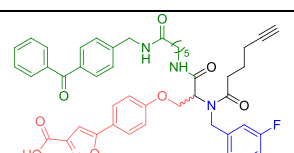
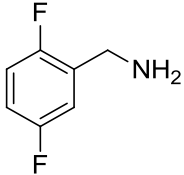
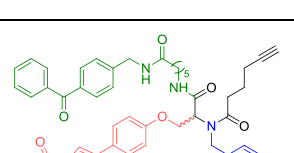
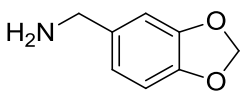
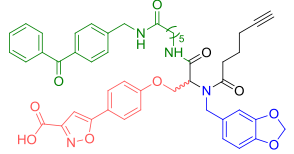
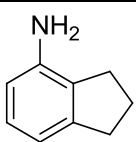
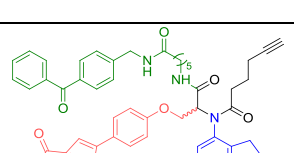
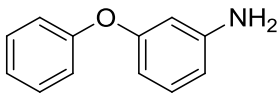
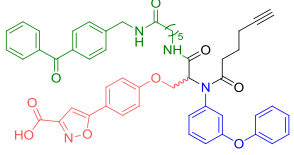
The acid (50 mg, 0.09 mmol; prepared based on reference 121) was pre-activated with HOBT (13.2 mg, 0.1 mmol), HBTU (37.2 mg, 0.1 mmol) and DIEA (18.55 mg, 0.1 mmol) in DMF for 10 min. Subsequently, 3-azidopropan-1-amine (9 mg, 0.09 mmol) was added and the resulting mixture was stirred for 2 h. Upon DMF removal *in vacuo*, the residue was purified by flash chromatography (0.1% to 5% Methanol in DCM) to afford the desired rhodamine-N₃ (85%). ESI: m/z [M+1]⁺ calcd: 694.4, found: 694.3.

6.8 MCR reaction

6.8.1 General procedure for MCR reaction

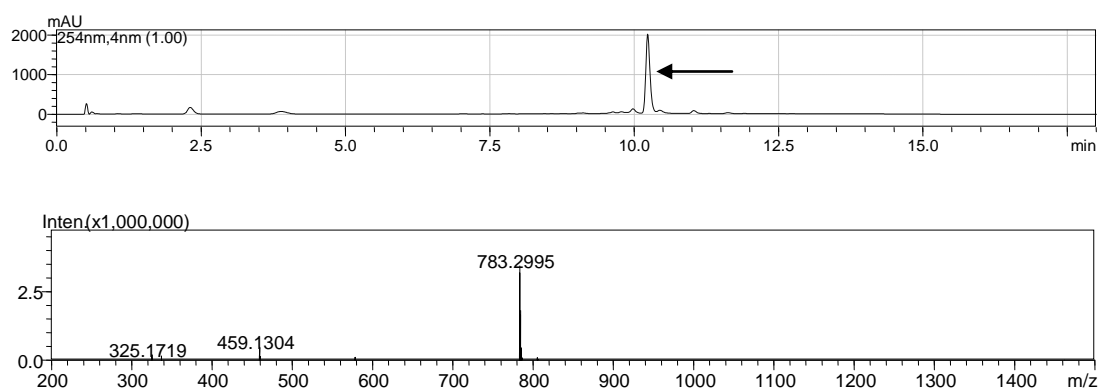
In 1.5 mL eppendorf tubes, a solution of amine (0.1 mmol, Figure 4.2) was added into 400 μ L MeOH. Then aldehyde **4-1** (0.1 mmol in 100 μ L DMF) was added. After 1 h, isonitrile **4-2** (0.1 mmol in 100 μ L MeOH) and 5-Hexynoic acid (0.1 mmol in 100 μ L MeOH) were added into the mixture. The reaction was stirred for ~9 h. The crude residue was purified by semi-preparative reverse-phase HPLC. The product was next treated with 1 mL of a mixture of TFA/DCM (50/50) for 2 h. TFA/DCM was next removed under reduced pressure with a GeneVac HT-4X Series II parallel evaporation system, affording the final product which was sufficiently pure and could be directly used in subsequent screening experiments. Yield (in two steps): 20 - 30%. Both LCMS and ¹H NMR were carried out to further characterize the final products and ensure the correct ID and purity (Table 6.1).

 <p>4g</p>	 <p>P7</p>	Calcd: 873.38 Found: 873.35	> 95%	N
 <p>4h</p>	 <p>P8</p>	Calcd: 887.39 Found: 887.37	> 90%	N
 <p>4i</p>	 <p>P9</p>	Calcd: 817.29 Found: 817.26	> 90%	Y
 <p>4j</p>	 <p>P10</p>	Calcd: 801.32 Found: 801.30	> 85%	Y
 <p>4k</p>	 <p>P11</p>	Calcd: 861.24 Found: 863.21 (Br)	> 85%	Y
 <p>4l</p>	 <p>P12</p>	Calcd: 839.39 Found: 839.37	> 90%	Y
 <p>4m</p>	 <p>P13</p>	Calcd: 851.32 Found: 851.29	> 95%	Y

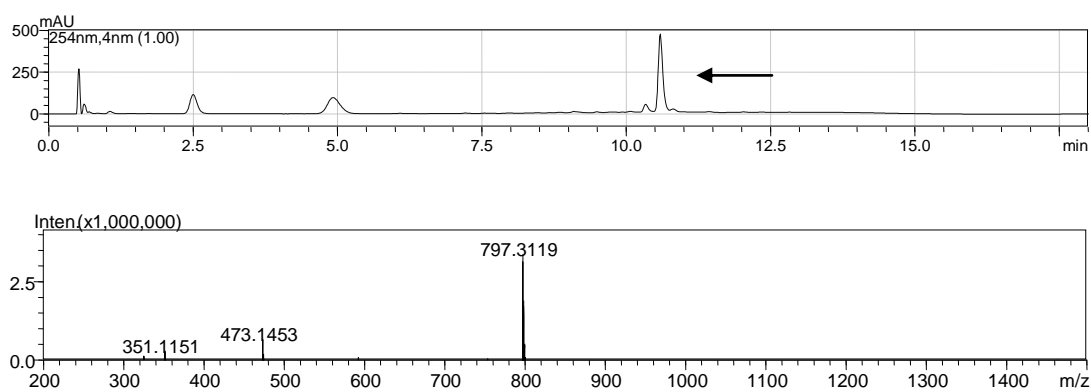
 <p>4n</p>	 <p>P14</p>	Calcd: 813.34 Found: 813.33	> 90%	Y
 <p>4o</p>	 <p>P15</p>	Calcd: 851.32 Found: 851.28	> 90%	Y
 <p>4p</p>	 <p>P16</p>	Calcd: 819.31 Found: 819.29	> 90%	Y
 <p>4q</p>	 <p>P17</p>	Calcd: 819.31 Found: 819.29	> 90%	N
 <p>4r</p>	 <p>P18</p>	Calcd: 827.32 Found: 827.30	> 85%	Y
 <p>4s</p>	 <p>P19</p>	Calcd: 809.35 Found: 809.32	> 95%	Y
 <p>4t</p>	 <p>P20</p>	Calcd: 861.34 Found: 861.30	> 95%	Y

6.8.2 LCMS Data

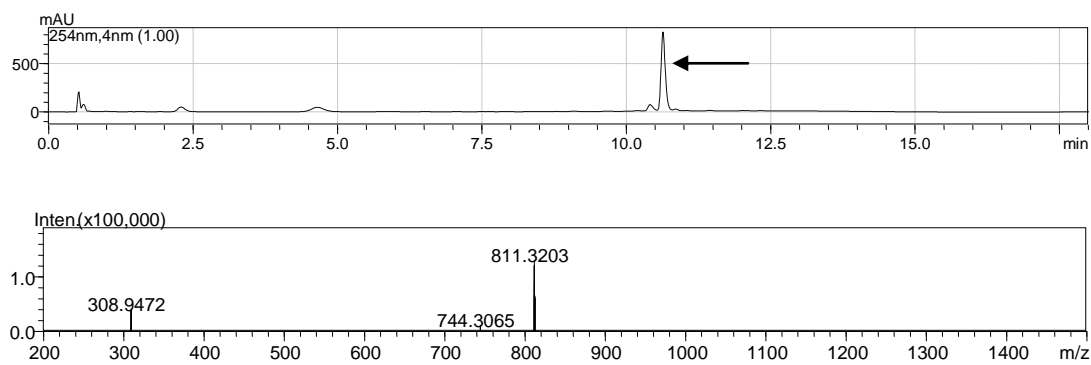
P1



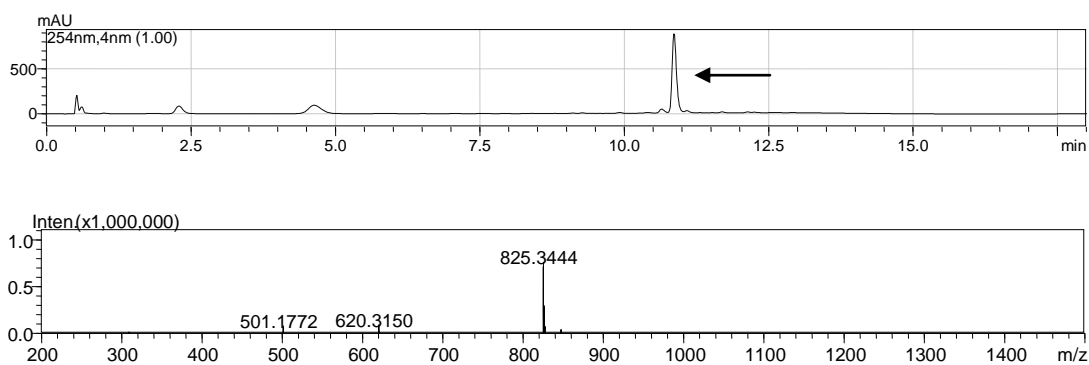
P2



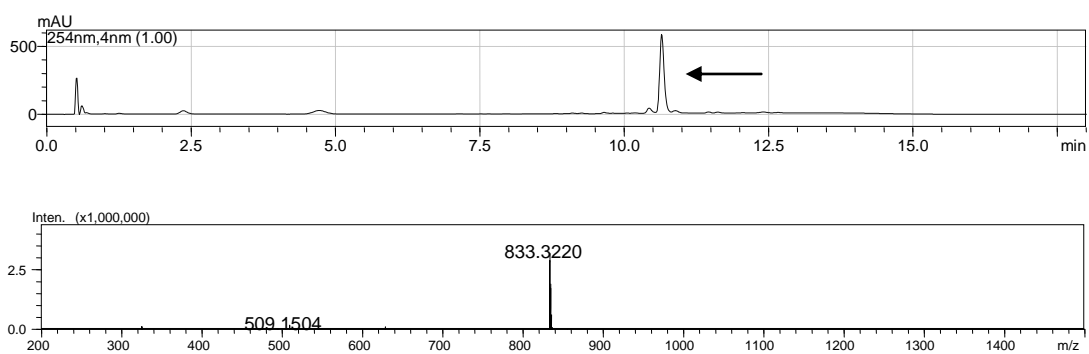
P3



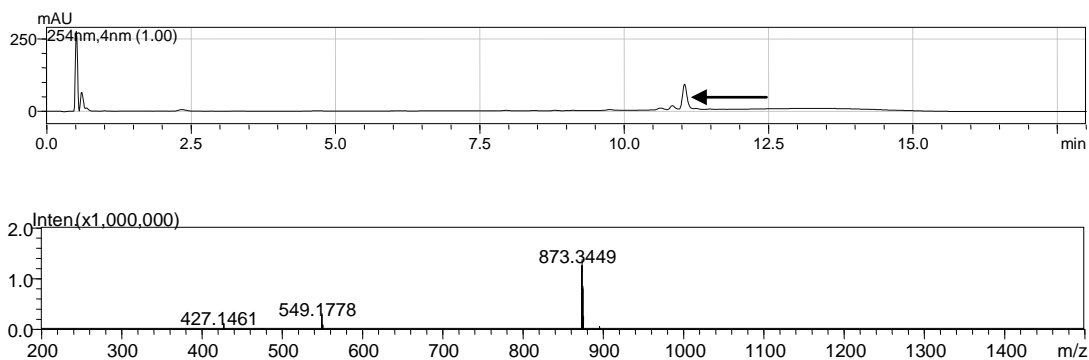
P4



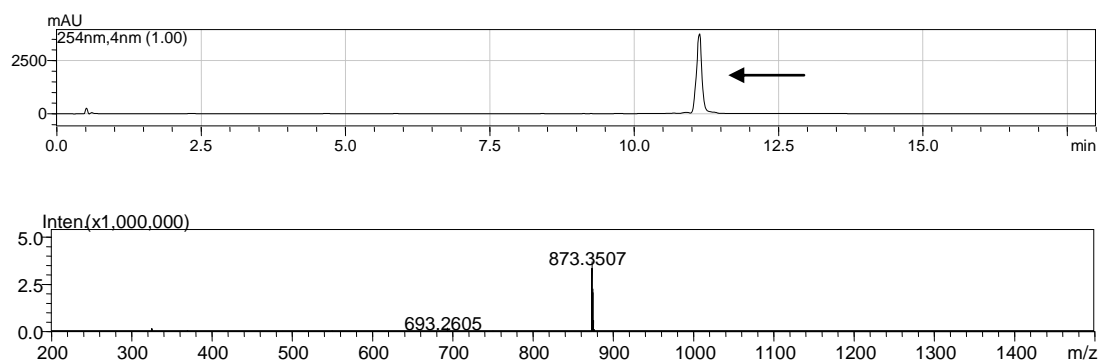
P5



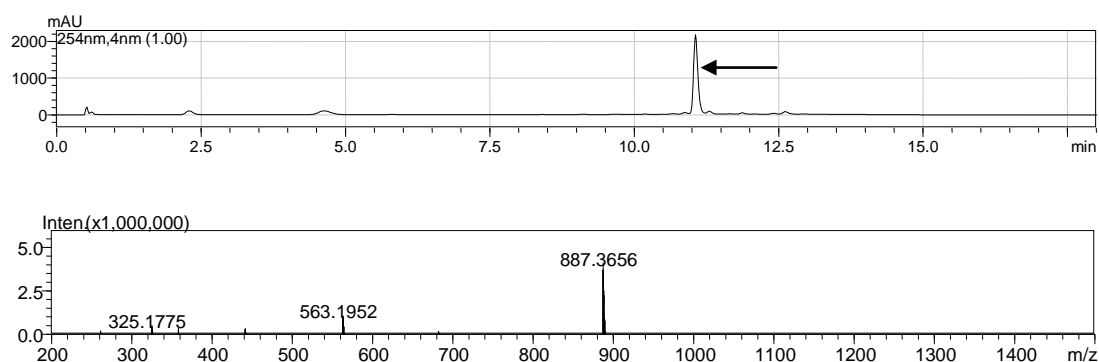
P6



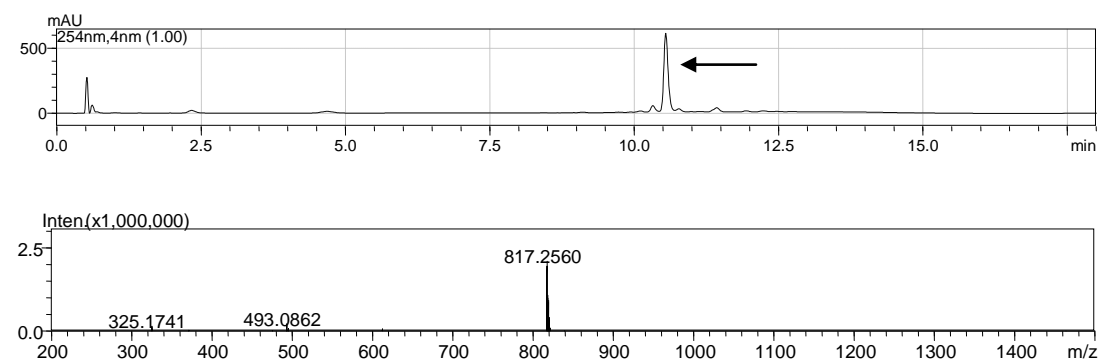
P7



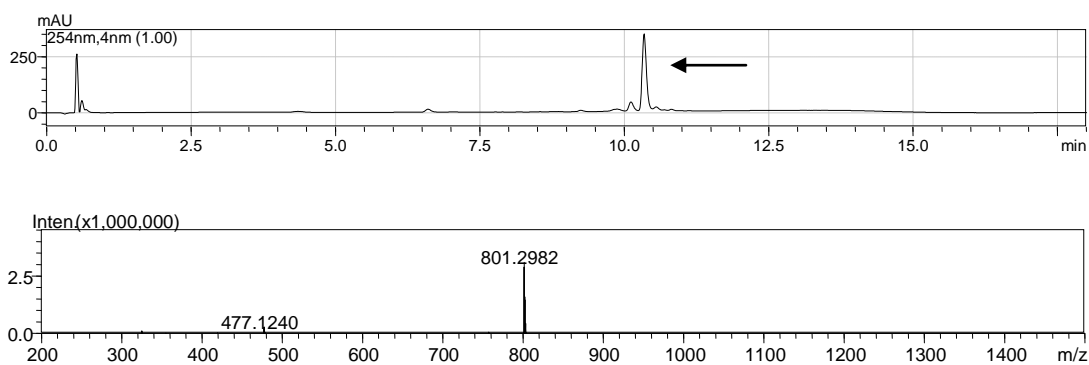
P8



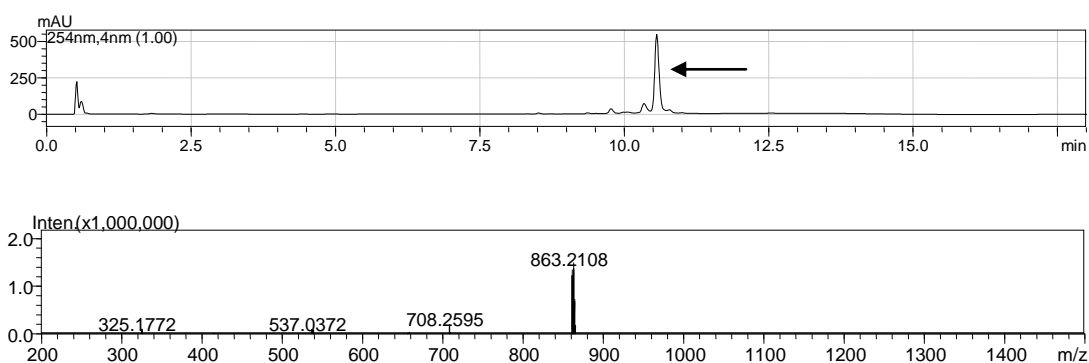
P9



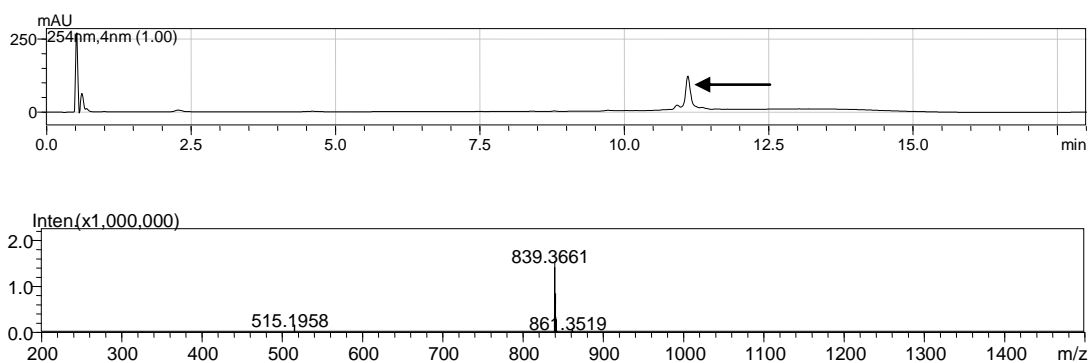
P10



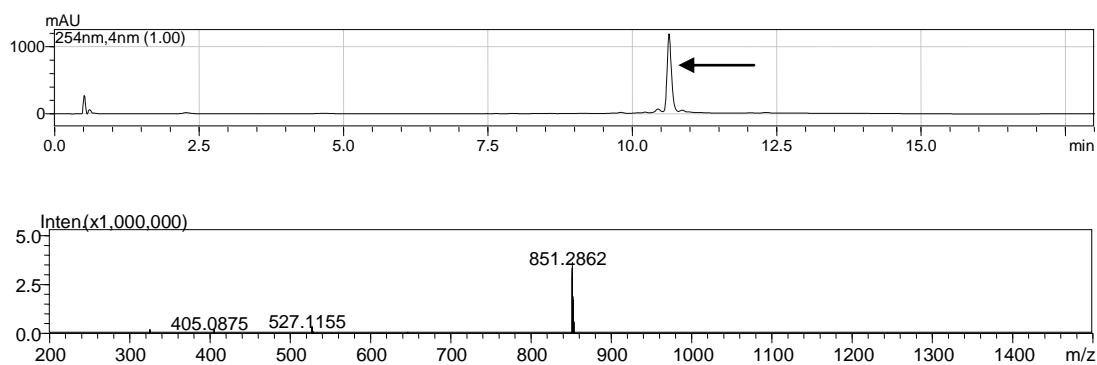
P11



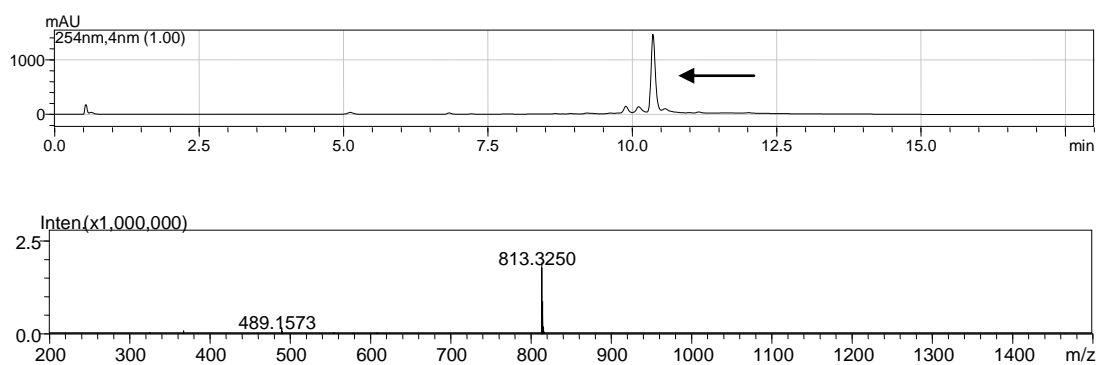
P12



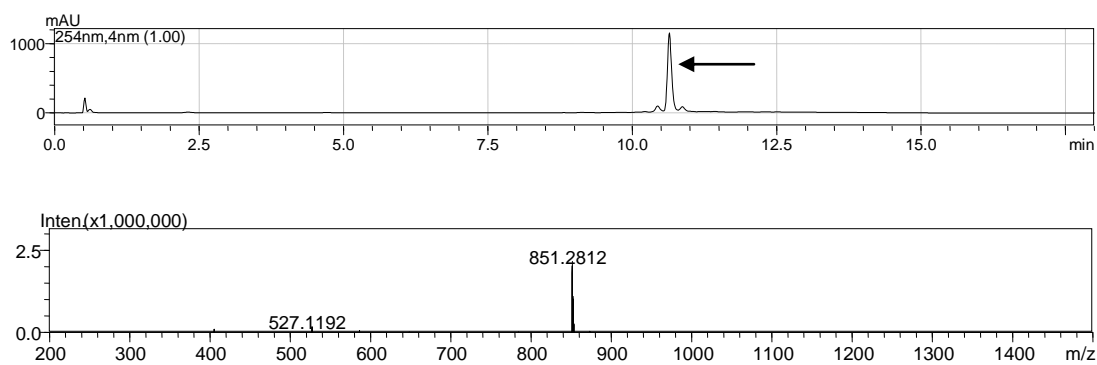
P13



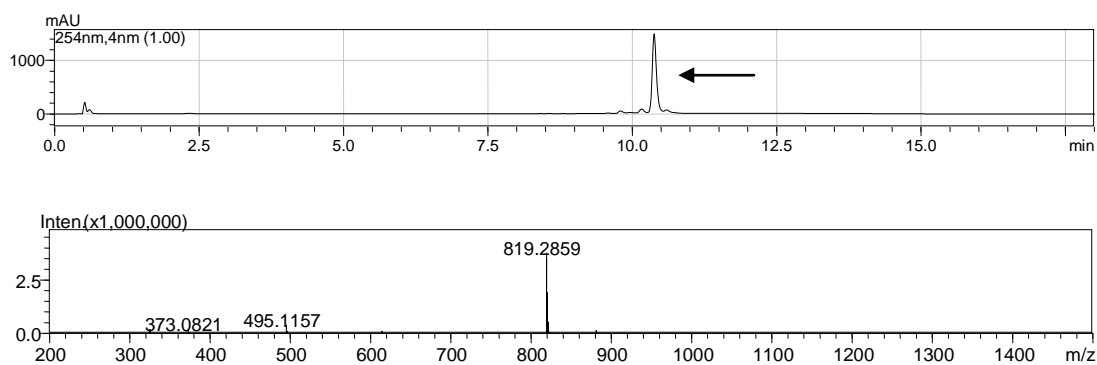
P14



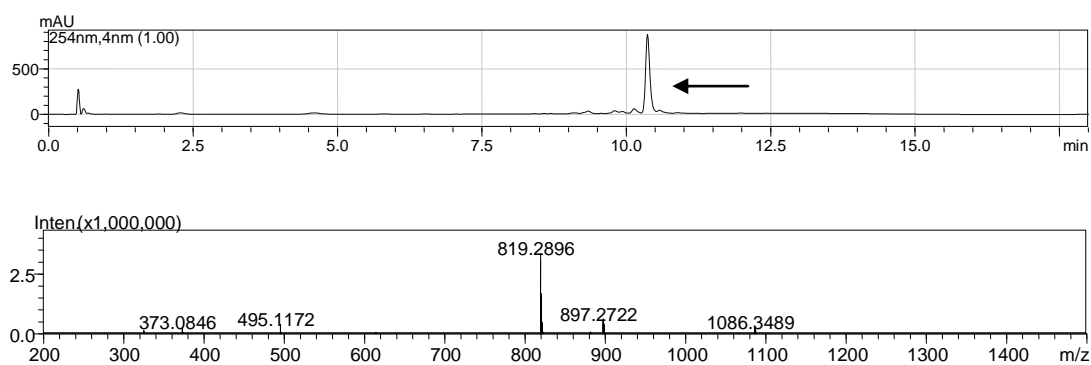
P15



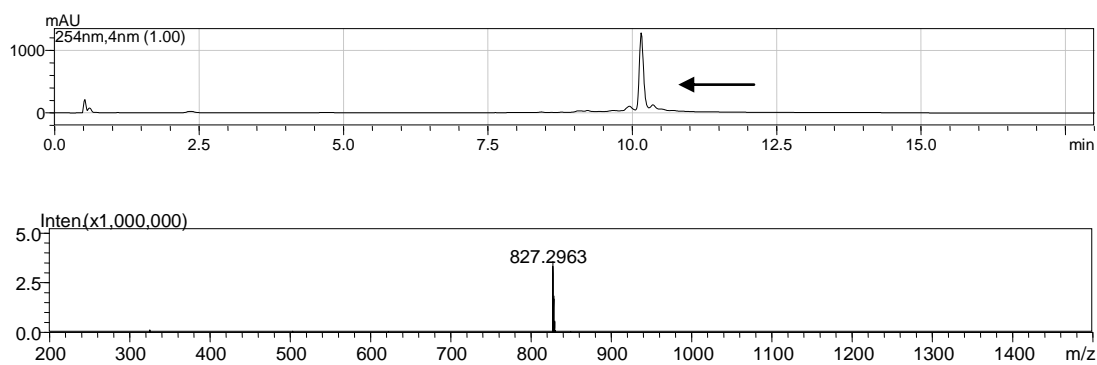
P16



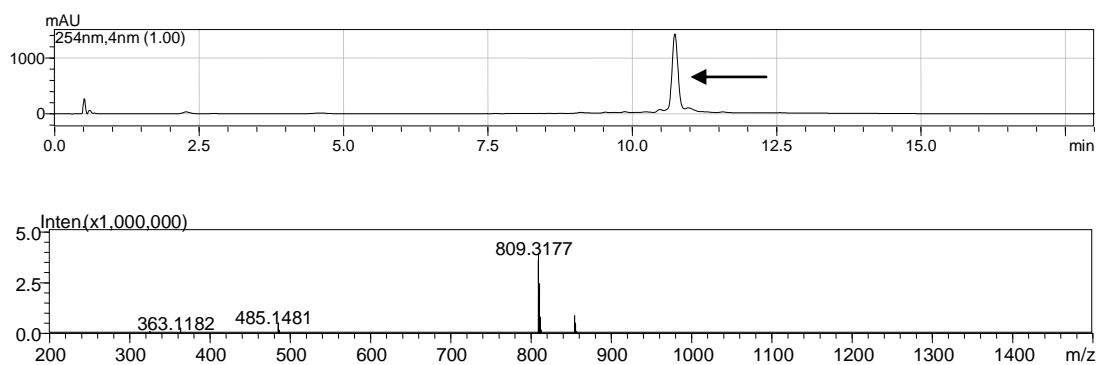
P17



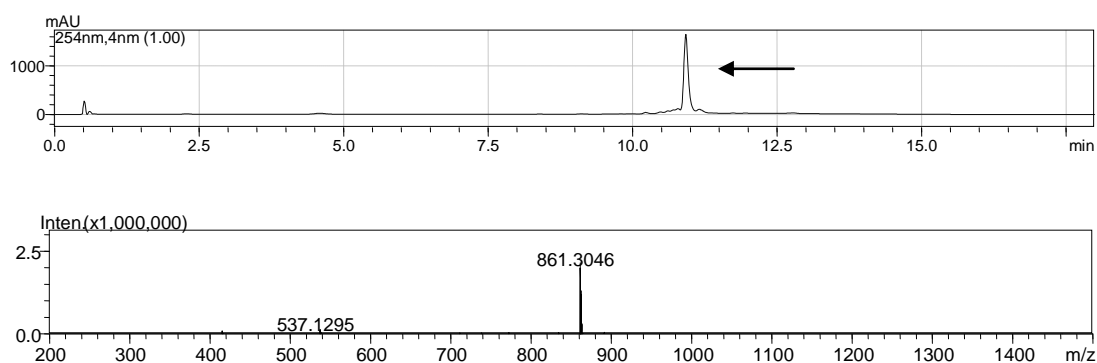
P18



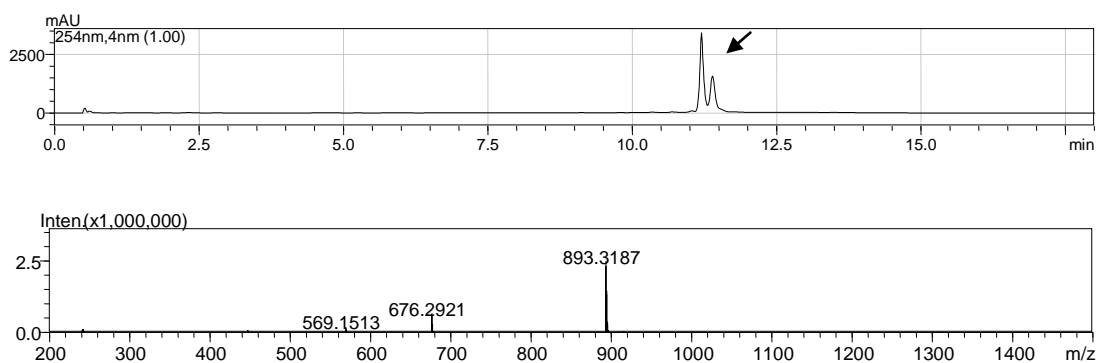
P19



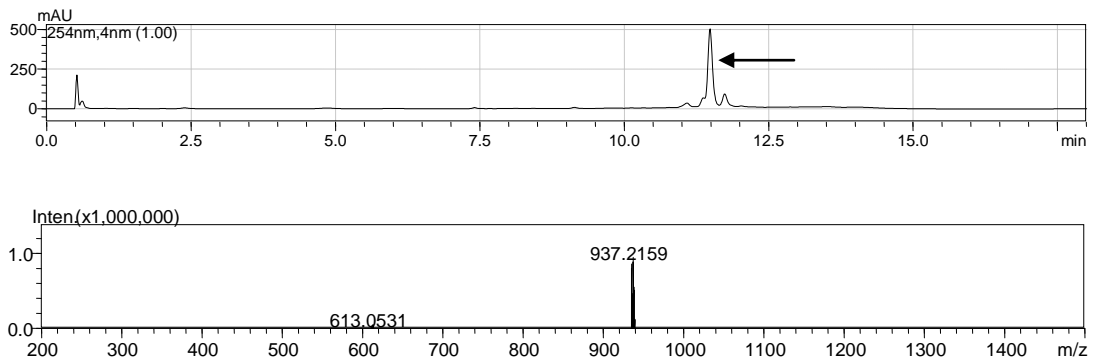
P20



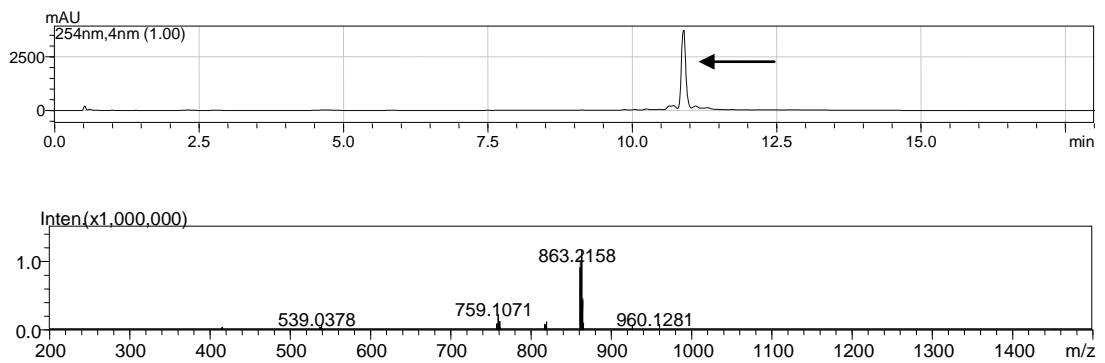
P21



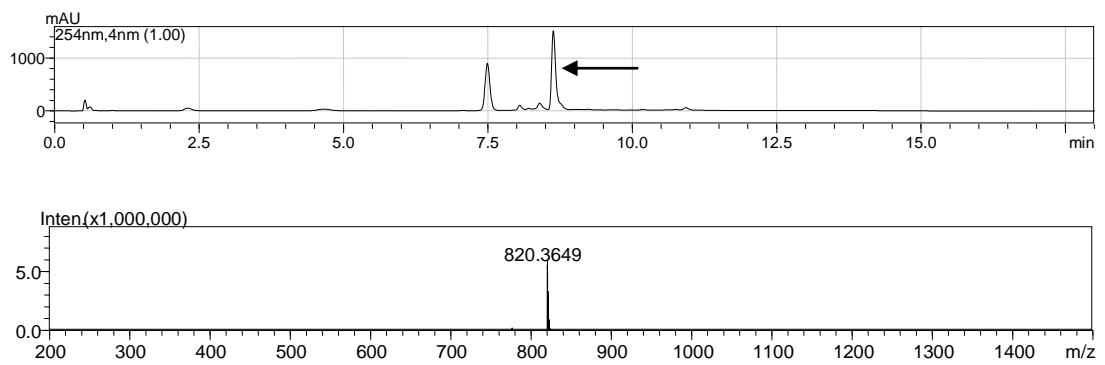
P22



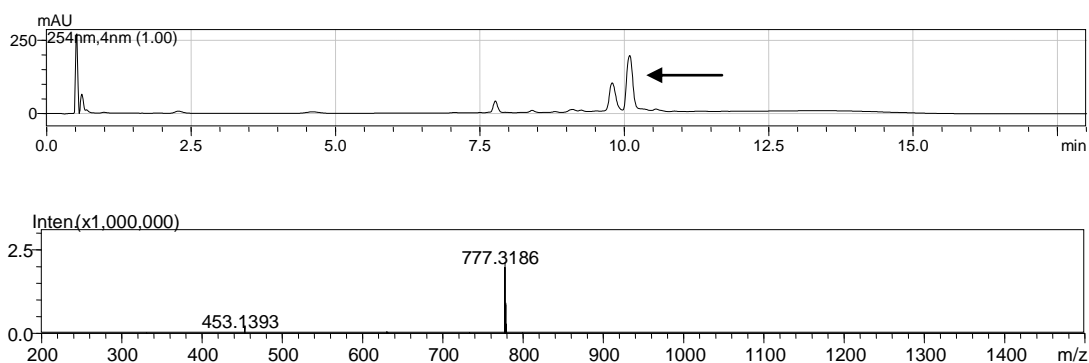
P23



P24



P25



6.8.3 ^1H NMR Data

P2

^1H -NMR (500 MHz, d_6 -DMSO) δ 8.41 - 8.42 (m, 1H), 7.91 - 8.28 (m, 1H), 7.65 - 7.72 (m, 7H), 7.55 (t, $J = 7.55$ Hz, 2H), 7.41 (d, $J = 7.55$ Hz, 2H), 7.11 - 7.29 (m, 6H), 6.88 (m, 2H), 4.80 - 5.04 (m, 1H), 4.23 - 4.46 (m, 4H), 3.53 (m, 2H), 3.06 - 3.11 (m, 2H), 2.78 - 2.85 (m, 3H), 2.11 - 2.21 (m, 6H), 1.69 (m, 2H), 1.50 - 1.51 (m, 2H), 1.39 - 1.42 (m, 2H), 1.33 (m, 2H). IT-TOF: m/z $[\text{M}+1]^+$ calcd: 797.35, found: 797.31.

P4

^1H -NMR (500 MHz, d_6 -DMSO) δ 8.41 (m, 1H), 7.82 - 8.20 (m, 1H), 7.65 - 7.72 (m, 7H), 7.55 (t, $J = 7.62$ Hz, 2H), 7.41 (d, $J = 8.20$ Hz, 2H), 7.07 - 7.24 (m, 5H), 6.88 (m, 3H), 4.75 - 4.90 (m, 1H), 4.22 - 4.36 (m, 4H), 3.03 (m, 2H), 2.76 (m, 1H), 2.50 (m, 2H), 2.15 (m, 4H), 1.67 (m, 2H), 1.41 - 1.52 (m, 8H), 1.23 (m., 2H). IT-TOF: m/z $[\text{M}+1]^+$ calcd: 825.38, found: 825.34.

P9

¹H-NMR (500 MHz, *d*₆-DMSO) δ 8.41 (m, 1H), 8.08 - 8.30 (m, 1H), 7.70 - 8.07 (m, 7H), 7.66 (m, 2H), 7.42 (m, 2H), 7.28 (m, 2H), 7.14 (m, 2H), 6.83 - 6.88 (m, 3H), 4.94 - 5.25 (m, 1H), 4.58 - 4.80 (m, 2H), 4.15 - 4.39 (m, 4H), 2.99 (m, 2H), 2.70 (m, 1H), 2.34 - 2.40 (m, 1H), 2.25 (m, 1H), 2.14 - 2.18 (m, 4H), 1.67 - 1.80 (m, 2H), 1.52 (m, 2H), 1.33 (m, 2H), 1.24 (m, 2H). IT-TOF: *m/z* [M+1]⁺ calcd: 817.29, found: 817.26.

P10

¹H-NMR (500 MHz, *d*₆-DMSO) δ 8.42 (m, 1H), 8.02 - 8.2 (m, 1H), 7.65 - 7.72 (m, 7H), 7.55 (t, *J* = 7.62 Hz, 2H), 7.41 (d, *J* = 8.20 Hz, 2H), 7.28 (m, 1H), 7.15 (m, 1H), 6.98 (dt, *J*₁ = 34.9, *J*₂ = 8.77 Hz, 2H), 6.81 (m, 3H), 4.89 - 5.21 (m, 1H), 4.63 - 4.78 (m, 2H), 4.09 - 4.37 (m, 4H), 2.98 (m, 2H), 2.68 (m, 1H), 2.23 - 2.32 (m, 2H), 2.08 - 2.15 (m, 4H), 1.65 - 1.80 (m, 4H), 1.51 (m, 2H), 1.33 (m, 2H), 1.22 (m, 2H). IT-TOF: *m/z* [M+1]⁺ calcd: 801.32, found: 801.30.

P11

¹H-NMR (500 MHz, *d*₆-DMSO) δ 8.42 (m, 1H), 8.02-8.28 (m, 1H), 7.66 - 7.72 (m, 7H), 7.55 (t, *J* = 7.65 Hz, 2H), 7.41 (m, 3H), 7.20 - 7.32 (m, 2H), 7.07 (m, 1H), 6.82 (m, 3H), 4.91 - 5.23 (m, 1H), 4.63 - 4.77 (m, 2H), 4.15 - 4.37 (m, 4H), 3.00 (m, 2H), 2.70 (m, 1H), 2.23 - 2.34 (m, 2H), 2.14 (m, 4H), 1.66 - 1.68 (m, 2H), 1.52 (m, 2H), 1.32 (m, 2H), 1.21 (m, 2H). IT-TOF: *m/z* [M+1]⁺ calcd: 861.24, found: 863.21. (Br isotope)

P12

¹H-NMR (500 MHz, *d*₆-DMSO) δ 8.42 (m, 1H), 7.96 - 8.22 (m, 1H), 7.63 - 7.96 (m, 7H), 7.55 (t, *J* = 7.55 Hz, 2H), 7.41 (d, *J* = 8.20 Hz, 2H), 7.22 (m, 1H), 7.04 - 7.15 (m, 3H), 6.75 (m, 3H), 4.89 - 5.17 (m, 1H), 4.67 (m, 1H), 4.03 - 4.36 (m, 4H), 2.97 (m, 2H), 2.68 (s, 1H), 2.24 - 2.40 (m, 2H), 2.08 - 2.14 (m, 4H), 1.67 - 1.76 (m, 2H), 1.51 (m, 2H), 1.34 (m, 2H), 1.14 - 1.19 (m, 11H). IT-TOF: *m/z* [M+1]⁺ calcd: 839.39, found: 839.37.

P13

¹H-NMR (500 MHz, *d*₆-DMSO) δ 8.41 (t, *J* = 5.80 Hz, 1H), 8.11 - 8.39 (m, 1H), 7.67 - 7.71 (m, 7H), 7.54 - 7.58 (m, 4H), 7.40 - 7.47 (m, 4H), 7.32 (d, *J* = 7.55 Hz, 1H), 6.84 (m, 2H), 4.96 - 5.30 (m, 1H), 4.77 - 4.91 (m, 2H), 4.16 - 3.37 (m, 4H), 2.93 - 2.99 (m, 2H), 2.66 - 2.69 (m, 1H), 2.23 - 2.26 (m, 1H), 2.15 (m, 1H), 2.13 - 2.14 (m, 4H), 1.68 - 1.79 (m, 2H), 1.50 (m, 2H), 1.32 (m, 2H), 1.22 (m, 2H). IT-TOF: *m/z* [M+1]⁺ calcd: 851.32, found: 851.27.

P14

¹H-NMR (500 MHz, *d*₆-DMSO) δ 8.41 (m, 1H), 8.03 - 8.26 (m, 1H), 7.63 - 7.71 (m, 7H), 7.55 (m, 2H), 7.41 (d, *J* = 8.20 Hz, 2H), 6.61 - 7.15 (m, 7H), 4.91 - 5.22 (m, 1H), 4.61 - 4.78 (m, 2H), 4.13 - 4.37 (m, 4H), 3.57 - 3.65 (m, 3H), 2.99 (m, 2H), 2.67 (m, 1H), 2.24 - 2.35 (m, 2H), 2.11 - 2.15 (m, 4H), 1.65 - 1.69 (m, 2H), 1.51 (m, 2H), 1.35 (m, 2H), 1.23 (m, 2H). IT-TOF: *m/z* [M+1]⁺ calcd: 813.34, found: 813.32.

P15

¹H-NMR (500 MHz, *d*₆-DMSO) δ 8.40 (m, 1H), 8.15 - 8.34 (m, 1H), 7.65 - 7.72 (m, 7H), 7.34 - 7.57 (m, 8H), 6.81 (m, 3H), 4.96 - 5.31 (m, 1H), 4.50 - 4.91 (m, 2H), 4.21 - 4.50 (m, 4H), 2.94 - 2.99 (m, 2H), 2.70 (m, 1H), 2.34 (m, 1H), 2.23 (m, 1H), 2.13 (m, 4H), 1.76 (m, 1H), 1.67 (m, 1H), 1.50 (m, 2H), 1.32 (m, 2H), 1.2 (m, 2H). IT-TOF: m/z [M+1]⁺ calcd: 851.32, found: 851.28.

P16

¹H-NMR (500 MHz, *d*₆-DMSO) δ 8.42 (m, 1H), 8.08 - 8.31 (m, 1H), 7.65 - 7.72 (m, 7H), 7.56 (t, *J* = 7.60 Hz, 2H), 7.41 (d, *J* = 8.20 Hz, 2H), 7.29 (m, 1H), 7.11 (m, 2H), 6.98 (m, 1H), 6.83 (m, 2H), 4.92 - 5.24 (m, 1H), 4.65 - 4.74 (m, 2H), 4.17 - 4.37 (m, 4H), 2.99 (m, 2H), 2.67 (m, 1H), 2.36 (m, 1H), 2.24 (m, 1H), 2.14 (m, 4H), 1.65 - 1.78 (m, 2H), 1.51 (m, 2H), 1.34 (m, 2H), 1.24 (m, 2H). IT-TOF: m/z [M+1]⁺ calcd: 819.31, found: 819.29.

P18

¹H-NMR (500 MHz, *d*₆-DMSO) δ 8.43 (m, 1H), 7.97 - 8.23 (m, 1H), 7.65 - 7.72 (m, 7H), 7.56 (t, *J* = 7.60 Hz, 2H), 7.41 (d, *J* = 8.20 Hz, 2H), 6.61 - 6.82 (m, 6H), 5.77 - 5.92 (m, 2H), 4.87 - 5.14 (m, 1H), 5.53 - 4.67 (m, 2H), 4.06 - 4.37 (m, 4H), 2.30 (m, 2H), 2.64 (m, 1H), 2.23 - 2.37 (m, 2H), 2.14 (m, 4H), 1.66 - 1.68 (m, 2H), 1.51 (m, 2H), 1.34 (m, 2H), 1.22 (m, 2H). IT-TOF: m/z [M+1]⁺ calcd: 827.32, found: 827.30.

P19

$^1\text{H-NMR}$ (500 MHz, d_6 -DMSO) δ 8.41 (m, 1H), 8.02 - 8.15 (m, 1H), 7.61 - 7.72 (m, 7H), 7.55 (t, $J = 7.55$ Hz, 2H), 7.41 (m, 3H), 7.12 - 7.22 (m, 2H), 6.93 - 7.03 (m, 1H), 6.73 (m, 1H), 5.21 - 5.26 (m, 1H), 3.84 - 4.36 (m, 4H), 2.81 - 2.88 (m, 4H), 2.68 - 2.71 (m, 2H), 1.98 - 2.18 (m, 8H), 1.52 - 1.63 (m, 4H), 1.40 (m, 2H), 1.25 (m, 2H). IT-TOF: m/z $[\text{M}+1]^+$ calcd: 809.35, found: 809.32.

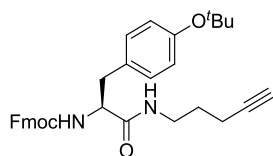
P20

$^1\text{H-NMR}$ (500 MHz, d_6 -DMSO) δ 8.41 (m, 1H), 8.10 (m, 1H), 7.65 - 7.72 (m, 7H), 7.55 (t, $J = 7.55$ Hz, 2H), 7.32 - 7.42 (m, 6H), 6.89 - 7.12 (m, 8H), 5.31 (m, 2H), 4.00 - 4.36 (m, 4H), 3.02 (m, 2H), 2.68 (s, 1H), 2.07 - 2.17 (m, 6H), 1.61 (m, 2H), 1.52 (m, 2H), 1.38 (m, 2H), 1.23 (m, 2H). IT-TOF: m/z $[\text{M}+1]^+$ calcd: 861.34, found: 861.30.

P23

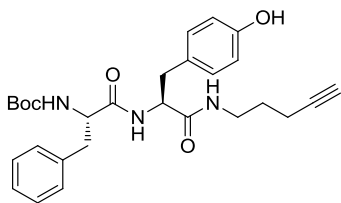
$^1\text{H-NMR}$ (500 MHz, d_6 -DMSO) δ 8.40 (m, 1H), 8.15 (m, 1H), 7.65 - 7.72 (m, 7H), 7.55 (m, 3H), 7.40 (m, 3H), 7.19 (s, 1H), 6.87 - 6.94 (m, 3H), 5.30 (m, 1H), 3.94 - 4.37 (m, 4H), 3.07(m, 2H), 2.68 (s, 1H), 2.05 - 2.23 (m, 9H), 1.53 - 1.63 (m, 4H), 1.41 (m, 2H), 1.27 (m, 2H). IT-TOF: m/z $[\text{M}+1]^+$ calcd: 861.24, found: 863.22. (Br isotope)

6.9 Synthesis of four probes



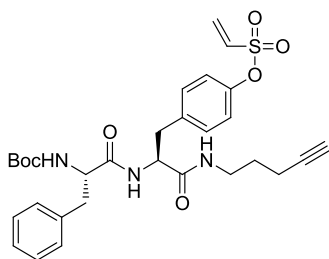
(*S*)-(9H-fluoren-9-yl)methyl (3-(4-(*tert*-butoxy)phenyl)-1-oxo-1-(pent-4-yn-1-ylamino)propan-2-yl)carbamate (**5-5**)

FmocNH-Tyr(*O*^{*t*}Bu)-COOH (1g, 2.17 mmol) was first activated by EDC (372 mg, 2.40 mmol), HOBT (323 mg, 2.40 mmol) and DIEA (486 μ L, 2.80 mmol) in DMF solution. Then the alkyne-based amine (460 mg, 2.3 mmol) was added into the solution. The solvent was evaporated after 1 h reaction. Then the residue was extracted by EA and washed by 10% citric acid, saturated NaHCO₃ and brine. The product was purified by chromatography column (94%, 1.1 g). ¹H NMR (500 MHz, CDCl₃) δ 7.75 (d, *J* = 7.6 Hz, 2H), 7.55 (dd, *J* = 7.4, 3.1 Hz, 2H), 7.39 (t, *J* = 7.5 Hz, 2H), 7.30 (t, *J* = 7.4 Hz, 2H), 7.07 (m, 2H), 6.91 (d, *J* = 8.2 Hz, 2H), 6.00 (s, 1H), 5.54 (s, 1H), 4.33 - 4.40 (m, 3H), 4.18 (t, *J* = 6.7 Hz, 1H), 3.20 - 3.30 (m, 2H), 2.96 - 3.05 (m, 1H), 2.05 - 2.11 (m, 2H), 1.90 (s, 1H), 1.51 - 1.65 (m, 2H), 1.31 (s, 9H). ¹³C NMR (126 MHz, CDCl₃) δ 170.80, 155.87, 154.39, 143.68, 143.66, 141.24, 131.18, 129.70, 127.69, 127.04, 124.93, 124.20, 119.93, 83.16, 78.38, 69.24, 66.95, 56.50, 47.08, 38.53, 38.16, 28.77, 27.75, 15.90. IT-TOF: *m/z* [M+H]⁺ calcd: 525.27, found: 525.25



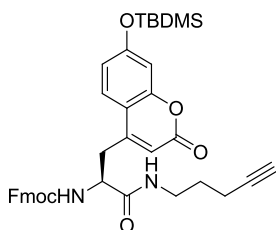
tert-butyl ((*S*)-1-(((*S*)-3-(4-hydroxyphenyl)-1-oxo-1-(pent-4-yn-1-ylamino)propan-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (**5-6**)

The *t*Bu group in **5-5** compound (0.3 mmol, 157 mg) was deprotected under TFA : DCM (1 : 1, 10 mL). After the reaction was finished, TFA and DCM were evaporated *in vacuo*. Piperidine (0.3 mL in 5 mL DMF, 3 mmol) was added into the resulted residue. After 30 min, DMF was also removed completely *in vacuo*. The residue was redissolved in DMF. BocNH-Phe-COOH (252 mg, 0.33 mmol) activated by EDC (51 mg, 0.33 mmol), HOBT (44 mg, 0.33 mmol), DIEA (76 μ L, 0.36 mmol) in DMF was dropwise added into the solution. DMF was evaporated after 1 h reaction. Then the residue was extracted by EA and washed by 10% citric acid, saturated NaHCO₃ and brine. The product was purified by chromatography column (67%, 100 mg). ¹H NMR (500 MHz, MeOD) δ 7.26 (t, *J* = 7.2 Hz, 2H), 7.17 - 7.21 (m, 3H), 6.98 (d, *J* = 8.0 Hz, 2H), 6.70 (d, *J* = 8.3 Hz, 2H), 4.44 (t, *J* = 6.8 Hz, 1H), 4.23 - 4.26 (m, 1H), 3.19 - 3.28 (m, 1H), 3.10 - 3.15 (m, 1H), 3.1 (dd, *J* = 13.9, 5.2 Hz, 1H), 2.90 (d, *J* = 4.1 Hz, 2H), 2.79 (dd, *J* = 13.6, 9.2 Hz, 1H), 2.22 (t, *J* = 2.6 Hz, 1H), 2.05 - 2.07 (m, 2H), 1.55 - 1.62 (m, 2H), 1.35 (s, 9H). ¹³C NMR (126 MHz, MeOD) δ 173.90, 173.07, 157.85, 157.44, 138.39, 131.40, 130.34, 129.42, 128.50, 127.74, 116.32, 84.15, 80.91, 70.03, 57.70, 56.09, 39.38, 38.84, 38.08, 29.18, 28.63, 16.53. IT-TOF: *m/z* [M+H]⁺ calcd: 494.25, found: 494.26



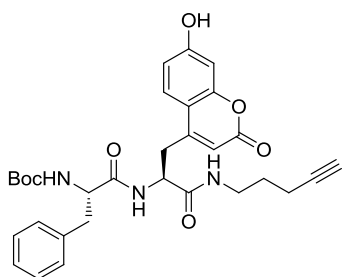
4-((*S*)-2-((*S*)-2-((*tert*-butoxycarbonyl)amino)-3-phenylpropanamido)-3-oxo-3-(pent-4-yn-1-ylamino)propyl)phenyl ethenesulfonate (**5-1**)

Compound **5-6** (50 mg, 0.1 mmol) and triethylamine (42 μ L, 0.3 mmol) was dissolved in anhydrous DCM at 0 $^{\circ}$ C. And 2-chloro-1-ethane-sulfonylchloride (11 μ L, 0.12 mmol) in DCM was dropwise added into the solution. After 2 h, the reaction was concentrated and the crude product was purified by chromatography column (**5-1**, 48 mg, 72%). ^1H NMR (500 MHz, CDCl_3) δ 7.21 - 7.31 (m, 3H), 7.15 (d, $J = 7.0$ Hz, 2H), 6.82 - 7.14 (m, 4H), 6.60 - 6.66 (m, 2H), 6.42 (s, 1H), 6.33 (d, $J = 16.6$ Hz, 1H), 6.14 (d, $J = 9.9$ Hz, 1H), 5.04 (s, 1H), 4.60 - 4.65 (m, 1H), 4.30 - 4.31 (m, 1H), 3.12 - 3.32 (m, 2H), 3.06 - 3.10 (m, 1H), 2.92 - 3.04 (m, 3H), 2.07 - 2.10 (m, 2H), 1.94 (t, $J = 2.6$ Hz, 1H), 1.55 - 1.61 (m, $J = 6.9$ Hz, 2H), 1.34 (s, 9H). ^{13}C NMR (126 MHz, CDCl_3) δ 171.21, 170.00, 155.73, 148.26, 136.05, 135.78, 131.98, 131.70, 130.71, 129.18, 128.75, 127.18, 122.31, 83.25, 80.70, 69.12, 56.02, 53.82, 38.53, 37.84, 37.09, 28.12, 27.77, 15.85. IT-TOF: m/z $[\text{M}+\text{H}]^+$ calcd: 583.24, found: 584.26



(*S*)-(9H-fluoren-9-yl)methyl (3-(7-((*tert*-butyldimethylsilyl)oxy)-2-oxo-2H-chromen-4-yl)-1-oxo-1-(pent-4-yn-1-ylamino)propan-2-yl)carbamate (**5-7**)

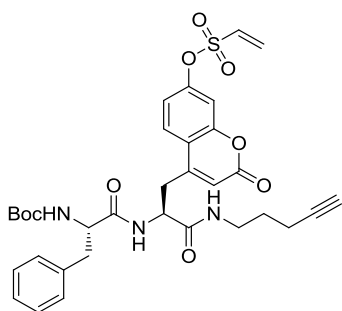
Prepared according to the general procedure using TBDMS protected amino acid (634 mg, 1.08 mmol), EDC (186 mg, 1.20 mmol), HOBT (161 mg, 1.20 mmol) and DIEA (243 μ L, 1.40 mmol) and 4-pentyn-1-amine (156 mg, 1.3 mmol). The product was purified by chromatography column (90%, 631 mg). ^1H NMR (500 MHz, CDCl_3) δ 7.74 (d, $J = 7.5$ Hz, 2H), 7.70 (d, $J = 8.5$ Hz, 1H), 7.55 (d, $J = 6.9$ Hz, 2H), 7.38 (t, $J = 7.4$ Hz, 2H), 7.24 - 7.34 (m, 2H), 6.77 - 6.82 (m, 2H), 6.70 - 6.72 (m, 1H), 6.22 (s, 1H), 6.03 (d, $J = 8.1$ Hz, 1H), 4.52 (d, $J = 6.8$ Hz, 1H), 4.39 (d, $J = 6.2$ Hz, 2H), 4.16 (t, $J = 6.4$ Hz, 1H), 3.17 - 3.38 (m, 3H), 3.13 - 3.15 (m, 1H), 2.13 - 2.17 (m, 2H), 1.87 (s, 1H), 1.56 - 1.68 (m, 2H), 1.00 (s, 9H), 0.24 (s, 6H). ^{13}C NMR (126 MHz, CDCl_3) δ 170.02, 161.06, 159.56, 156.06, 155.31, 151.68, 143.60, 141.30, 127.81, 127.12, 125.64, 124.97, 120.03, 117.54, 113.06, 108.05, 83.29, 69.42, 67.23, 54.52, 47.04, 38.93, 34.31, 27.65, 25.55, 18.24, 16.06, -4.41. IT-TOF: m/z $[\text{M}+\text{H}]^+$ calcd: 651.28, found: 651.28



tert-butyl ((*S*)-1-(((*S*)-3-(7-hydroxy-2-oxo-2H-chromen-4-yl)-1-oxo-1-(pent-4-yn-1-yl)amino)propan-2-yl) amino)-1-oxo-3-phenylpropan-2-yl)carbamate (**5-8**)

5-7 (100 mg, 0.15 mmol) was first treated with TBAF (1.5 mmol) in THF. After 30 min, THF was evaporated and the residue was redissolved in DMF. BocNH-Phe-COOH (137 mg, 0.18 mmol) activated by EDC (28 mg, 0.18mmol), HOBT (25 mg, 0.18 mmol), DIEA (42 μ L, 0.2 mmol) in DMF was dropwise added into the solution. DMF was evaporated after 1 h reaction. Then the residue was extracted with EA and

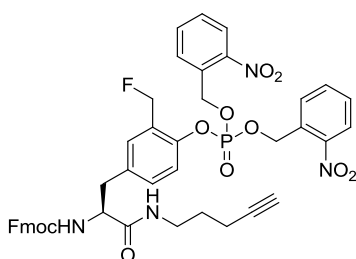
washed by 10% citric acid, saturated NaHCO₃ and brine. The product was purified by chromatography column (55%, 48 mg). ¹H NMR (500 MHz, MeOD) δ 7.75 (d, *J* = 8.8 Hz, 1H), 7.22 - 7.24 (m, 2H), 7.18 (t, *J* = 6.8 Hz, 3H), 6.86 (dd, *J* = 8.8, 2.3 Hz, 1H), 6.72 (d, *J* = 2.4 Hz, 1H), 6.10 (s, 1H), 4.64 (t, *J* = 7.1 Hz, 1H), 4.19 - 4.24 (m, 1H), 3.16 - 3.20 (m, 1H), 3.08 (dd, *J* = 14.2, 8.2 Hz, 1H), 2.98 (dd, *J* = 13.7, 5.9 Hz, 1H), 2.79 (dd, *J* = 14.71, 7.78 Hz, 1H), 2.23 (t, *J* = 2.63 Hz, 1H), 2.14 (td, *J* = 7.06, 2.61 Hz, 2H), 1.38 (s, 9H). ¹³C NMR (126 MHz, MeOD) δ 172.95, 170.57, 161.98, 161.73, 156.45, 155.53, 153.06, 136.90, 128.85, 128.07, 126.42, 126.04, 113.14, 111.51, 111.02, 102.40, 82.63, 79.55, 68.73, 56.39, 52.41, 38.25, 37.53, 33.58, 27.79, 27.26, 15.19. IT-TOF: *m/z* [M+Na]⁺ calcd: 583.25, found: 583.23



4-((*S*)-2-((*S*)-2-((tert-butoxycarbonyl)amino)-3-phenylpropanamido)-3-oxo-3-(pent-4-yn-1-ylamino) propyl)-2-oxo-2H-chromen-7-yl ethenesulfonate (**5-2**)

Prepared according to the general procedure compound **5-8** (40 mg, 0.07 mmol) and triethylamine (30 μL, 0.21 mmol) and 2-chloro-1-ethane-sulfonylchloride (8.2 μL, 0.9 mmol). The crude product was purified by chromatography column (**5-2**, 34 mg, 75%). ¹H NMR (500 MHz, CDCl₃) δ 7.92 (d, *J* = 9.3 Hz, 1H), 7.29 (dd, *J* = 14.7, 7.4 Hz, 2H), 7.10 -7.24 (m, 2H), 7.16 (d, *J* = 7.0 Hz, 2H), 6.72 (dd, *J* = 16.6, 10.0 Hz, 1H), 6.67 (s, 1H), 6.49 (s, 1H), 6.45 (d, *J* = 16.6 Hz, 1H), 6.26 (d, *J* = 9.9 Hz, 1H), 6.22 (s, 1H), 4.95 (s, 1H), 4.63 (dd, *J* = 14.2, 7.0 Hz, 1H), 4.28 (dd, *J* = 13.2, 6.5 Hz, 1H),

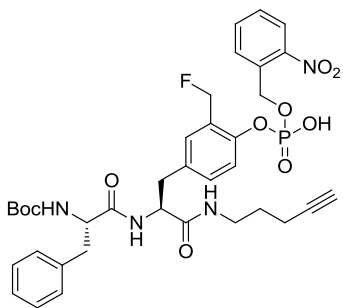
3.19 - 3.38 (m, 3H), 2.99 - 3.08(m, 3H), 2.17 (td, $J = 6.8, 2.5$ Hz, 0H), 1.89 - 1.90 (m, 1H), 1.61 - 1.66(m, 2H), 1.39 (s, 9H). ^{13}C NMR (126 MHz, CDCl_3) δ 171.64, 169.08, 159.42, 154.36, 151.38, 150.73, 135.87, 132.65, 131.90, 129.08, 128.92, 127.39, 126.27, 118.82, 117.75, 115.95, 111.25, 83.29, 81.16, 69.28, 56.35, 52.30, 39.01, 37.57, 34.25, 28.23, 28.17, 27.67, 16.04. IT-TOF: m/z $[\text{M}+\text{Na}]^+$ calcd: 674.23, found: 674.22



(*S*)-(9H-fluoren-9-yl)methyl (3-(4-((bis((2-nitrobenzyl)oxy)phosphoryl)oxy)-3-(fluoromethyl) phenyl)-1-oxo-1-(pent-4-yn-1-ylamino)propan-2-yl)carbamate (**5-10**)

Prepared according to the general coupling reaction procedure using **5-9** (500 mg, 0.63 mmol), EDC (117 mg, 0.75 mmol), HOBT (100 mg, 0.75 mmol) and DIEA (153 μL , 0.88 mmol), 4-pentyn-1-amine (140 mg, 0.70 mmol). The product was purified by chromatography column (87%, 466 mg). ^1H NMR (500 MHz, CDCl_3) δ 8.08 (d, $J = 8.2$ Hz, 2H), 7.71 (d, $J = 7.5$ Hz, 2H), 7.67 (d, $J = 7.6$ Hz, 2H), 7.62 (t, $J = 7.5$ Hz, 2H), 7.51 (d, $J = 7.3$ Hz, 2H), 7.46 (t, $J = 7.7$ Hz, 2H), 7.35 (t, $J = 7.4$ Hz, 2H), 7.32 (d, $J = 8.4$ Hz, 2H), 7.24 - 7.28 (m, 3H), 7.16 (d, $J = 6.1$ Hz, 1H), 6.36 (s, 1H), 5.66 (s, 1H), 5.60 (d, $J = 7.3$ Hz, 2H), 5.37 (d, $J = 47.4$ Hz, 2H), 4.35 - 4.44 (m, 2H), 4.25 - 4.32 (m, 1H), 4.10 - 4.14 (m, 1H), 3.22 - 3.35 (m, 2H), 3.97 - 3.11 (m, 2H), 2.05 - 2.12 (s, 2H), 1.89 (s, 1H), 1.54 - 1.63 (m, 2H). ^{13}C NMR (126 MHz, CDCl_3) δ 170.49, 155.85, 147.17, 146.63, 143.53, 141.15, 134.25, 134.07, 131.92, 131.44, 131.21, 130.80, 129.13, 128.56, 127.65, 127.42, 126.98, 125.02, 124.84, 119.89, 83.21, 79.53

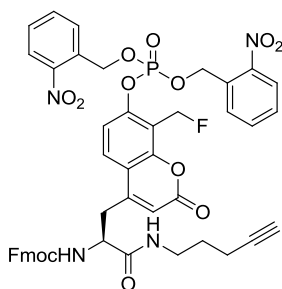
(d, $J = 167.3$ Hz, 1C, C-F), 69.19, 66.93, 66.91, 56.10, 46.98, 38.57, 37.81, 27.69, 15.82. ^{19}F NMR (282 MHz, CDCl_3) δ -136.06 (t, $J = 47.5$ Hz, 1F). ^{31}P NMR (121 MHz, CDCl_3) δ -6.40. IT-TOF: m/z $[\text{M}+\text{H}]^+$ calcd: 851.24, found: 851.30



tert-butyl ((2*S*)-1-(((2*S*)-3-(3-(fluoromethyl)-4-((hydroxy((2-nitrobenzyl)oxy)phosphoryl)oxy)phenyl)-1-oxo-1-(pent-4-yn-1-ylamino)propan-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (**5-3**)

Compound **5-10** (100 mg, 0.12 mmol) was first treated with DBU (1.2 mmol) in 5 mL DMF to deprotect Fmoc group for 10 min. BocNH-Phe-COOH (106 mg, 0.14 mmol) activated by EDC (22 mg, 0.14 mmol), HOBT (20 mg, 0.14 mmol), DIEA (31 μL , 0.15 mmol) in DMF was dropwise added into the solution. After 1 h, the residue was extracted with EA and washed by 10% citric acid, saturated NaHCO_3 and brine. Then the crude product was treated with 2 mL 1 M thiophenol and triethylamine solution in acetonitrile for 1 h. Subsequently the solvent was evaporated and the product was purified by preparative HPLC (30%, 26 mg). ^1H NMR (500 MHz, MeOD) δ 8.05 (d, $J = 7.8$ Hz, 1H), 7.82 (d, $J = 7.7$ Hz, 1H), 7.63 (t, $J = 7.5$ Hz, 1H), 7.39 - 7.49 (m, 2H), 7.11 - 7.24 (m, 7H), 5.35 - 5.45 (m, 4H), 4.47 (t, $J = 7.0$ Hz, 1H), 4.22 (dd, $J = 9.0, 5.3$ Hz, 1H), 3.08 - 3.20 (m, 2H), 2.93 - 3.02 (m, 2H), 2.83 (ddd, $J = 23.0, 13.7, 8.3$ Hz, 2H), 2.19 (t, $J = 2.6$ Hz, 1H), 2.04 - 2.09 (m, 2H), 1.52 - 1.60 (m, 2H), 1.33 (s, 9H). ^{13}C NMR (126 MHz, MeOD) δ 171.83, 170.62, 155.61, 148.54, 145.87, 136.18,

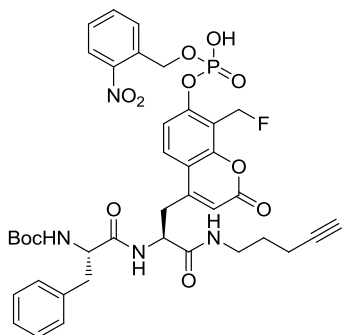
133.40, 132.78, 131.17, 129.49, 128.85, 128.10, 127.30, 127.21, 127.17, 125.52, 123.44, 118.68, 81.93, 78.90 (d, $J = 165.0$ Hz, 1C, C-F), 78.68, 78.25, 67.86, 63.69, 55.53, 53.79, 37.23, 36.77, 35.95, 26.99, 26.44, 14.33. ^{19}F NMR (282 MHz, MeOD) δ -135.95 (t, $J = 47.8$ Hz, 1F). ^{31}P NMR (121 MHz, MeOD) δ -6.67. IT-TOF: m/z $[\text{M}+\text{H}]^+$ calcd: 741.26, found: 741.26



(*S*)-(9H-fluoren-9-yl)methyl 3-(7-((bis((2-nitrobenzyl)oxy)phosphoryl)oxy)-8-(fluoromethyl)-2-oxo-2H-chromen-4-yl)-1-oxo-1-(pent-4-yn-1-ylamino)propan-2-yl carbamate (**4-2**)

Prepared according to the general coupling reaction procedure using **5-11** (300 mg, 0.35 mmol), EDC (65 mg, 0.42 mmol), HOBT (56 mg, 0.42 mmol) and DIEA (85 μL , 0.50 mmol), 4-pentyn-1-amine (90 mg, 0.45 mmol). The product was purified by chromatography column (79%, 254 mg). ^1H NMR (500 MHz, CDCl_3) δ 8.12 (d, $J = 7.4$ Hz, 2H), 7.94 (d, $J = 7.94$ Hz, 1H), 7.64 - 7.78 (m, 6H), 7.45 - 7.60 (m, 5H), 7.37 - 7.40 (m, 2H), 7.30 (t, $J = 7.4$ Hz, 2H), 6.31 (s, 1H), 6.27 (s, 1H), 5.58 - 5.71 (m, 7H), 4.44 (d, $J = 6.13$ Hz, 2H), 4.37 - 4.40 (m, 1H), 4.18 (t, $J = 6.5$ Hz, 1H), 3.30 - 3.34 (m, 2H), 3.06 - 3.26 (m, 2H), 2.14 - 2.18 (m, 2H), 1.85 (s, 1H), another 2H is inside of water peak (1.65 ppm). ^{13}C NMR (75 MHz, CDCl_3) δ 169.33, 158.95, 155.91, 153.43, 153.04, 152.94, 152.28, 151.72, 150.65, 146.82, 143.51, 141.33, 134.25, 131.26, 129.43, 128.86, 127.85, 127.14, 125.22, 124.86, 120.06, 116.28, 115.68, 115.62, 104.98, 83.21, 77.20, 72.47 (d, $J = 166.5$ Hz, C-F), 69.39, 67.41, 67.15, 54.28, 47.09,

39.04, 27.59, 16.08. ^{19}F NMR (282 MHz, CDCl_3) δ -131.76 (t, $J = 48.2$ Hz, 1F). ^{31}P NMR (121 MHz, CDCl_3) δ -7.05. IT-TOF: m/z $[\text{M}+\text{H}]^+$ calcd: 919.23, found: 919.23

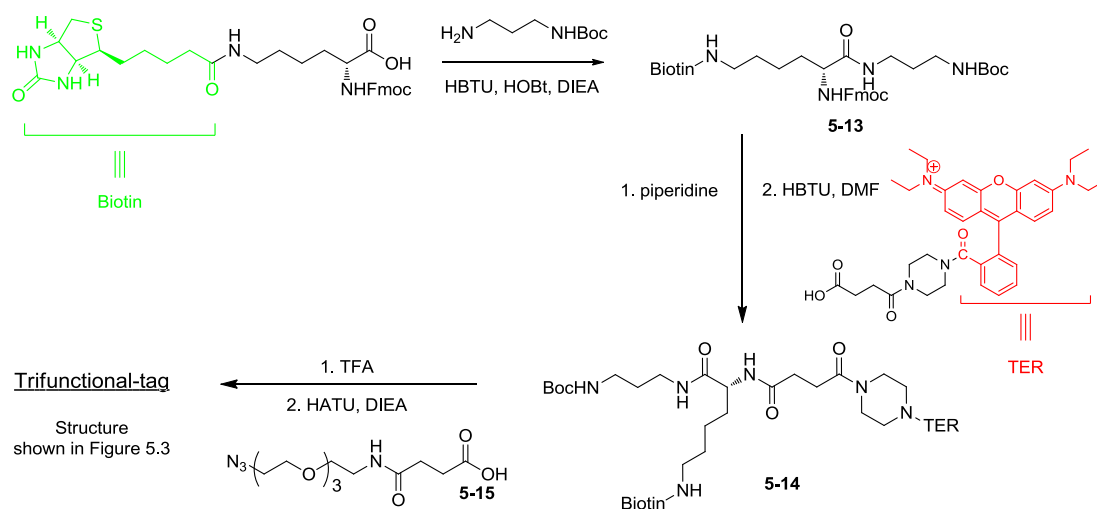


tert-butyl ((2*S*)-1-(((2*S*)-3-(8-(fluoromethyl)-7-((hydroxy((2-nitrobenzyl)oxy)phosphoryl)oxy)-2-oxo-2H-chromen-4-yl)-1-oxo-1-(pent-4-yn-1-ylamino)propan-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (**5-4**)

Compound **5-12** (80 mg, 0.09 mmol) was first treated with DBU in DMF to deprotect Fmoc group for 10 min. BocNH-Phe-COOH (83 mg, 0.11 mmol) activated by EDC (17 mg, 0.11 mmol), HOBT (15 mg, 0.11 mmol), DIEA (31 μL , 0.15 mmol) in DMF was dropwise added into the solution. After 1 h, the residue was extracted with EA and washed by 10% citric acid, saturated NaHCO_3 and brine. Then the crude product was treated with 1 mL 1 M thiophenol and triethylamine solution in acetonitrile for 1 h. Subsequently the solvent was evaporated and the product was purified by preparative HPLC (25%, 18 mg). ^1H NMR (500 MHz, d_6 -DMSO) δ 8.14 (d, $J = 6.1$ Hz, 1H), 8.06 (d, $J = 8.1$ Hz, 1H), 7.90 (d, $J = 9.1$ Hz, 1H), 7.70 - 7.81 (m, 3H), 7.64 (d, $J = 8.8$ Hz, 1H), 7.52 (t, $J = 7.5$ Hz, 1H), 7.15 - 7.23 (m, 5H), 6.95 (d, $J = 8.3$ Hz, 1H), 6.19 (s, 1H), 5.50 (d, $J = 48.2$ Hz, 2H), 5.20 (d, $J = 7.7$ Hz, 2H), 4.58 (dd, $J = 13.88, 7.95$ Hz, 1H), 4.08 - 4.12 (m, 1H), 3.16 - 3.19 (m, 1H), 3.09 - 3.13 (m, 2H), 3.00 (dd, $J = 13.9, 8.2$ Hz, 1H), 2.87 (dd, $J = 13.8, 4.3$ Hz, 1H), 2.75 (t, $J = 2.47$ Hz, 1H), 2.65 - 2.72 (m, 1H), 2.12 (dd, $J = 7.0, 4.7$ Hz, 2H), 1.54 (td, $J = 13.3, 6.6$ Hz,

2H), 1.27 (s, 9H). ^{13}C NMR (126 MHz, d_6 -DMSO) δ 171.56, 169.72, 159.32, 156.36, 155.23, 153.10, 152.12, 146.49, 137.93, 135.18, 133.90, 130.24, 129.05, 128.41, 128.10, 128.05, 127.97, 126.75, 126.14, 124.33, 120.24, 116.39, 113.22, 113.07, 112.66, 83.85, 78.23, 73.46, 72.81(d, $J = 166.5$ Hz, C-F), 63.55, 56.02, 51.41, 37.78, 37.35, 34.35, 28.03, 27.81, 15.27. ^{19}F NMR (282 MHz, d_6 -DMSO) δ -135.28 (t, $J = 47.4$ Hz, 1F). ^{31}P NMR (121 MHz, d_6 -DMSO) δ -5.45. IT-TOF: m/z $[\text{M-H}]^+$ calcd: 807.25, found: 807.22

6.10 Synthesis of trifunctional tag



Scheme 6.1 Synthetic scheme of trifunctional tag

Synthesis of 5-13

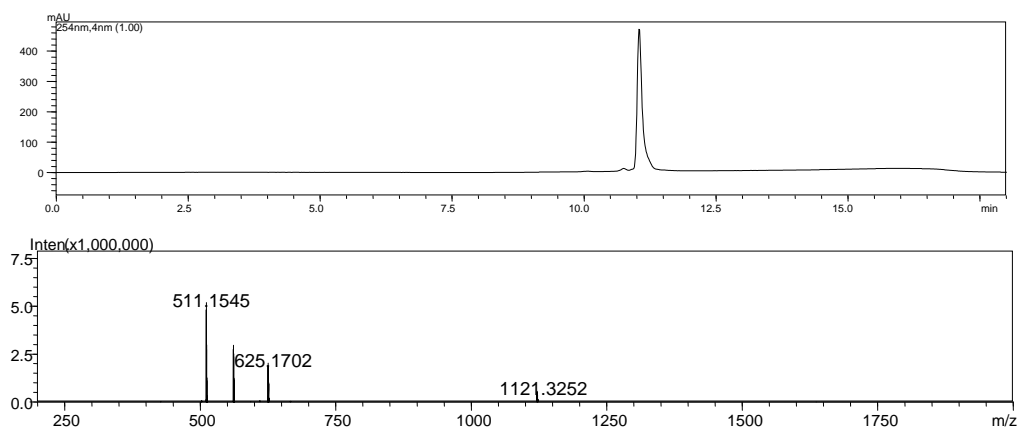
The solution of Fmoc-Lys(Biotin)-OH (4.0 g, 6.7 mmol)¹²², HBTU (3.04 g, 8.0 mmol) and HOBT (1.08 g, 8.0 mmol) was stirred at room temperature for 10 min. Subsequently, *N*-Boc-propyldiamine (1.76 g, 10 mmol) and DIEA (2.36 mL, 13.6 mmol) were added. After the reaction was stirred over night at room temperature, the product was precipitated in water as a yellowish solid. (3.5 g, 69%). ^1H NMR (300 MHz, d_6 -DMSO) δ 7.95-7.85 (m, 3H), 7.74 (s, 3H), 7.44-7.39 (m, 3H), 7.32 (t, $J = 7.3$

Hz, 2H), 6.76 (s, 1H), 6.42 (s, 1H), 6.35 (s, 1H), 4.49-4.22 (m, 4H), 4.11-4.09 (m, 1H), 3.92-3.85 (m, 1H), 3.15-2.75 (m, 9H), 2.04 (t, $J = 6.8$ Hz, 2H), 1.68-1.10 (m, 23H).
 ^{13}C NMR (75 MHz, d_6 -DMSO) δ 171.8, 162.7, 155.9, 155.6, 143.9, 143.8, 140.7, 127.6, 127.0, 125.3, 120.1, 108.4, 77.4, 65.6, 61.0, 59.2, 55.4, 54.7, 46.7, 38.3, 37.4, 36.2, 35.8, 35.2, 31.6, 30.8, 29.5, 28.9, 28.2, 28.0, 25.3, 23.0.

Synthesis of **5-14**

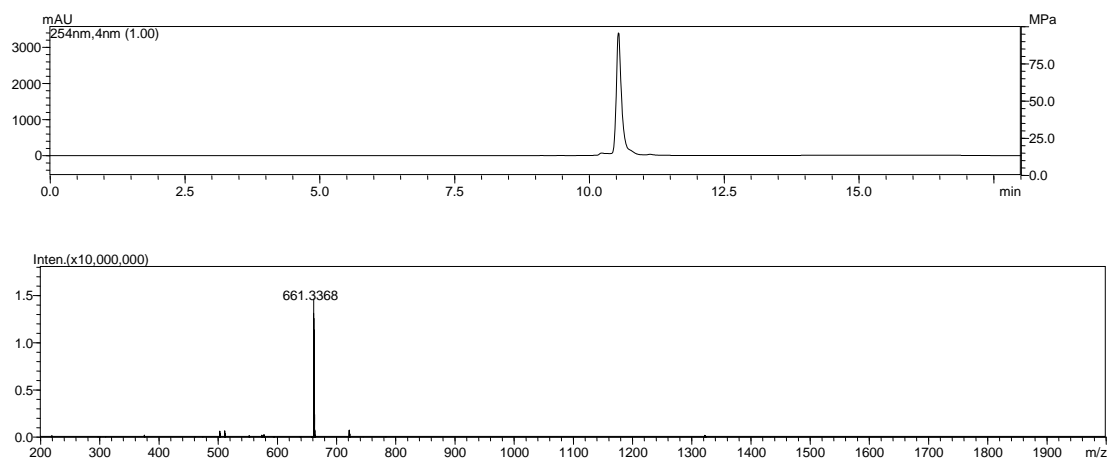
The solution of compound **5-13** (3.5 g, 4.7 mmol) in piperidine-DMF (1: 4, 15 mL) was stirred at room temperature for 1 h. After the volatiles were removed under reduced pressure, water (~ 40 mL) was then added into the solution until no more precipitate formed. The precipitate was filtered and the filtrate was concentrated under reduced pressure to give the product as yellowish oil which was slowly changed to a yellowish solid. (2.2 g, 89%). The product was pure enough to use in next step.

The crude product (1.8 g, 3.4 mmol) and DIEA (0.74 mL) in DMF were added into a solution of TER-acid (1.6 g, 2.2 mmol), HBTU (1.0 g, 2.6 mmol) and HOBT (0.36 g, 2.8 mmol) in DMF (18 mL). After the reaction was stirred at room temperature overnight, the compound **5-14** was purified by prepHPLC. The purity of **5-14** was shown in the following LCMS. IT-TOF: m/z $[\text{M}]^+$ calcd: 1121.6, found: 1121.3



Synthesis of Trifunctional tag

To the solution compound **5-14** (300 mg, 0.27 mmol) in DCM (10 mL) was added TFA (0.5 mL) and the reaction was stirred at room temperature for 1 h. The volatiles were removed under reduced pressure. The residual TFA was removed by azeotropy with toluene. After the solution of the residue above, DIEA (400 μ L), HATU (137 mg, 0.36 mmol) and **5-15** (0.11 g, 0.35 mmol) in DMF (10 mL) was stirred at room temperature overnight, the tag was purified by prepHPLC. IT-TOF: m/z $[M+H]^+$ calcd: 661.3, found: 661.3



6.11 Mass results

Table 6.2 The complete list of **P23** Labeled proteins¹²³ identified from MCF-7 cell lysate.

Name (Gene, Protein)	M.W.	Score	Queries matched	emPAI	Family
FKBP4 FK506-binding protein 4	52057	738	29	3.14	immunophilin protein family
VAT1 Synaptic vesicle membrane protein VAT-1 homolog	42122	387	7	0.31	zinc-containing alcohol dehydrogenase family
WARS Tryptophanyl-tRNA synthetase, cytoplasmic	53474	354	10	0.45	class-I aminoacyl-tRNA synthetase family
CTSD Cathepsin D precursor	45037	338	13	0.46	peptidase family
PSMD6 26S proteasome non-ATPase regulatory subunit 6	45787	313	16	1.1	proteasome subunit S10 family
VDAC1 Voltage-dependent anion-selective channel protein 1	30868	289	9	0.85	eukaryotic mitochondrial porin family
DCTN2 dynactin 2	44906	275	8	0.46	dynactin subunit 2 family
SEC22B Vesicle-trafficking protein SEC22b	24896	271	7	1.13	synaptobrevin family
PRKCSH Glucosidase 2 subunit beta precursor	60357	251	6	0.15	
SSB Lupus La protein	46979	245	7	0.35	
PCK2 Phosphoenolpyruvate carboxykinase [GTP], mitochondrial precursor	71447	242	7	0.25	phosphoenolpyruvate carboxykinase [GTP] family.
HM13 Isoform 1 of Minor histocompatibility antigen H13	41747	229	10	0.58	peptidase family
SLC7A5 Large neutral amino acids transporter small subunit 1	55659	223	6	0.11	amino acid-polyamine-organocation (APC) superfamily
BCAP31 B-cell receptor-associated protein 31	34901	217	7	0.89	BCAP29/BCAP31 family
BZW1 similar to basic leucine zipper and W2 domains 1	51420	208	15	0.55	BZW family
YBX1 Nuclease sensitive element-binding protein 1	35903	208	5	0.17	may involve in cancer
TRIM28 Isoform 1 of Transcription intermediary factor 1-beta	90261	177	8	0.21	TRIM/RBCC family
CTTN Src substrate cortactin	61896	175	7	0.32	
CAP1 Adenylyl cyclase-associated protein 1	52222	173	5	0.18	CAP family

Chloride intracellular channel protein 1	27248	169	4	0.41	chloride channel CLIC family.
PSME1 Proteasome activator complex subunit 1	28876	150	11	1.4	PA28 family
IDH2 Isocitrate dehydrogenase [NADP], mitochondrial precursor	51333	149	7	0.32	isocitrate and isopropylmalate dehydrogenases family
RAB1A Isoform 1 of Ras-related protein Rab-1A	22891	149	6	0.51	Rab family
SHMT2 Serine hydroxymethyltransferase, mitochondrial precursor	56414	147	5	0.22	SHMT family
HEXA Beta-hexosaminidase alpha chain precursor	61106	145	9	0.32	glycosyl hydrolase 20 family
AARS Alanyl-tRNA synthetase, cytoplasmic	107484	145	8	0.17	class-II aminoacyl-tRNA synthetase family
SELENBP1 Selenium-binding protein 1	52928	144	8	0.38	selenium-binding protein family
CS Citrate synthase, mitochondrial precursor	51908	141	4	0.18	citrate synthase family
CD63 CD63 antigen	26474	140	10	0.81	tetraspanin (TM4SF) family
LCPI Plastin-2	70815	135	5	0.22	actin-binding proteins
TMED10 Transmembrane emp24 domain-containing protein 10 precursor	25131	133	6	0.65	EMP24/GP25L family.
Ribonuclease inhibitor	51766	131	3	0.12	
CDC37 Hsp90 co-chaperone Cdc37	44953	131	7	0.46	CDC37 family
RAB10 Ras-related protein Rab-10	22755	129	5	0.51	Rab family
Annexin A11	54697	127	7	0.3	Annexin family
LAMP1 lysosomal-associated membrane protein 1	45367	126	6	0.32	LAMP family
SLC9A3R1 Ezrin-radixin-moesin-binding phosphoprotein 50	39130	124	6	0.43	solute carrier family
TXNDC4 Thioredoxin domain-containing protein 4 precursor	47341	121	3	0.2	
OLA1 Isoform 1 of Putative GTP-binding protein 9	44943	121	6	0.37	GTP1/OBG family
PSMD11 Proteasome 26S non-ATPase subunit 11 variant (Fragment)	47790	120	3	0.19	proteasome family.
PHB Prohibitin	29843	118	5	0.7	prohibitin family.
SUCLG2 Succinyl-CoA ligase [GDP-forming] beta-chain, mitochondrial precursor	46824	116	4	0.27	succinate/malate CoA ligase beta subunit family
USP5 Isoform Long of Ubiquitin carboxyl-terminal hydrolase 5	96638	116	5	0.09	peptidase family

CAND1 Isoform 1 of Cullin-associated NEDD8-dissociated protein 1	137999	115	7	0.12	CAND family
BASP1 Brain acid soluble protein 1	22680	111	5	0.74	BASP1 family
VISA Isoform 1 of Mitochondrial antiviral-signaling protein	57063	110	4	0.16	
USP14 Ubiquitin carboxyl-terminal hydrolase 14	56489	109	4	0.22	peptidase family
DERP12	38340	108	6	0.39	
VDAC3 Isoform 1 of Voltage-dependent anion-selective channel protein 3	30981	108	3	0.23	eukaryotic mitochondrial porin family
VAPB Isoform 1 of Vesicle-associated membrane protein-associated protein B/C	27439	107	6	0.77	VAMP-associated protein family
AIFM1 Isoform 1 of Apoptosis-inducing factor 1, mitochondrial precursor	67144	107	5	0.15	FAD-dependent oxidoreductase family
PICALM Isoform 2 of Phosphatidylinositol-binding clathrin assembly protein	68892	106	2	0.09	
CNDP2 Cytosolic non-specific dipeptidase	41770	103	5	0.24	peptidase family
BRI3BP BRI3-binding protein	27932	102	3	0.12	
PTPN1 Tyrosine-protein phosphatase non-receptor type 1	50505	101	5	0.25	protein-tyrosine phosphatase family
ALDH7A1 Similar to Antiquitin	59020	99	2	0.1	aldehyde dehydrogenase family
SRP54 Signal recognition particle 54 kDa protein	55953	99	4	0.23	GTP-binding SRP family
YWHAB Isoform Long of 14-3-3 protein beta/alpha	28179	98	4	0.4	14-3-3 family
YWHAE 14-3-3 protein epsilon	29326	98	4	0.38	14-3-3 family
TACSTD1 Tumor-associated calcium signal transducer 1 precursor	35582	97	3	0.17	EPCAM family
MARCKS Myristoylated alanine-rich C-kinase substrate	31707	96	3	0.22	MARCKS family.
ANXA7 Isoform 1 of Annexin A7	52991	95	6	0.24	annexin family
ARFIP1 Isoform A of Arfaptin-1	38632	92	3	0.24	
RANBP5 127 kDa protein	127923	91	5	0.08	importin beta family
SERPINH1 Serpin H1 precursor	46525	91	3	0.2	serpin family
PSMB5 Putative uncharacterized protein DKFZp686I0180 (Fragment)	28962	90	2	0.11	peptidase family
RAD23B UV excision repair protein RAD23 homolog B	43202	90	8	0.39	RAD23 family

CLCC1 Isoform 1 of Chloride channel CLIC-like protein 1 precursor	62667	87	4	0.11	chloride channel MCLC family
GNAS Isoform XLas-1 of Guanine nucleotide-binding protein G(s) subunit alpha isoforms XLas	111697	85	5	0.08	G-alpha family
RAB3A Ras-related protein Rab-3A	25196	85	3	0.28	Rab family
PSMC6 26S protease regulatory subunit S10B	44430	85	3	0.21	AAA ATPase family
HMGCS1 Hydroxymethylglutaryl-CoA synthase, cytoplasmic	57828	85	5	0.16	HMG-CoA synthase family
SH3GLB2 Isoform 1 of SH3 domain GRB2-like protein B2	44175	84	4	0.21	endophilin family
RUVBL1 Isoform 1 of RuvB-like 1	50538	84	5	0.25	RuvB family
RAB11B Ras-related protein Rab-11B	24588	84	3	0.46	Rab family
TOMM40 Isoform 1 of Probable mitochondrial import receptor subunit TOM40 homolog	38211	83	4	0.34	Tom40 family
ATP6V1A Vacuolar ATP synthase catalytic subunit A	68660	82	3	0.09	ATPase alpha/beta chains family
OAT Ornithine aminotransferase, mitochondrial precursor	48846	81	3	0.19	class-III pyridoxal-phosphate-dependent aminotransferase family
FUBP1 Isoform 1 of Far upstream element-binding protein 1	67690	80	5	0.18	
EIF4G1 EIF4G1 variant protein (Fragment)	178843	77	4	0.06	
TMPO Isoform Beta of Lamina-associated polypeptide 2, isoforms beta/gamma	50696	77	3	0.18	LEM family
CCT6A T-complex protein 1 subunit zeta	58444	77	2	0.12	TCP-1 chaperonin family
ASAH1 Acid ceramidase precursor	45077	77	2	0.15	acid ceramidase family
ATP1B3 Sodium/potassium-transporting ATPase subunit beta-3	31834	77	2	0.22	potassium ATPases subunit beta family
FBP1 Fructose-1,6-bisphosphatase 1	37190	76	5	0.29	FBPase class 1 family
PSMB3 Proteasome subunit beta type-3	23219	76	2	0.31	peptidase family
SERPINB1 Leukocyte elastase inhibitor	42829	76	2	0.14	serpin family
PCYOX1 Prenylcysteine oxidase 1 precursor	57003	75	4	0.22	prenylcysteine oxidase family
CLPTM1 Isoform 2 of Cleft lip and palate transmembrane protein 1	79791	74	2	0.07	CLPTM1 family
NCSTN Isoform 1 of Nicastrin precursor	79103	74	4	0.18	nicastrin family

GNS N-acetylglucosamine-6-sulfatase precursor	62840	74	3	0.15	sulfatase family
ATP6V1H Isoform 1 of Vacuolar ATP syntha	56417	73	2	0.05	V-ATPase H subunit family
ATP6AP1 Vacuolar ATP synthase subunit S1 precursor	52164	72	4	0.24	vacuolar ATPase subunit S1 family
SNX2 Sorting nexin-2	58549	72	3	0.18	sorting nexin family
ERP29 Endoplasmic reticulum protein ERp29 precursor	29032	72	4	0.54	
S100A11 Protein S100-A11	11847	72	3	0.66	S-100 family
BAT3 Isoform 1 of Large proline-ric	120639	71	3	0.08	
F3B Isoform 1 of Eukaryotic translation initi	92833	71	2	0.04	eIF-3 subunit B family.
TXNDC5;MUTED Thioredoxin domain-containing protein 5 precursor	48283	71	5	0.26	protein disulfide isomerase family
CKAP4 Isoform 1 of Cytoskeleton-associated protein 4	66097	70	6	0.03	
RPL5 60S ribosomal protein L5	34569	70	2	0.2	ribosomal protein L18P family
EIF2S2 Eukaryotic translation initiation factor 2 subunit 2	38706	69	2	0.16	eIF-2-beta/eIF-5 family.
RTN4 Isoform 1 of Reticulon-4	130420	68	5	0.09	
PPT1 Palmitoyl-protein thioesterase 1 precursor	76216	68	4	0.13	palmitoyl-protein thioesterase family
KYNU Kynureninase	52831	68	2	0.13	kynureninase family.
UBQLN4 Ubiquilin-4	63869	67	2	0.05	
SCARB2 Lysosome membrane protein 2	54712	62	3	0.11	CD36 family
FLJ22662 hypothetical protein LOC79887	63499	62	2	0.09	phospholipase B-like family
LAP3 Isoform 1 of Cytosol aminopeptidase	56530	61	2	0.11	peptidase family
IMMT Isoform 1 of Mitochondrial inner membrane protein	84026	61	4	0.11	
UQCRC2 Ubiquinol-cytochrome-c reductase complex core protein 2, mitochondrial precursor	48584	59	2	0.12	peptidase family
LAMP2 Isoform LAMP-2A of Lysosome-associated membrane glycoprotein 2 precursor	45503	59	2	0.13	LAMP family
RUVBL2 RuvB-like 2	51296	58	2	0.12	peptidase family
PRKACA Isoform 2 of cAMP-dependent protein kinase, alpha-catalytic subunit	39911	55	2	0.15	protein kinase superfamily

Chapter 7 Conclusion remarks

In summary, we developed some new unnatural amino acids which mimic *p*Tyr in this study. Protein phosphorylation, targeted to tyrosine residues, plays a critical role in regulating many fundamental cellular processes including growth differentiation, metabolism, cell cycle progression, cell adhesion and migration, ion channel activity and immune response. Current knowledge of *p*Tyr related proteins and domains have been mostly derived from experiments using gene knockout or overexpression techniques. However, these approaches do not provide an adequate view of the temporal, spatial and dynamic properties of proteins and domains in cell signaling. Given the large number and complexity of cellular system, we focused on novel *p*Tyr mimic with better bioavailability and new mechanism-based probes.

As described in Chapter 2, we synthesized a novel unnatural amino acid which mimics phosphotyrosine. The *p*Tyr was replaced with a non-hydrolyzable isoxazole carboxylic acid. The new mimic possesses better cell permeability and hydrolytic stability. This unnatural amino acid can be readily incorporated into SH2 domain inhibitors, such as **ISS 610** (a known STAT3 PPI inhibitor), to derive a new inhibitor which has better pharmacological properties without losing most of its inhibitory activity against STAT3. Its incorporation into a known STAT3 SH2 domain inhibitor further confirmed our design principle. This opens up opportunities to study the incorporation of this amino acid into other *p*Tyr-containing higher biological active small molecules and proteins to further understand the studies of PTP biology. Furthermore, this mimic could potentially be incorporated into proteins using the nonsense suppression technology.

In Chapter 3, since most of PTP fluorogenic substrates diffuse from the site of dephosphorylation making it not suitable for cell based assays, we synthesized the first self-immobilizing, fluorogenic unnatural amino acid that mimics *p*Tyr. Using solid-phase peptide synthesis, it was subsequently incorporated into peptide-based localization probes which were applied in bioimaging and fluorescence-activated cell sorting. However, due to the highly *pH* sensitivity and low quantum yield of coumarin, the mimic could be modified to improve the properties. Nevertheless, as this study is based on an unnatural amino acid, the incorporation of this mimic to peptides, proteins and other biological active compounds could be explored to improve the recognition and specificity in detecting PTP activity in cellular systems.

Activity-based protein profiling (ABPP) has emerged as a powerful chemical proteomic approach for fingerprinting and identifying targets/off-targets of known/unknown drugs. ABPPs have already been widely used in PTP family, however, due to the high homology of PTP active site, the development of specific probes to target specific PTP still remains to be addressed. In Chapter 4, inspired from the Ugi reaction and the bidentate ligand concept developed by Zhang group, we rapidly assembled a small library of PTP probes to explore the fingerprinting of different PTPs and profiling its cellular targets. In future, ABPs could be assembled in other high-through put manner, like the microwave solid-phase synthesis to generate a large library with different scaffolds. Furthermore, other multi-component reactions (MCR) can also be used in ABPs development. The combination of MCR and ABPP will become another useful chemical tool in “Catalomic” field.

Moreover, in Chapter 5, we continued to develop new phosphotyrosine mimics and converted a known PTP1B inhibitor to four activity based probes. The four “clickable” and cell-permeable probes containing either sulfonate or quinone

methide-base reactive group were successfully synthesized. Subsequently, proteome-wide profiling of potential cellular targets in the live MCF-7 cancer cells was performed. However, based on the pull-down experiments followed by mass spectrometry validation, PDI family proteins were found to be the major and specific targets of vinyl sulfonate probes not PTP family. As this VS probe is failed to be a suitable probe for PTP family, new probes should be done to further investigate PTPs function in cellular process. Furthermore, this VS probe gave a bright prospective view in developing new inhibitors of PDI for better understanding its role in biological systems.

It is hoped that the *p*Tyr mimics and MCR assembling probe approach developed in this study, could provide valuable chemical biological tools for investigation of the phosphoproteome network.

Chapter 8 References

- 1 (a) J. Drews, *Science*, 2000, **287**, 1960-1964; (b) J. G. Robertson, *Curr. Opin. Struc. Biol.*, 2007, **17**, 674-679; (c) M. Rask-Andersen, M. S. Alm n, H. B. Schi th, *Nature Rev. Drug Disc.*, 2011, **10**, 579-590.
- 2 (a) T. Hunter, *Cell*, 1995, **80**, 225-236; (b) T. Hunter, *Cell*, 2000, **100**, 113-127; (c) L. N. Johnson, R. J. Lewis, *Chem. Rev.*, 2001, **101**, 2209-2242; (d) P. Cohen, *Nature Cell Biol.*, 2002, **4**, 127-130; (e) M. K. Tarrant, P. A. Cole, *Annu. Rev. Biochem.*, 2009, **78**, 797-825; (f) C. H. S. Lu, K. Liu, L. P. Tan, S. Q. Yao, *Chem. Eur. J.*, 2012, **18**, 28-29.
- 3 (a) G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, *Science*, 2002, **298**, 1912-1934; (b) P. Cohen, *Nature Rev. Drug Disc.*, 2002, **1**, 309-315.
- 4 (a) D. Barford, *Trends Biochem.Sci.*, 1996, **21**, 407-412; (b) Y. G. Shi, *Cell*, 2009, **139**, 468-484; (c) J. L. McConnel, B. E. Wadzinski, *Mol. Pharmacol.*, 2009, **75**, 1249-1261; (d) A. Alonso, J. Sasin, N. Bottini, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon, T. Mustelin, *Cell*, 2004, **117**, 699-711; (e) S. G. Julien, N. Dub  S. Hardy, M. L. Tremblay, *Nature Rev. Cancer*, 2011, **11**, 35-49; (f) N. K. Tonks, *Nature Rev. Mole. Cell Biol.*, 2006, **7**, 833-846.
- 5 (a) J. Kuriyan, D. Cowburn, *Annu. Rev. Biophys. Biomol. Struct.*, 1997, **26**, 259-288; (b) S. E. Shoelson, *Curr. Opin. Chem. Biol.*, 1997, **1**, 227-234; (c) M. B. Yaffe, A. E. Elia, *Curr. Opin. Cell Biol.*, 2001, **13**, 131-138; (d) M. B. Yaffe, *Nat. Rev. Mol. Cell Biol.*, 2002, **3**, 177-186.

- 6 (a) K. Machida, B. J. Mayer, *Biochim. Biophys. Acta.*, 2005, **1747**, 1-25; (b) T. Pawson, G. D. Gish, P. Nash, *Trends Cell Biol.*, 2001, **11**, 504-511.
- 7 (a) P. Blaikie, D. Immanuel, J. Wu, N. Li, V. Yajnik, B. Margolis, *J. Biol. Chem.*, 1994, **269**, 32031-32034; (b) M. M. Zhou, S. W. Fesik, *Prog. Biophys. Mol. Biol.*, 1995, **64**, 221-235; (c) J. Kuriyan, D. Cowburn, *Annu. Rev. Biomol. Struct.*, 1997, **26**, 259-288.
- 8 (a) K. Rittinger, J. Budman, J. Xu, S. Volinia, L. C. Cantley, S. J. Smerdon, S. J. Gamblin, M. B. Yaffe, *Mol. Cell*, 1999, **4**, 153-166; (b) H. Hermeking, *Nat. Rev. Cancer*, 2003, **3**, 931-943; (c) P. Mhawech, *Cell Res.*, 2005, **15**, 228-236.
- 9 (a) A. Östman, C. Hellberg, F. D. Böhmer, *Nature Rev. Cancer*, 2006, **6**, 307-320; (b) L. Bialy, H. Waldmann, *Angew. Chem. Int. Ed.*, 2005, **44**, 3814-3839.
- 10 (a) J. N. Andersen, O. H. Mortensen, G. H. Peters, P. G. Drake, L. F. Iversen, O. H. Olsen, P. G. Jansen, H. S. Andersen, N. K. Tonks, N. P. H. Møller, *Mol. Cell Biol.*, 2001, **21**, 7117-7136; (b) Z. -Y. Zhang, *Acc. Chem. Res.*, 2003, **36**, 385-392.
- 11 (a) A. P. Ducruet, A. Vogt, P. Wipf, J. S. Lazo, *Annu. Rev. Pharmacol. Toxicol.*, 2005, **45**, 725-750; (b) K. I. Patterson, T. Brummer, P. M. O'Brien, R. J. Daly, *Biochem. J.*, 2009, **418**, 475-489.
- 12 (a) K. Kristjansdottir, J. Rudolph, *Chem. Biol.*, 2004, **11**, 1043-1051; (b) R. Boutros, V. Lobjois, B. Ducommun, *Nature Rev. Cancer*, 2007, **7**, 495-507.
- 13 G. Raugei, G. Ramponi, P. Chiarugi, *Cell Mol. Life Sci.*, 2002, **59**, 941-949.
- 14 T. R. Burke, Jr., Z. -Y. Zhang, *Biopolymers (Pept. Sci.)*, 1998, **47**, 225-241.
- 15 (a) Z. Jia, D. Barford, A. J. Flint, N. K. Tonks, *Science*, 1995, **268**, 1754-1758; (b) A. D. B. Pannifer, A. J. Flint, N. K. Tonks, D. Barford, *J. Biol. Chem.*, 1998, **273**, 10454-10462.

- 16 (a) K. L. Guan, J. E. Dixon, *J. Biol. Chem.*, 1991, **266**, 17026-17030; (b) M. Streuli, N. X. Krueger, T. Thai, M. Tang, H. Saito, *EMBO J.*, 1990, **9**, 2399-2407.
- 17 (a) Z.-Y. Zhang, Y. Wang, E. J. Dixon, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 1624 -1627; (b) J. M. Denu, D. L. Lohse, J. Vijayalakshmi, M. A. Saper, J. E. Dixon, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 2493-2498.
- 18 (a) I. Sadowski, J. C. Stone, T. Pawson, *Mol. Cell Biol.*, 1986, **6**, 4396-4408; (b) A. Koch, M. Moran, I. Sadowski, T. Pawson, *Mol. Cell Biol.*, 1989, **9**, 4131-4140.
- 19 (a) G. Waksman, S. E. Shoelson, N. Pant, D. Cowburn, J. Kuriyan, *Cell*, 1993, **72**, 779-790; (b) T. K. Sawyer, *Biopolymers (Pept. Sci.)*, 1998, **47**, 243-261; (c) W. C. Shakespeare, *Curr. Opin. Chem. Biol.*, 2001, **5**, 409-415; (d) T. Pawson, *Cell*, 2004, **116**, 191-203; (d) K. Machida, B. J. Mayer, *BBA-Proteins Proteomics*, 2005, **1747**, 1-25
- 20 (a) G. Waksman, D. Kominos, S. C. Robertson, N. Pant, D. Baltimore, R. B. Birge, D. Cowburn, H. Hanafusa, B. J. Mayer, M. Overduin, M. D. Resh, C. B. Rios, L. Silverman, J. Kuriyan, *Nature*, 1992, **358**, 646-653; (b) M. J. Eck, S. E. Shoelson, S. C. Harrison, *Nature*, 1993, **362**, 87-91; (c) G. Waksman, S. E. Shoelson, N. Pant, D. Cowburn, J. Kuriyan, *Cell*, 1993, **72**, 779-790; (d) S. Zhou, S. E. Shoelson, M. Chaudhuri, G. Gish, T. Pawson, W. G. Haser, F. King, T. Roberts, S. Ratnofsky, R. J. Lechleider, B. G. Neel, M. M. Chou, H. Hanafusa, B. Schaffhausen, L. C. Cantley, *Cell*, 1993, **72**, 767-778.
- 21 (a) F. Poy, M. B. Yaffe, J. Sayos, K. Saxena, M. Morra, J. Sumegi, L. C. Cantley, C. Terhorst, M. J. Eck, *Mol. Cell*, 1999, **4**, 555-561; (b) J. L. Cannons, S. G. Tangye, P. L. Schwartzberg, *Annu. Rev. Immu.*, 2011, **9**, 665-705.

- 22 (a) M. L. Galisteo, I. Dikic, A. G. Batzer, W. Y. Langdon, J. Schlessinger, *J. Biol.Chem.*, 1995, **270**, 20242-20245; (b) G. Swaminathan, A. Y. Tsygankov, *J. Cell. Physiol.*, 2006, **209**, 21-43.
- 23 (a) P. Hof, S. Pluskey, S. Dhe-Paganon, M. J. Eck, S. E. Shoelson, *Cell*, 1998, **92**, 441-450; (b) M. Tartaglia, E. L. Mehler, R. Goldberg, G. Zampino, H. G. Brunner, H. Kremer, I. van der Burgt, A. H. Crosby, A. Ion, S. Jeffery, K. Kalidas, M. A. Patton, R. S. Kucherlapati, B. D. Gelb, *Nature Genet.*, 2001, **29**, 465-468; (c) U. Lorenz, *Immunol. Rev.*, 2009, **228**, 342-359.
- 24 (a) J. Nachman, G. Gish, C. Virag, T. Pawson, R. Pomes, E. Pai, *PLoS One*, 2010, **5**, e11215; (b) D. H. Anderson, P. M. Ismail, *Oncogene*, 1998, **16**, 2321-2331.
- 25 (a) G. S. Feng, *Exp. Cell Res.*, 1999, **253**, 47-54; (b) B. G. Neel, H. Gu, L. Pao, *Trends Biochem. Sci.*, 2003, **28**, 284-293; (c) Y. M. Agazie, N. Movilla, I. Ischenko, M. J. Hayman, *Oncogene*, 2003, **22**, 6909-6918; (d) D. Barua, J. R. Faeder, J. M. Haugh, *Biophys. J.* 2007, **92**, 2290-2300.
- 26 (a) T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci., U.S.A.*, 1998, **95**, 6705-6710; (b) T. W. Muir, *Annu. Rev. Biochem.*, 2003, **72**, 249-289.
- 27 (a) L. Wang, P. G. Schultz, *Angew. Chem. Int. Ed.*, 2005, **44**, 34-66; (b) L. Wang, J. Xie, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.*, 2006, **35**, 225-249; (c) C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.*, 2010, **79**, 413-444.
- 28 (a) C. M. Salisbury, D. J. Maly, J. A. Ellman, *J. Am. Chem. Soc.*, 2002, **124**, 14868-14870; (b) Q. Zhu, M. Uttamchandani, D. Li, M. L. Lesaicherre, S. Q. Yao, *Org. Lett.*, 2003, **5**, 1257-1260.
- 29 J. Montalibet, K.I. Skorey, B.P. Kennedy, *Methods*, 2005, **35**, 2-8.

- 30 S. Welte, K.-H. Baringhaus, W. Schmider, G. Muller, S. Petry, N. Tennagels, *Anal. Biochem.*, 2005, **338**, 32-38.
- 31 (a) S. Mitra, A. M. Barrios, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 5142-5145; (b) S. Mitra, A. M. Barrios, *ChemBioChem*, 2008, **9**, 1216-1219.
- 32 S. Mitra, A. Barrios, *Anal. Biochem.*, 2007, **370**, 249-251
- 33 (a) K. Curley, D. S. Lawrence, *Curr. Opin. Chem. Biol.*, 1999, **3**, 84-88; (b) G. Mayer, A. Heckel, *Angew. Chem., Int. Ed.*, 2006, **45**, 4900-4921; (c) H. -M. Lee, D. R. Larson, D. S. Lawrence, *ACS Chem. Biol.*, 2009, **4**, 409-427.
- 34 K. Zou, W. T. Miller, R. S. Givens, H. Bayley, *Angew. Chem. Int. Ed.*, 2001, **40**, 3049-3051.
- 35 (a) D. M. Rothman, M. E. Vázquez, E. M. Vogel, B. Imperiali, *Org. Lett.*, 2002, **4**, 2865-2868; (b) D. M. Rothman, E. J. Petersson, M. E. Vázquez, G. S. Brandt, D. A. Dougherty, B. Imperiali, *J. Am. Chem. Soc.*, 2005, **127**, 846-847; (c) S. Lahiri, R. Seidel, M. Engelhard, C. F. Becker, *Mol. Biosyst.*, 2010, **6**, 2423-2429.
- 36 (a) E. M. Vogel, B. Imperiali, *Protein Sci.*, 2007, **16**, 550-556; (b) S. Lahiri, R. Seidel, M. Engelhard, C. F. Becker, *Mol. Biosyst.*, 2010, **6**, 2423-2429.
- 37 T. Kawakami, H. Cheng, S. Hashiro, Y. Nomura, S. Tsukiji, T. Furuta, T. Nagamune, *ChemBioChem*, 2008, **9**, 1583-1586.
- 38 (a) D. L. Pettit, S. S. -H. Wang, K. R. Gee, G. J. Augustine, *Neuron*, 1997, **19**, 465-471; (b) F. Helmchen, W. Denk, *Nat. Methods*, 2005, **2**, 932-940.
- 39 (a) I. Marseigne, B. P. Roques, *J. Org. Chem.*, 1988, **53**, 3621-3624; (b) T. R. Burke, Jr., P. Russ, B. Lim, *Synthesis*, 1991, **11**, 1019-1020.
- 40 (a) M. S. Smyth, H. Ford, Jr., T. R. Burke, Jr., *Tetrahedron Lett.*, 1992, **33**, 4137-4140; (b) T. R. Burke, Jr., *Curr. Top. Med. Chem.*, 2006, **6**, 1465-1471

- 41 (a) T. Gilmer, M. Rodriguez, S. Jordan, R. Crosby, K. Alligood, M. Green, M. Kimery, C. Wagner, D. Kinder, P. Charifson, A. M. Hassell, D. Willard, M. Luther, D. Rusnak, D. D. Sternbach, M. Mehrotra, M. Peel, L. Shampine, R. Davis, J. Robbins, I. R. Patel, D. Kassel, W. Burkhart, M. Moyer, T. Bradshaw, J. Berman, *J. Biol. Chem.*, 1994, **269**, 31711-31719.
- 42 K. Shen, Y. -F. Keng, L. Wu, X. -L. Guo, D. S. Lawrence, Z. -Y. Zhang, *J. Biol. Chem.*, 2001, **276**, 47311-47319.
- 43 I. G. Boutselis, X. Yu, Z. -Y. Zhang, B. F. Borch, *J. Med. Chem.*, 2007, **50**, 856-864.
- 44 G. Lui, Z. Xin, H. Liang, C. Abad-Zapatero, P. J. Hajduk, D. A. Janowick, B. G. Szczepankiewics, Z. Pei, C. W. Hutchins, S. J. Ballaron, M. A. Stashko, T. H. Lubben, C. E. Berg, C. M. Rondinone, J. M. Trevillyan, M. R. Jirousek, *J. Med. Chem.*, 2003, **46**, 3437-3440.
- 45 Z. L. Xin, G. Liu, C. Abad-Zapatero, Z. H. Pei, P. J. Hajduk, S. J. Ballaron, M. A. Stashko, T. H. Lubben, J. M. Trevillyan, M. R. Jirousek, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 3947-3950.
- 46 (a) Z. J. Yao, C. R. King, T. Cao, J. Kelley, G. W. A. Milne, J. H. Voigt, T. R. Burke, *J. Med. Chem.*, 1999, **42**, 25-35; (b) T. R. Burke, J. Luo, Z. J. Yao, Y. Gao, H. Zhao, G. W. A. Milne, R. B. Guo, J. H. Voigt, C. R. King, D. J. Yang, *Bioorg. Med. Chem. Lett.*, 1999, **9**, 347-352; (c) Z. -J. Yao, Y. Gao, J. H. Voigt, H. Ford, T. R. Burke, *Tetrahedron*, 1999, **55**, 2865-2874; (d) Y. Gao, J. Luo, Z. -J. Yao, R. Guo, H. Zou, J. Kelley, J. H. Voigt, D. Yang, T. R. Burke, *J. Med. Chem.*, 2000, **43**, 911-920.
- 47 C. Liljebris, S. D. Larsen, D. Ogg, B. J. Palazuk, J. E. Bleasdale, *J. Med. Chem.*, 2002, **45**, 1785-1798.

- 48 (a) P. J. Ala, L. Gonneville, M. C. Hillman, M. Becker-Pasha, M. Wei, B. G. Reid, R. Klabe, E. W. Yue, B. Wayland, B. Douty, P. Polam, Z. Wasserman, M. Bower, A. P. Combs, T. C. Burn, G. F. Hollis, R. Wynn, *J. Biol. Chem.*, 2006, **281**, 32784-32795; (b) E. W. Yue, B. Wayland, B. Douty, M. L. Crawley, E. McLaughlin, A. Takvorian, Z. Wasserman, M. J. Bower, M. Wei, Y. L. Li, P. J. Ala, L. Gonneville, R. Wynn, T. C. Burn, P. C. C. Liu, A. P. Combs, *Bioorg. Med. Chem.*, 2006, **14**, 5833-5849.
- 49 M. Abdo, S. Liu, B. Zhou, C. D. Walls, L. Wu, S. Knapp, Z. -Y. Zhang, *J. Am. Chem. Soc.*, 2008, **130**, 13196-13197.
- 50 K. A. Kalesh, L. P. Tan, K. Liu, L. Gao, J. Wang, S. Q. Yao, *Chem. Commun.*, 2010, **46**, 589-591.
- 51 (a) M. J. Evans, B. F. Cravatt, *Chem. Rev.*, 2006, **106**, 3279-3301; (b) M. Uttamchandani, J. Li, H. Sun, S. Q. Yao, *ChemBioChem*, 2008, **9**, 667-675; (c) M. Fonovic, M. Bogyo, *Curr. Pharm. Des.*, 2007, **13**, 253-261.
- 52 (a) Y. Liu, M. P. Patricelli, B. F. Cravatt, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 14694-14699; (b) D. Kidd, Y. Liu, B. F. Cravatt, *Biochemistry*, 2001, **40**, 6107-6155; (c) M. P. Patricelli, D. K. Giang, L. M. Stamp, J. J. Burbaum, *Proteomics*, 2001, **1**, 1067-1071.
- 53 (a) D. Greenbaum, K. F. Medzihradzsky, A. Burlingame, M. Bogyo, *Chem. Biol.*, 2000, **7**, 569-581; (b) A. Borodovsky, H. Ovaa, N. Kolli, T. Gan-Erdene, K. D. Wilkinson, H. L. Ploegh, B. M. Kessler, *Chem. Biol.*, 2002, **9**, 1149-1159; (c) L. E. Edginton, B. J. van Raam, M. Verdoes, C. Wierschem, G. S. Salvesen, M. Bogyo, *Chem. Biol.*, 2012, **19**, 340-350.
- 54 (a) A. E. Speers, G. C. Adam, B. F. Cravatt, *J. Am. Chem. Soc.*, 2003, **125**, 4686-4687; (b) H. Ovaa, P. F. Van Swieten, B. M. Kessler, M. A.

- Leeuwenburgh, E. Fiebiger, A. M. C. H. van den Nieuwendijk, P. J. Galardy, G. A. van der Marcel, H.L. Ploegh, H. S. Overkleeft, *Angew. Chem. Int. Ed.*, 2003, **42**, 3626-3629; (c) A. E. Speers, B. F. Cravatt, *Chem. Biol.*, 2004, **11**, 535-546.
- 55 J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 16793-16797.
- 56 S. T. Laughlin, J. M. Baskin, S. L. Amacher, C. R. Bertozzi, *Science*, 2008, **320**, 664-667.
- 57 (a) H. C. Hang, C. Yu, D. L. Kato, C. R. Bertozzi, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 14846-14851; (b) E. M. Sletten, C. R. Bertozzi, *Acc. Chem. Res.*, 2011, **44**, 666-676; (c) T. Kurpiers, H. D. Mootz, *Angew. Chem. Int. Ed.*, 2009, **48**, 1729-1731.
- 58 L. I. Willems, N. Li, B. I. Florea, M. Ruben, G. A. van der Marcel, H. S. Overkleeft, *Angew. Chem. Int. Ed.*, 2012, DOI: 10.1002/anie.201200923
- 59 K. A. Kalesh, H. Shi, J. Ge, S. Q. Yao, *Org. Biol. Chem.*, 2010, **8**, 1749-1762
- 60 Y. Tanaka, M. R. Bond, J. J. Kohler, *Mol. BioSyst.*, 2008, **4**, 473-480.
- 61 (a) A. Saghatelian, N. Jessani, A. Joseph, M. Humphrey, B. F. Cravatt, *Proc. Natl. Acad. Sci. USA.*, 2004, **101**, 10000-10005; (b) E. W. S. Chan, S. Chattopadhyaya, R. C. Panicker, X. Huang, S. Q. Yao, *J. Am. Chem. Soc.*, 2004, **126**, 14435-14446; (c) H. Fuwa, Y. Takahashi, Y. Konno, N. Watanabe, H. Miyashita, M. Sasaki, H. Natsugari, T. Kan, T. Fukuyama, T. Tomita, T. Iwatsubo, *ACS Chem. Biol.*, 2007, **2**, 408-418; (d) H. Shi, K. Liu, A. Xu, S. Q. Yao, *Chem. Commun.*, 2009, 5030-5032; (e) K. Liu, H. Shi, H. Xiao, A. G. L. Chong, X. Bi, Y. T. Chang, K. Tan, R. Y. Yada, S. Q. Yao, *Angew. Chem. Int.*

- Ed.*, 2009, **48**, 8293-8297; (f) K. A. Kalesh, S. B. D. Sim, J. Wang, K. Liu, Q. Lin, S. Q. Yao, *Chem. Commun.*, 2010, **46**, 1118-1120; (g) H. Shi, C. Zhang, G. Y. J. Chen, S. Q. Yao, *J. Am. Chem. Soc.*, 2012, **134**, 3001-3014.
- 62 (a) L. C. Lo, T. L. Pang, C. H. Kuo, Y. L. Chiang, H. Y. Wang, J. J. Lin, *J. Proteome Res.*, 2002, **1**, 35-40; (b) L. C. Lo, Y. L. Chiang, C. H. Kuo, H. K. Liao, Y. J. Chen, J. J. Lin, *Biochem. Biophys. Res. Commun.*, 2005, **326**, 30-35.
- 63 (a) J. R. Betley, S. Cesaro-Tadic, A. Mekhalia, J. H. Richard, H. Denham, L. J. Partridge, A. Pluckthun, G. M. Blackburn, *Angew. Chem. Int. Ed.*, 2002, **41**, 775-777; (b) Q. Zhu, X. Huang, G. Y. J. Chen, S. Q. Yao, *Tetrahedron Lett.*, 2003, **44**, 2669-2672; (c) Y. Y. Huang, C. C. Kuo, C. Y. Chu, Y. H. Huang, Y. L. Hu, J. J. Lin, L. C. Lo, *Tetrahedron*, 2010, **66**, 4521-4529; (d) J. Jiang, D. Zeng, S. Li, *ChemBioChem*, 2009, **10**, 635-638.
- 64 (a) S. Kumar, B. Zhou, F. Liang, W. -Q. Wang, Z. Huang, Z. -Y. Zhang, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 7943-7948; (b) S. Kumar, B. Zhou, F. Liang, H. Yang, W. -Q. Wang, Z. -Y. Zhang, *J. Proteome Res.*, 2006, **5**, 1898-1905.
- 65 S. Liu, B. Zhou, H. Yang, Y. He, Z. -X. Jiang, S. Kumar, L. Wu, Z. -Y. Zhang, *J. Am. Chem. Soc.*, 2008, **130**, 8251-8260.
- 66 (a) J. A. Wells, C. L. McLendon, *Nature*, 2007, **450**, 1001-1009; (b) H. Yin, A. D. Hamilton, *Angew. Chem. Int. Ed.*, 2005, **44**, 4130-4163.
- 67 (a) T. Berg, *Angew. Chem. Int. Ed.*, 2003, **42**, 2462-2481; (b) A. J. Wilson, *Chem. Soc. Rev.*, 2009, **38**, 3289-3300.
- 68 (a) J. Turkson, J. S. Kim, S. Zhang, J. Yuan, M. Huang, M. Glenn, E. Haura, S. Sebti, A. D. Hamilton, R. Jove, *Mol. Cancer. Ther.*, 2004, **3**, 261-269; (b) J. Turkson, D. Ryan, J. S. Kim, Y. Zhang, Z. Chen, E. Haura, A. Laudano, S.

- Sebti, A. D. Hamilton, R. Jove, *J. Biol. Chem.*, 2001, **276**, 45443-45455; (c) K. A. Z. Siddiquee, P. T. Gunning, M. Glenn, W. P. Katt, S. Zhang, C. Schroeck, S. M. Sebti, R. Jove, A. D. Hamilton, J. Turkson, *ACS Chem. Biol.*, 2007, **2**, 787-798; (d) K. Siddiquee, S. Zhang, W. C. Guida, M. A. Blaskovich, B. Greedy, H. R. Lawrence, M. L. Yip, R. Jove, M. M. McLaughlin, N. J. Lawrence, S. M. Sebti, J. Turkson, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 7391-7396; (e) P. K. Mandal, W. S. -L. Liao, J. S. McMurry, *Org. Lett.*, 2009, **11**, 3394-3394.
- 69 a) M. Chatterjee-Kishore, F. V. D. Akker and G. R. Stark, *Trends Cell Biol.*, 2000, **10**, 106-111; (b) D. E. Levy and J. E. Darnell, Jr, *Nat. Rev. Mol. Cell Biol.*, 2002, **3**, 651-662.
- 70 (a) J. Kuriyan, D. Cowburn, *Annu. Rev. Biophys. Biomol. Struct.*, 1997, **26**, 259-288; (b) P. Tony, R. Linding, *FEBS Lett.*, 2005, **579**, 1808-1814.
- 71 S. Zhou, S. E. Shoelson, M. Chaudhuri, G. Gish, T. Pawson, W. G. Haser, F. King, T. Roberts, S. Ratnofsky, R. J. Lechleider, B. G. Neel, R. B. Birge, J. E. Fajardo, M. M. Chou, H. Hanafusa, B. Schaffhausen, L. C. Cantley, *Cell*, 1993, **72**, 767-778.
- 72 (a) E. A. Lunney, K. S. Para, R. J. Ronald, C. Humblet, J. H. Fergus, J. S. Marks, T. K. Sawyer, *J. Am. Chem. Soc.*, 1997, **119**, 12471-12476; (b) R. Sundaramoorthi, N. Kawahata, M. G. Yang, W. C. Shakespeare, C. A. Metcalf. III, Y. Wang, T. Merry, C. J. Eyermann, R. S. Bohacek, S. Narula, D. C. Dalgarna, T. K. Sawyer, *Biopolymers*, 2003, **71**, 717-729; (c) Y. -Q. Long, T. Xue, Y. -L. Song, Z. -L. Liu, S. -X. Huang, Q. Yu, *J. Med. Chem.*, 2008, **51**, 6371-6380; (d) T. R. Burke, K. Lee, *Acc. Chem. Res.*, 2003, **36**, 426-433.
- 73 H. Yu, R. Jove, *Nat. Rev. Cancer*, 2004, **4**, 97-105.

- 74 J. F. Bromberg, J. F.; M. H. Wrzeszczynska, G. Devgan, Y. Zhao, R. G. Pestell, C. Albanese, J. E. Darnell, Jr., *Cell*, 1999, **98**, 295-303.
- 75 K. E. Muratore, P. A. Cole, *ACS Chem. Biol.*, 2007, **2**, 454-456.
- 76 G. Liu, Z. Xin, Z. Pei, P. J. Hajduk, C. A. Zapatero, C. W. Hutchins, H. Zhao, T. H. Lubben, S. J. Ballaron, D. L. Haasch, W. Kaszubska, C. M. Rondinone, J. M. Trevillyan, M. R. Jirousek, *J. Med. Chem.*, 2003, **46**, 4232-4235.
- 77 (a) R. Srinivasan, M. Uttamchandani, S. Q. Yao, *Org. Lett.*, 2006, **8**, 713-716;
(b) R. Srinivasan, L. P. Tan, H. Wu, S. Q. Yao, *Org. Lett.*, 2008, **10**, 2295-2298.
- 78 H. Y. Lei, M. S. Stoakers, K. P. B. Herath, J. H. Lee, A. W. Schwabacher, *J. Org. Chem.*, 1994, **59**, 4206-4210.
- 79 F. Himo, T. Lovell, R. Hilgraf, V. V. Rostovtsev, L. Noodleman, K. B. Sharpless, V. V. Fokin, *J. Am. Chem. Soc.*, 2005, **127**, 210-216.
- 80 M. B. Soellner, K. A. Rawls, C. Grundner, T. Alber, J. A. Ellman, *J. Am. Chem. Soc.*, 2007, **129**, 9613-9615.
- 81 M. E. Vazquez, J. B. Blanco, B. Imperiali, *J. Am. Chem. Soc.*, 2005, **127**, 1300-1306.
- 82 (a) S. Becker, B. Groner, C. W. Muller, *Nature*, 1998, **394**, 145-151; (b) D. R. Coleman, IV, Z. Ren, P. K. Mandal, A. G. Cameron, G. A. Dyer, S. Muranjan, M. Campell, X. Chen, J. S. McMurray, *J. Med. Chem.*, 2005, **48**, 6661-6670.
- 83 (a) C. H. S. Lu, H. Sun, F. B. A. Bakar, M. Uttamchandani, W. Zhou, Y. -C. Liou, S. Q. Yao, *Angew. Chem. Int. Ed.*, 2008, **47**, 7438-7441; (b) <http://www.openbiosystems.com> .
- 84 J. Schust, T. Berg, *Anal. Biochem.*, 2004, **330**, 114-118.

- 85 This value is slightly higher than the reported 42 μ M obtained from a DNA-binding assay,^{68a} likely due to the difference in the assay type.
- 86 J. D. Irvine, L. Takahashi, K. Lockhart, J. Cheong, J. W. Tolan, H. E. Selick, J. R. Grove, *J. Pharm. Sci.*, 1999, **88**, 28-33
- 87 (a) L. P. Tan, H. Wu, P. -Y. Yang, K. A. Kalesh, X. Zhang, M. Hu, R. Srinivasan, S. Q. Yao, *Org. Lett.*, 2009, **11**, 5102-5105; (b) H. Sun, L. P. Tan, L. Gao, S. Q. Yao, *Chem. Commun.*, 2009, 677-679.
- 88 Z. -Y. Zhang, *Annu. Rev. Pharmacol. Toxicol.*, 2002, **42**, 209-234.
- 89 I. A. Yudushkin, A. Schleifenbaum, A. Kinkhabwala, B. G. Neel, C. Schultz, P. I. H. Bastiaens, *Science*, 2007, **315**, 115-119.
- 90 M. Hu, L. Li, H. Wu, Y. Su, P. -Y. Yang, M. Uttamchandani, Q. H. Xu, S. Q. Yao, *J. Am. Chem. Soc.*, 2011, **133**, 12009-12020.
- 91 (a) S. Welte, K. -H. Baringhaus, W. Schmider, G. Müller, S. Petry, N. Tennagels, *Anal. Biochem.*, 2005, **338**, 32-38; (b) ELF 97, another popular PTP-detecting reagent (www.invitrogen.com), alleviate this problem by forming highly fluorescent precipitates near the reaction site. But this chemical works only on fixed cells due to its cell-impermeable property.
- 92 D. H. Kwan, H. -M. Chen, K. Ratananikom, S. M. Hancock, Y. Watanabe, P. T. Kongsaree, A. L. Samuels, S. G. Withers, *Angew. Chem. Int. Ed.*, 2011, **55**, 300-303.
- 93 D. D. Young, A. Deiters, *Org. Biomol. Chem.*, 2007, **5**, 999-1005
- 94 R. M. Naik, V. M. Thakor, *J. Org. Chem.*, 1957, **22**, 1626-1629.
- 95 J. -P. Pellois, T. W. Muir, *Angew. Chem. Int. Ed.*, 2005, **44**, 5713-5717.
- 96 L. Ellgaard, A. Hlenius, *Nature Rev. Mol. Cell Biol.*, 2003, **4**, 181-191.
- 97 Y. Loh, H. Shi, M. Hu, S. Q. Yao, *Chem. Commn.*, 2010, **46**, 8407-8409

- 98 T. O. Johnson, J. Ermolieff, M. R. Jirousek, *Nat. Rev. Drug Disc.*, 2002, **1**, 696-709
- 99 A. E. Greenstein, C. Grundner, N. Echols, L. M. Gay, T. N. Lombana, C. A. Miecskowski, K. E. Pullen, P. Y. Sung, T. Alber, *J. Mol. Microbiol. Biotechnol.*, 2005, **9**, 167-181.
- 100 (a) K. T. Barglow, B. F. Cravatt, *Nature Meth.*, 2007, **4**, 822-827; (b) M. Fonovic, M. Bogyo, *Exp. Rev. Proteomics*, 2008, **5**, 721-730; (c) W. P. Heal, T. H. T. Dang, E. W. Tate, *Chem. Soc. Rev.*, 2011, **40**, 246-257.
- 101 H. Shi, X. -M. Cheng, S. K. Sze, S. Q. Yao, *Chem. Commun.*, 2011, **47**, 11306-11308
- 102 X.-L. Guo, K. Shen, F. Wang, D. S. Lawrence, Z.-Y. Zhang, *J. Biol. Chem.*, 2002, **277**, 41014-41022.
- 103 (a) A. Dömling, *Chem. Rev.*, 2006, **106**, 17-89; (b) A. Irini, *Curr. Opin. Chem. Biol.*, 2008, **12**, 324-331; (c) A. Dömling, I. Ugi, *Angew. Chem. Int. Ed.*, 2000, **39**, 3168-3210.
- 104 (a) Z. Li, S. L. Yeo, C. J. Pallen, A. Ganesan, *Bioorg. Med. Chem. Lett.*, 1998, **8**, 2443-2446; (b) G. Bergnes, C. L. Gilliam, M. D. Boisclair, J. L. Blanchard, K. V. Blake, D. M. Epstein, K. Pal, *Bioorg. Med. Chem. Lett.*, 1999, **9**, 2854-2854.
- 105 (a) E. M. Sletten, C. R. Bertozzi, *Angew. Chem. Int. Ed.*, 2009, **48**, 6974-6998; (b) K. A. Kalesh, H. Shi, J. Ge, S. Q. Yao, *Org. Biomol. Chem.*, 2010, **8**, 1749-1762.
- 106 X. Ping, W. Lin, X. Zou, *Synth.*, 2002, **8**, 1071-1026.
- 107 P. -Y. Yang, K. Liu, M. H. Ngai, M. J. Lear, M. Wenk, S. Q. Yao, *J. Am. Chem. Soc.*, 2010, **132**, 656-666.

- 108 (a) H. Sun, S. Chattopadhyaya, J. Wang, S. Q. Yao, *Anal. Bioanal. Chem.*, 2006, **386**, 416-426; (b) M. Uttamchandani, C. H. S. Lu, S. Q. Yao, *Acc. Chem. Res.*, 2009, **42**, 1183-1192.
- 109 B. F. Cravatt, A. T. Wright, J. W. Kozarich, *Annu. Rev. Biochem.*, 2008, **77**, 383-414.
- 110 D. Krishnamurthy, A. M. Barrios, *Curr. Opin. Chem. Biol.*, 2009, **13**, 375-381.
- 111 J. Ge, L. Li, S. Q. Yao, *Chem. Commun.*, 2011, **47**, 10939-10941.
- 112 W. P. Heal, M. H. Wright, E. Thinon, E. W. Tate, *Nature Prot.*, 2012, **7**, 105-117
- 113 P. -Y. Yang, M. Wang, C. Y. He, S. Q. Yao, *Chem. Commun.*, 2012, **48**, 835-837.
- 114 (a) B. Wilkinson, H. F. Gilbert, *Biochim. Biophys. Acta.*, 2004, **1699**, 35-44; (b) L. Ellgaard, L. W. Ruddock, *EMBO Rep.*, 2005, **6**, 28-32; (c) A. R. Karala; A. K. Lappi, L. W. Ruddock, *J. Mol. Biol.*, 2010, **396**, 883-892.
- 115 A. Gallina, T. M. Hanley, R. Mandel, M. Trahey, C. C. Broder, G. A. Viglianti H. J. -P. Ryser, *J. Biol. Chem.*, 2002, **277**, 50579-50688.
- 116 (a) K. J. Conn, W. Gao, A. McKee, M. S. Lan, M. D. Ullman, P. B. Eisenhauer, R. E. Fine, J. M. Wells, *Brain Res.*, 2004, **1022**, 164-172; (b) J. D. Atkin, M. A. Farg, B. J. Turner, D. Tomas, J. A. Lysaght, J. Nunan, A. Rembach, P. Nagley, P. M. Beart, S. S. Cheema, M. K. Horne, *J. Biol. Chem.*, 2006, **281**, 30152-30165; (c) B. G. Hoffstrom, A. Kaplan, R. Letso, R.S. Schmid, G. J. Turmel, D. C. Lo, B. R. Stockwell, *Nature Chem. Biol.*, 2010, **6**, 900-906.
- 117 A. M. Smith, J. Chan, D. Oksenberg, R. Urfer, D. S. Wexler, A. Ow, L. Gao, A. McAlorum, S. -G. Huang, *J. Biomol. Screen*, **2004**, **9**, 614-620

- 118 J. Lundström, A. Holmgren, *J. Biol. Chem.*, 1990, **265**, 9114-9120.
- 119 N. J. Darby, R. B. Freedman, T. E. Creighton, *Biochemistry*, 1994, **33**, 7937-7947
- 120 T. Finkel, *Curr. Opin. Cell Biol.*, 2003, **15**, 247-254.
- 121 M. B. Francis, T. Nguyen, *Org. Lett.*, 2003, **5**, 3245-3248.
- 122 J. Wang, M. Uttamchandani, L. P. Sun, S. Q. Yao, *Chem. Commun.*, 2006, 717-719.
- 123 <http://www.uniprot.org/uniprot/>