Synthesis of Targeted Probes for Chemical Proteomics Studies of the EGFR Signaling Network

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

by

Ivan de Jesus Salazar Estrada

Master of Research (Chemistry)

MACQUARIE
University

Department of Molecular Sciences

Macquarie University

Sydney, Australia

July 2019

TABLE OF CONTENTS

ABSTRACT

The Epidermal growth factor receptor (EGFR) kinase is a prototypical receptor kinase with critical function in normal homeostasis and disease progression. Eight EGFR inhibitors spanning three generations are FDA-approved for a few types of cancers and ultimately limited by resistance. Drug resistance has recently been associated to EGFR roles beyond its kinase activity and dependent on subcellular localization. Elucidating endogenous EGFR signaling characteristics may lead to more efficacious therapies.

In this thesis work we report the design and synthesis of EGFR-directed probes for taking a chemical proteomics approach to finding potential protein partners of EGFR in the cell. The probes were based on selective EGFR inhibitors and designed to covalently label a lysine residue in the solvent-exposed region of EGFR. Several trifunctional linkers for the modification of other drugs were also obtained. In reverse chemical proteomics using human cDNA libraries from various cancer cells, DNA topoisomerase I was identified as a potential target of the EGFR inhibitor Gefitinib, supporting recent hypothesis of EGFR-topoisomerase crosstalk in drug resistance. In forward chemical proteomics studies in MDA-MB-468 cells, several ATP-utilizing enzymes were identified as potential EGFR interactors. These proteins may in the future be further investigated to validate their interactions with EGFR and understand the biological roles of these interactions to assist the development of new combination therapies.

STATEMENT OF ORIGINALITY

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

July 3, 2019

Ivan de Jesus Salazar Estrada

(SN 43831702)

DECLARATION

The research presented in this thesis was approved by Macquarie University Institutional Biosafety Committee:

- Approval number NLRD 5201600329 (T7 phage display)

- Project ID 0340 – 520180340968 (mammalian cell culture)

ACKNOWLEDGEMENTS

I'm grateful to the numerous people that have contributed directly to the realization of this project, and those who have supported me with their love and friendship.

To my supervisor, Dr. Fei Liu. You made possible one of my longest-desired dreams. Thanks for introducing me to the fascinating world of EGFR and chemical biology. Thanks also for your patience with this annoying student that loved to showed up at your office quite often. Thanks also for all the encouragement that you gave me during the project, especially on times of "failed" experiments when I thought I should not do research any more.

A special thanks also to my co-supervisor, Prof. Peter Karuso as well as Dr. Andrew Piggott and Associate Prof. Koushik Venkatesan for having adopted me into their labs after ours was consumed by a fire, literally! The interdisciplinary nature of this project was both challenging and exciting at the same time. I am thankful to all the research staff that trained me into the different techniques and guided the first steps of this chemist to do Biology: Dr. Jenny Vo (phage display); Dr. Charlie Ahn and Dr. Abidali Mohamedali (SDS-PAGE, mammalian cell culture); Dr. Karthik Kamath and Dr. Matthew McKay (proteomics). Thanks also to the technical staff for helping me solve the hands-on problems with the different instrumentation.

To my fellow labmates: Ryan, Harry, Alex, Sviat, Payal and Lubna. Thanks for your advice on the different experimental problems I encountered, for the philosophical talks, for encouragement in times of distress and in general for keeping a nice environment in the lab.

Thanks also to Macquarie University for my scholarship, without which I could have not started my graduate studies. Thanks to the people of Australia, who with their hard work made possible

this scholarship expecting nothing in return but the advancement of science. Be assured your efforts have not been in vain.

Pursuing an experimental doctoral degree can be a very isolating process; I was blessed with many friends that make of my graduate studies an incredible experience. Thanks to my friends from the Capuching Young Adults, Macquarie University Catholic Society and SVD Community (Divine Music Lovers): Vy, Bernie, Dara, Chris and Cath, Johanne and Krystle, Vanessa, Elva, Tony, Herson, Nicholas, Arturo, Alejandro, Juan, Christine, Pia, Bianca. Thanks for reminding me that what matters most in this life is not how many things I accomplish, but how much I can love; that I am loved for who I am, and not for what I do. Your faith has had a great impact in my life. You have made visible for me the communion of saints. I strongly believe we are starting to live heaven here on earth. I look forward to that day when we will praise the Lord without ceasing in the wedding banquet he has prepared for us (remember to intercede for me if you go first).

Thanks also to the numerous priest (diocesan, carmelites, dominicans, capuchins) that have nurtured me with the body and blood of Christ. Special thanks to Fr. Epeli, Fr. Iñigo, Fr. Thomas Azzi, Fr. Dom Murphy, Fr. Thomas McFadden, Fr. Vincy, Fr. Pio, Fr. Justin and Fr. Daniele, who have followed more closely my journey with the Lord. Thanks for showing me with your example what it means to be a true father and a true man, and the joy that can be found in laying down your life for Christ.

To my friends in heaven for their intercession, especially St. Francis, St. Dominic, St. Thomas Aquinas, St. Albert the Great and St. John Paul II. A special thanks to blessed Guadalupe Ortiz de Landazuri, who in these last months has been a great inspiration for my work. There is a great joy in knowing that every experiment I do, every hour of planning and studying is an opportunity to live in sainthood, to praise the Lord and to lead souls to him. I'm so happy to know I have a

colleague officially in heaven, can't wait to be there as well! Thanks also to my mother Mary, refuge of sinners, who as a good mother intercedes quietly and constantly for me.

To my professors back in Mexico, who encouraged me to pursue this path of life which I was given the opportunity to pursue. Thanks as well to my grandparents, uncles, aunts and relatives that have supported me physically or spiritually at different stages in my formation.

To my parents and my sister, for your unconditional love and for encouraging me to always follow my dreams. I love you. I know there have been times of great difficulty and fear, but the Lord is always with us, do not be afraid. You have given to me the greatest gift that I have ever received, and without which nothing else would make sense: my faith.

Last to be mentioned but first in order, thanks to the mastermind behind everything that exist and is truth, existence and truth itself, the origin and end of all creatures: God. Thanks for my body that manifests your glory making visible the invisible. Thanks for my soul that allows me to know and seek truth, beauty and goodness. Thanks for Chemistry and Mathematics, with which you have written the universe and life, and for your Divine Providence that governs them. Thanks because through the beauty of nature you have called me "to contemplate and share with others the fruit of the contemplation". Above all else, thanks for your unconditional love, from which everything else emerges. Everything I have and am I have received from you. I give it all back to you, so that in everything you may be glorified. *Non nisi te, Domine.*

DEDICATION

In loving memory of Gloria Estrada,

Oneida De la Rosa and Juan De la Rosa:

In hope that my work may one day contribute to alleviate the suffering caused by cancer and with faith that I'll see you again. May my work be always guided by love and expressed in service to my brothers and sisters.

Ad maiorem Dei gloriam
inque hominum salute

"For it is he who gave me unerring knowledge of what exists, that I may know the structure of the world and the activity of its elements. I learned both what is secret and what is manifest, for Wisdom, the artificer of all, taught me"

Wis 7:17, 21-22

ABBREVIATIONS

°C	degrees celsius
4Å-MS	4Å molecular sieves
Å	angstrom
ABP	activity based probe
ABPP	activity-based proteome profiling
AcOH	acetic acid
ADAMs	A desintegrin and metalloproteases
A*f*BP	affinity based probe
AHR	aryl hydrocarbon receptor
AKT	protein kinase B
ALDH1A1	Retinal dehydrogenase
aq.	aqueous
AR	Amphiregulin
ATP	adenosine 5'-triphosphate
AXL	receptor tyrosine protein kinase UFO
BC	breast cancer
BLAST	basic local alignment search tool
Bn	benzyl
bp	Base pair
br	broad
BTC	betacellulin
CAN	ammonium cerium(IV) nitrate
Cbl	casita B-lineage lymphoma ubiquitin protein ligase
CCCP	compound centric chemical proteomics
cDNA	complementary DNA
CDS	coding region
CRC	colorectal cancer
CuAAC	copper(I)-catalyzed azide alkyne cicloadition
d	doublet
d₆	hexadeuterated
DCM	dichloromethane
dd	doublet of doublets
DDQ	2,3-dichloro-5,6-dicyano-1,4-benzoquinone
DFG	Asp-Phe-Gly motif
DIPEA	*N,N*-diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	*N,N*-dimethylformamide
DMSO	DMSO
DNA	deoxyribonucleic acid
DTT	dithiothreitol

DUS2L	tRNA-dihydrouridine(20) synthase [NAD(P)+]-like protein
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
EGN	epigen
EPR	epiregulin
ER	estrogen receptor
ErbB	epidermal growth factor receptor family of proteins
ErbB2/3/4	human epidermal growth factor receptor 2/3/4
ERK	extracellular-signal regulated kinase
ESI	electrospray ionisation
Et_2O	diethyl ether
ETDA	Ethylenediaminetetraacetic acid
EtOAc	ethyl acetate
F1,6BP	fructose 1,6-biphosphate
FA	formic acid
FDA	United States Food and Drug Administration
g	gram
GBM	glioblastoma
GC	gastric cancer
h	hour
HB-EGF	heparin-binding EGF-like receptor
HEPES	N-(2-Hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid)
HER2/3/4	human epidermal growth factor receptor 2/3/4
HK2	hexokinase 2
hnRNP	heteronuclear ribonucleoprotein
HNSCC	head and neck squamous cell carcinoma
HPLC	high-performance liquid chromatography
HSQC	heteronuclear single bond correlation spectroscopy
Hz	hertz
ICAT	Isotope-coded affinity tag
IPTG	β-D-1-thiogalactopyranoside
ITRAQ	Isobaric tag for relative and absolute quantitation
J	coupling constant
K_d	dissociation constant
LC-MS	liquid chromatography-mass spectrometry
LC-MS/MS	liquid chromatography-tandem mass spectrometry
LRIG1	leucine-rich repeats and immunoglobulin-like domains-1
m	multiplet
m/z	mass to charge ratio
Me	methyl
MEK	mitogen activated protein kinase

MeOH	methanol
MET	hepatocyte growth factor receptor
mg	milligram
MHz	megahertz
MIG-6	mitogen-inducible gene-6 protein
min	minute
mmol	milimole
mRNA	messenger RNA
Ms	methanesulfonyl
MS	mass spectrometry
NHS	*N*-hydroxysuccinimide
nM	nanomolar
NMR	Nuclear magnetic resonance spectroscopy
NRG	neuregulin
NSAF	Normalized spectral abundance factor
NSCLC	non-small-cell lung cancer
OD_{600}	absorbance at 600 nm
PBS	phosphate buffered saline
PC	pancreatic cancer
PCR	Polymerase chain reaction
PEG	polyethylene glycol
PFKP	phosphofructokinase, platelet type
PI3K	phosphoinositide 3-kinase
PIP	phosphatidylinositol phosphate
PIP4K2C	phosphatidylinositol 5-phosphate 4-kinase
PK	protein kinase
PKM2	pyruvte kinase M2
PMB	*p*-methoxybenzyl
PPI	protein-protein interactions
ppm	parts per million
PR	progesteron receptor
PTB	phosphotyrosine binding domain
PTC	phase transfer catalyts
PTGES2	Prostaglandin E 2 synthase
PTM	post-translational modification
PWB	phage wash buffer
Raf	RAF serine/threonine protein kinase family
Ras	rat sarcoma GTPase family
RNA	ribonucleic acid
RNAi	small interfering RNA
rt	room temperature
RTPK	receptor tyrosine protein kinase

s	singlet
SDS	sodium dodecylsulfate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SFK	Src family kinase
SH2	Src homology 2 domain
SILAC	stable isotope labelling by amino acids in cell culture
S_N2	bimolecular nucleophilic substitution
S_NAr	nucleophilic aromatic substitution
SPRY	sprouty adaptor protein
Src	sarcoma family kinases
STAT	signal transducer and activator of transcription protein
t	triplet
TAE	tris acetate EDTA buffer
TBAF	tetra-n-butylammonium fluoride
TBAI	tetra-n-butylammonium iodide
TBS	*tert*-butyldimethylsilyl
TBSOTf	*tert*-butyldimethylsilyl trifluoromethanesulfonate
TBTA	tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine
*t*BuOH	*tert*-butanol
TCEP	tris(2-carboxyethyl)phosphine
TFA	trifluoroacetic acid
TGA-α	transforming growth factor alpha
THF	tetrahydrofuran
TKD	tyrosine kinase domain
TKI	tyrosine kinase inhibitor
TNBC	Triple negative breast cancer
TOP1	DNA topoisomerase I
TsOH	*p*-toluenesulfonic acid
UV	ultraviolet
WT	wild type
XL-MS	cross-linking mass spectrometry
δ	NMR chemical shift in parts per million
μ	micro
μM	micromolar

LIST OF FIGURES

LIST OF SCHEMES

LIST OF TABLES

CHAPTER 1

Introduction: EGFR Signaling Network

and Targeted Therapies

1. 1 EGFR STRUCTURE AND PHYSIOLOGY

The cell, the basic unit of living organisms, is a highly complex chemical system with unique emergent properties. In order to perform its activities, the cell must be able to recognize, decode and translate environmental signals. Membrane receptors are a fundamental component of this cell signaling process. Along with G-protein coupled receptors (GPCRs) and cytokine receptors, receptor tyrosine protein kinases (RTK) are one of the major groups of membrane receptors. Among RTKs, the epidermal growth factor receptor (EGFR) is one of the first and most extensively studied protein kinases. EGFR plays an essential role in the normal physiology and disease progression in cells of epithelial, mesenchymal and neuronal origin. EGFR is also a validated drug target for certain types of cancer e.g. lung and colorectal cancer), with several small-molecule and antibody inhibitors currently approved for clinical use. However, drug resistance associated with these drugs is an issue that has not been possible to overcome. In this chapter, we describe EGFR structure and physiology, the signaling networks regulated by EGFR, EGFR-targeted therapies and mechanisms of drug resistance as well as chemical and biological approaches used for studying EGFR.

1.1.1 EGFR structure

The EGFR family of RTKs comprises four closely related members: EGFR (also known as ErbB1, HER1), ErbB2 (neu, HER2), ErbB3 (HER3) and ErbB4 (HER4). These proteins consist of approximately 1200 – 1300 residues. The protein architecture comprises three defined regions: an extracellular, transmembrane and intracellular domain (**Fig. 1.1**). The crystal structure of each region has been extensively reported and reviewed.[1-2] The extracellular domain is involved in ligand binding and comprises four subdomains: subdomains I and III participate directly in ligand binding, while subdomains II and IV are involved in receptor dimerization. The intracellular region comprises the enzymatic tyrosine kinase domain (TKD) flanked by a juxtamembrane segment and a C-terminal tail.
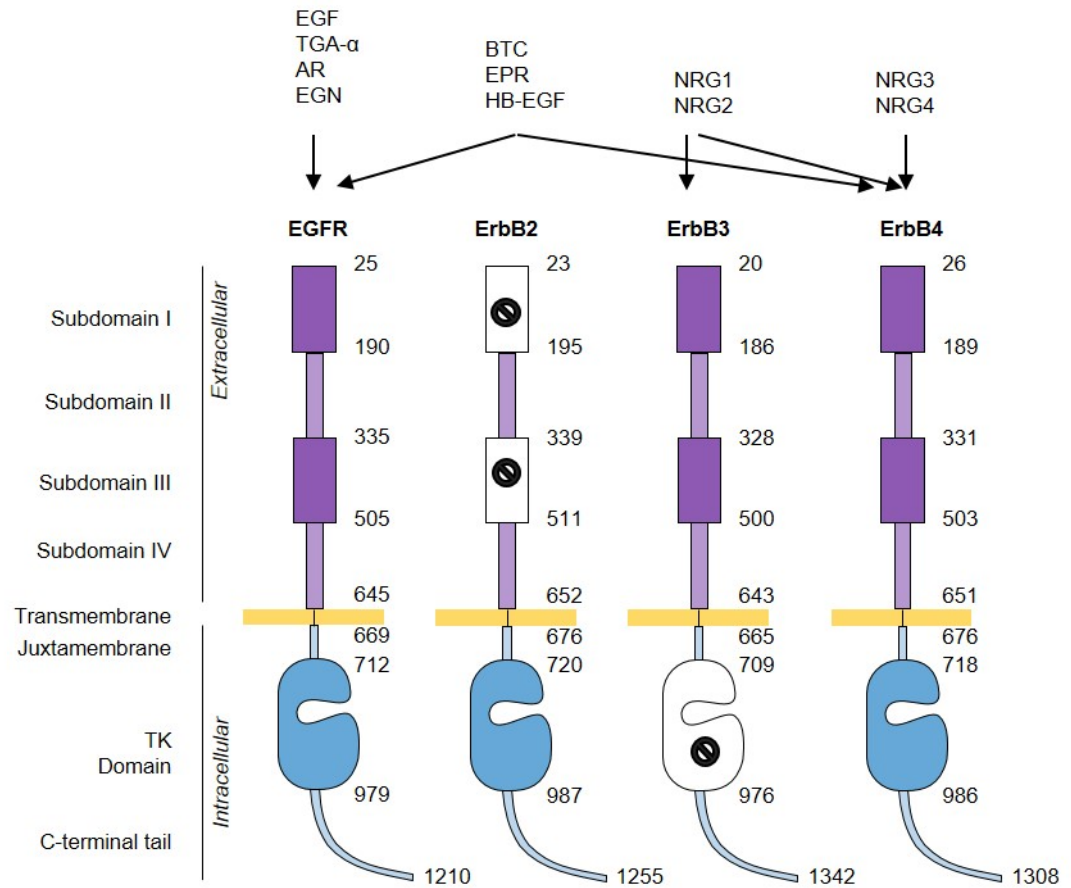
Fig. 1.1 Architecture of EGFR family of proteins

The ligands of the EGFR family comprise at least 11 members with different specificity for each receptor: epidermal growth factor (EGF), transforming growth factor α (TGF-α), amphiregulin (AR) and epigen (EGN) bind exclusively to EGFR; betacellulin (BTC), epiregulin (EPR) and heparin-binding EGF-like growth factor (HB-EGF) can bind to EGFR and HER4; neuregulins (NRG1–4) can bind to ErbB3 and ErbB4.[3] The ligands precursors are expressed as transmembrane proteins. The release and assembly of fully functional ligands is controlled by posttranslational modifications mediated by members of the A desintegrin and metalloproteases (ADAMs) family.

1.1.2 EGFR activation

As is the case for other RTKs, EGFR activation occurs through receptor dimerization triggered by ligand binding.[4-7] The elucidation of a step-by-step mechanisms has been

possible after several structural studies in the last two decades and has been recently reviewed by Kuriyan, Lemmon and Schlessinger.[8-11] A simplified schematic model is depicted in **Fig. 1.2**.
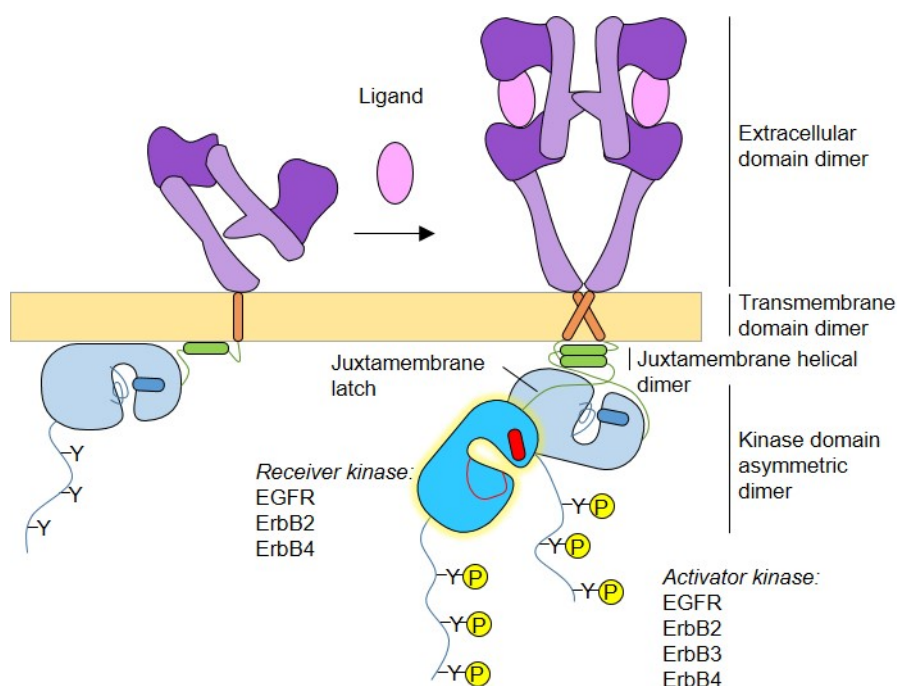
Fig. 1.2 Activation of the TKD in EGFR upon ligand binding[11]
(Adapted from Kovacs, E. et al, *Annu. Rev. Biochem.*, **2015**, *84*, 739-764)

In the absence of ligand, EGFR is mostly monomeric and enzymatically inactive. Several molecular interactions make possible this auto inhibited state: the extracellular domain is in a closed conformation, with interactions between subdomains II and IV hiding a dimerization arm.[12-13] The N and C lobes of the TKD are displaced in a Src-like inactive conformation[14] and docked against the membrane, along with a segment of the juxtamembrane, through electrostatic interactions.[15-16] Ligand binding promotes a conformational change of the extracellular domain, releasing the dimerization arm and allowing the homo or hetero-dimerization of this domain.[13, 17] This extracellular change is transmitted to the transmembrane, which also dimerizes and releases the juxtamembrane and TKD from the membrane.[15-16, 18] The TKD is then able to form an asymmetric dimer in which the C-lobe of one (the activator) docks onto the N-lobe the other (the receiver).[14] This asymmetric dimer is stabilized by the juxtamembrane through a helical dimer and a latch. This dimerization

4

promotes conformational changes in the receiver kinase, switching "on" its enzymatic activity.

Although in theory the formation of all homo and hetero dimers of the family is possible, EGFR and HER4 are the only fully functional members of the family on their own. ErbB2 is always present in an open conformation, has no known ligands and is the preferred dimerization partner of the family, while HER3 kinase activity is considered null and therefore can only act as the receiver kinase.[19]

1.1.3 EGFR signaling networks

Once activated, EGFR can phosphorylate tyrosine residues on the C-tail (either receiver or activator kinase) of the dimer, as well as on cytosolic proteins. The phosphorylated tyrosines in the receptor are then able to recruit adaptors, enzymes, transcription factors and other molecules through interactions with SH2 or PTB domains.[5, 20-21] The recruited molecules activate different signaling pathways, including the Ras/Raf/MEK/ERK pathway (involved in cell proliferation) and the PI3K/Akt pathway (involved in cell survival) (**Fig. 1.3**).[22-23] These signaling pathways in turn modify the activity of specific transcription networks. EGFR-dependent regulation of signaling pathways and transcriptional networks is finally translated into a specific physiological outcome such as: cell proliferation,[24] differentiation,[25] survival, survival or death,[26] adhesion and migration.[27]

Each receptor of the EGFR family has a unique phosphorylation pattern and therefore recruits a different set of proteins that, in part, determine different cellular outcomes (**Fig. 1.4**).[28-29] For example, EGFR is the only member that recruits Cbl, the E3 ligase enzyme involved in ubiquitination and subsequent receptor degradation. In EGFR, PI3K binds indirectly through adaptor proteins, while ErbB3 has multiple sites that recruite PI3K directly. These signalling differences are dependent on the nature of the homo or hetero dimer, as well as several other

factors such as the nature of the ligand, cellular localization and levels of proteins expression (e.g ligands, receptors and signaling proteins).[30] Different posttranslational modifications (PTMs), especially phosphorylation and ubiquitination, are also essential for signaling specificity. Ultimately, through the regulation of protein-protein interactions (PPI) these PTMs orchestrate and integrate the various signaling pathways and regulatory mechanisms in space and time, creating a unique message for each cellular context.[31-33]
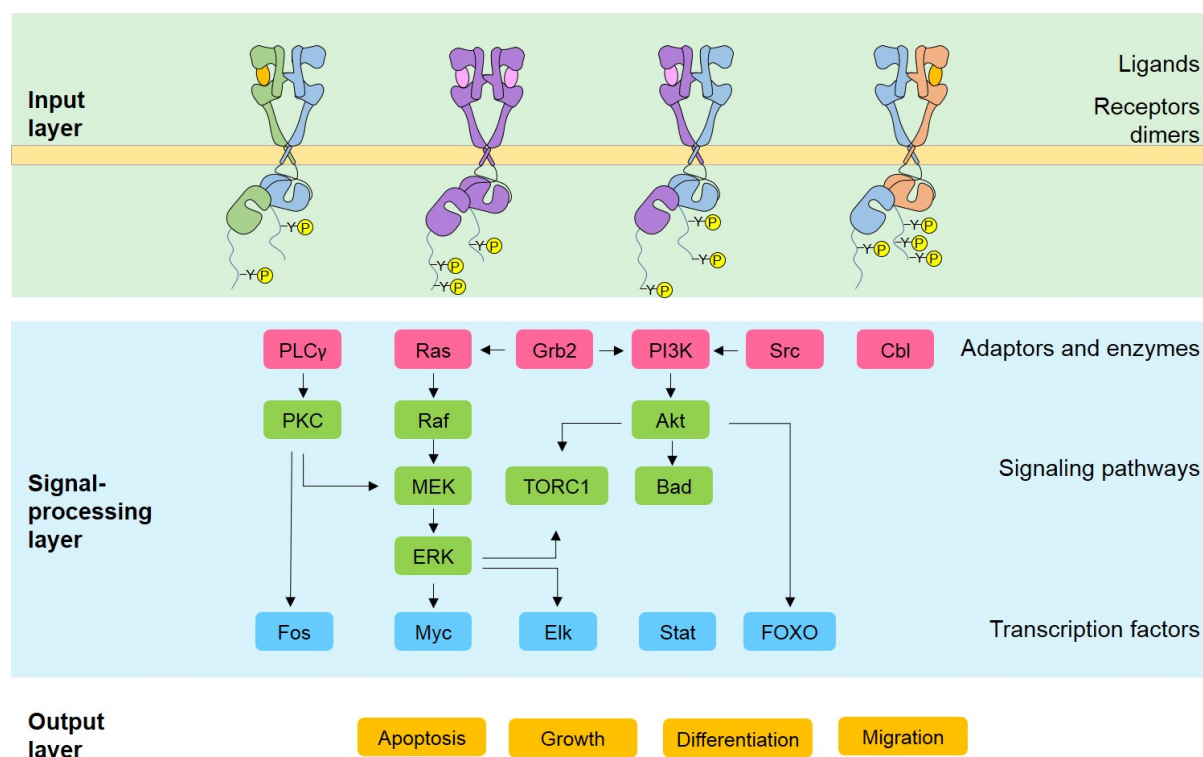
Fig. 1.3 Key binding partners and signaling pathways activated by EGFR

EGFR signaling activity must be tightly regulated to ensure the proper homeostasis of the cell. As described in activation (1.1.2) and downregulation (1.1.4) sections, multiple and diverse inhibitory mechanisms, from conformational to transcriptional control, regulate EGFR signaling activity. These mechanisms provide robustness and diversity to the EGFR signaling network and allow it to respond properly to external signals. Release of EGFR from these regulatory mechanisms often results in aberrant signaling leading to tumor formation and progression.
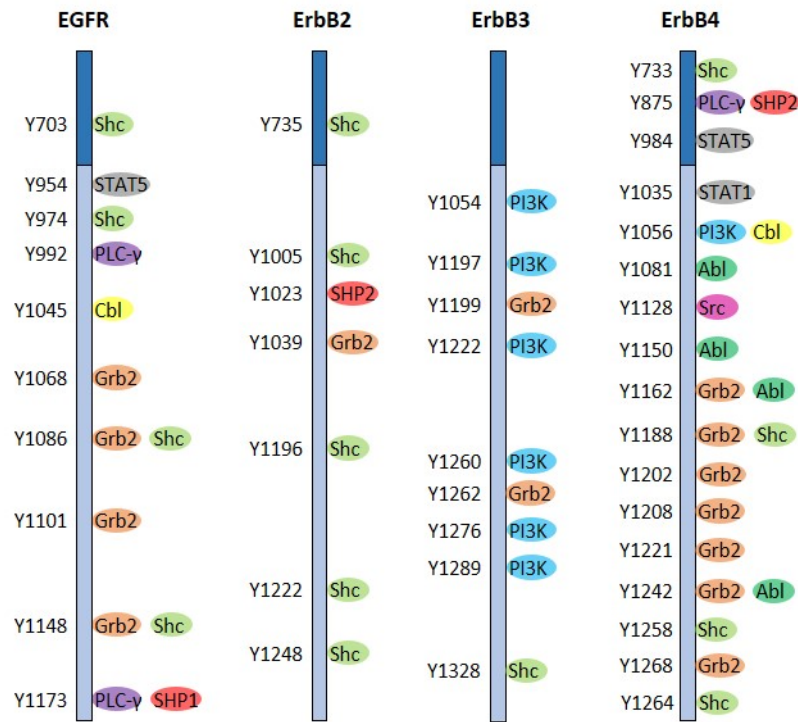
Fig. 1.4 Phosphorylation pattern and major recruited proteins in EGFR family proteins

1.1.4 EGFR downregulation

Diverse negative regulatory mechanisms are involved in "turning off" EGFR signal activity. These mechanisms can be either irreversible, involving receptor degradation, or reversible, which interfere with the capacity of the receptor to respond to ligand binding activation. Inhibitory mechanisms can also be classified according to the time upon which they act to modulate signaling activity. They can act early in the signaling process making use of molecules constitutively present in the cell, or in a delayed or feedback-mediated process which requires the transcription and synthesis of new proteins.

Reversible inhibitory mechanisms comprise both early and feedback-mediated events. Receptor compartmentalization and receptor dephosphorylating by phosphatases are early events that disrupt the physical interaction of EGFR with other effector proteins. Several proteins whose transcription is feedback regulated by EGFR activation are known to bind

EGFR and inhibit its activity; these include SPRY, LRIG1 and MIG-6 (RALT), among others.[34-37]

Irreversible EGFR degradation occurs almost exclusively through receptor ubiquitination by Cbl, which is recruited to phosphotyrosine Y1045 in EGFR. Among all downregulation mechanisms, EGFR ubiquitination is the most extensively studied; specific PTMs of numerous proteins involved in receptor trafficking and sorting, especially phosphorylation and ubiquitination, play a key role in receptor degradation.[38-41]

1.2   EGFR-TARGETED THERAPIES

Since its identification as an oncogene in the mid-1980s,[42] aberrant EGFR activity has been associated with the development of solid tumours, especially in non-small-cell lung cancer (NSCLC), glioblastoma (GBM), colorectal cancer (CRC), gastric cancer (GC), head and neck squamous cell carcinoma (HNSCC), breast cancer (BC) and pancreatic cancer (PC).[1, 43] Diverse molecular mechanisms can cause dysregulated EGFR activity.[44] These mechanisms can be intrinsic (e.g. activating mutations, overexpression) or extrinsic (e.g. ligand overexpression, alterations in regulatory proteins) to EGFR. Current EGFR-targeted therapies include antibodies targeting the extracellular ligand-binding domain and small-molecule tyrosine kinase inhibitors (TKIs) targeting the TKD. In accordance to their mode of action, antibodies are more effective in tumours dependent on ligand-activated EGFR, while TKI are more effective in tumours expressing mutant EGFR. Despite their initial effectiveness in certain types of cancer, resistance to EGFR inhibitors inevitably develops. Other types of cancer are intrinsically resistant to EGFR inhibitors in spite of presenting aberrant EGFR signaling. Understanding the mechanisms behind aberrant EGFR signaling and EGFR-targeted drug resistance is indispensable for the development of more efficacious therapies. Some of the well-established mechanisms are described below.

1.2.1 Mechanisms of EGFR dysregulation in cancer

1.2.1.1 Receptor overexpression

Elevated levels of EGFR have been reported in several types of carcinomas including head and neck, ovarian, cervical, bladder and oesophageal cancer, where is often correlated with a poor prognosis.[45] EGFR overexpression can result from numerous mechanisms, including gene amplification,[46-48] increased transcription[49] and enhanced posttranslational recycling.[50]

High levels of surface EGFR can result in constitutive receptor activation in the absence of ligand.[15] Although both receptor overexpression and ligand-binding result in an active tyrosine-phosphorylated EGFR dimer, the downstream signaling pathways activated by the dimer can be divergent and mutually exclusive.[51-52] For example, in the absence of ligand activation, EGFR overexpression does not activate the canonical pathways ERK and Akt, but regulates a non-canonical pathway through the transcription factor IRF3. Treatment of these cells with EGF disrupts the IRF3 pathway and switches the receptor towards ERK and Akt activation.

1.2.1.2 *EGFR* mutations

Oncogenic mutations in EGFR commonly arise from the deletion of segments in the extracellular domain (common in GBM)[53] or point mutations in the TKD (common in NSCLC).[54] These mutations disrupt the inhibitory mechanisms embedded in the structure of the receptor, resulting in constitutive receptor activation and signaling, usually accompanied by impaired downregulation.

EGFRvIII is the most frequent and most extensively studied mutant form of EGFR.[55] *EGFRvIII* results from the in frame deletion of exons 2-7, which encodes residues 2–273 comprising the totality of subdomain I and two thirds of subdomain II in the extracellular domain. *EGFRvIII* is found in about 40% of GBM with wild-type *EGFR* amplification.

EGFRvIII has also been reported in a fraction of head and neck, prostate, breast, ovarian and lung cancer. Several other mutant forms resulting from deletions at the extracellular (*EGFRvI/II*) or the C-tail (*EGFRvIV/V*) are found exclusively in GBM.

While deletions in the extracellular domain are typical for GBM, mutations in the TKD are characteristic for NSCLC. The most common mutations arise from in frame small deletions in exon 19, which almost always include residues L747 to E749 (ΔLRE), and the single point mutation L858R in exon 21.[56] Together, these constitute about 90% of all EGFR activating mutations in NSCLC. These mutations were discovered in 2004 after identifying a group of patients that were more responsive to small-molecule TKI.[57-59]

1.2.1.3 Alterations in EGFR ligands

The various EGFR ligands have different effects in receptor activation and processing which have important implications in their oncogenic potential. For example, EGF binding promotes EGFR lysosomal degradation limiting its signaling amplitude, while TGF-α binding leads to EGFR recycling and continuous signalling, making TGF-α a more potent mitogen. These differences have been attributed to different ligand-receptor binding stabilities in the endosome: EGF binding to EGFR is relatively stable and thus promotes continuous receptor ubiquitination and lysosomal trafficking, while TGF-α dissociates rapidly leading to receptor deubiquitination and recycling.[60-61] In addition to these well-known roles of EGF and TGF-α ligands, HB-EGF and BTC were reported to target EGFR for lysosomal degradation, while AR and EPR were associated with receptor recycling.[62]

Overexpression of EGFR ligands with concurrent expression of EGFR has been observed in several cancer types.[63-64] Ligand overexpression can activate EGFR persistently through autocrine loops. From the different ligands, the relevance of TGF-α in cancer is best

characterized.[65] Compared with other carcinomas with normal level of ligands, TGF-α overexpression is often associated with a poor prognosis.

In addition to ligand overexpression, altered ligand trafficking can also promote persistent receptor activation. EGFR ligands are normally trafficked to the basolateral domain of polarized epithelial cells.[66] Altered trafficking to the apical domain results in prolonged EGFR phosphorylation resulting in increased proliferation and tumour invasiveness.[67]

1.2.2 EGFR inhibitors

Current FDA-approved EGFR inhibitors include six small-molecule TKIs and two antibodies (**Table 1.1**).[43] Small-molecule EGFR inhibitors comprise three different generations (**Fig. 1.5**). First generation EGFR inhibitors include Gefitinib, Erlotinib and Lapatinib, which are based on the same ATP antagonist: 4-anilinoquinazoline. Gefitinib and Erlotinib bind preferentially to EGFR in an active conformation (Type I TKI) and do not present a high affinity for other members of the ErbB family.[68] Gefitinib and Erlotinib were originally approved for the treatment of NSCL in general. Later studies showed that patients harbouring exon 19 mutations or exon L858R were more sensitive to these inhibitors.[57-59] This sensitivity has been attributed to a higher inhibitor affinity and lower ATP affinity of the mutant forms compared with WT EGFR,[69-70] as well as oncogenic addiction to the mutant forms.[71] Unlike Gefitinib and Erlotinib, an extra aromatic ring in the aniline core of Lapatinib allows a preferential binding to the DFG-out inactive conformation of both EGFR and HER2 (Type II TKI).[72] Thus, Lapatinib is approved for the treatment of BC overexpressing HER2. Afatinib and Dacomitinib are second generation EGFR inhibitors with FDA approval.[73] Like the first generation inhibitors, they are also based on the 4-anilinoquinazoline scaffold but contain an electrophilic group which allows irreversible inhibition by forming a covalent bond through C797.[74] This allows greater efficacy with lower therapeutic concentrations. Other second generation inhibitors are currently under clinical trials.[75]

Table 1.1 FDA-approved EGFR inhibitors

Name	Type of molecule	FDA approval
Gefitinib	Small molecule, type I TKI	2003, NSCLC 2015, EGFR mutant NSCLC
Erlotinib	Small molecule, type I TKI	2004, NSCLC 2005, PC 2013 EGFR mutant NSCLC
Lapatinib	Small molecule, type II TKI	2006, HER2-overexpressing advanced BC 2010, hormone and HER2-positive BC
Afatinib	Small molecule, irreversible TKI	2013, EGFR mutant metastatic NSCLC 2018, EGFR mutant NSCLC
Dacomitinib	Small molecule, irreversible TKI	2018, EGFR mutant NSCLC
Osimertinib	Small molecule, irreversible TKI	2017, T790M EGFR mutant NSCLC
Cetuximab	Human-murine chimeric IgG2	2004, EGFR-positive CRC 2006, advanced HNSCC 2011, metastatic HNSCC 2012, KRAS-WT EGFR-positive CRC
Panitumumab	Human IgG1	2006, EGFR-positive metastatic CRC 2017, KRAS-WT EGFR-positive CRC

About 50% of patients treated with EGFR TKI develop resistance through a secondary T790M mutation.[76-77] While first and second generation inhibitors target both WT and mutant-activating forms of EGFR, third generation inhibitors are designed to bind exclusively to mutant resistant T790M EGFR. Third generation inhibitors are no longer based on the 4-anilinoquinazoline scaffold but on diverse 2-substituted pyrimidines. Similar to the second-generation inhibitors, third-generation inhibitors are irreversible. In 2017, Osimertinib was the first third generation EGFR inhibitor to received FDA approval, and other inhibitors are currently under clinical trials.[78-79]
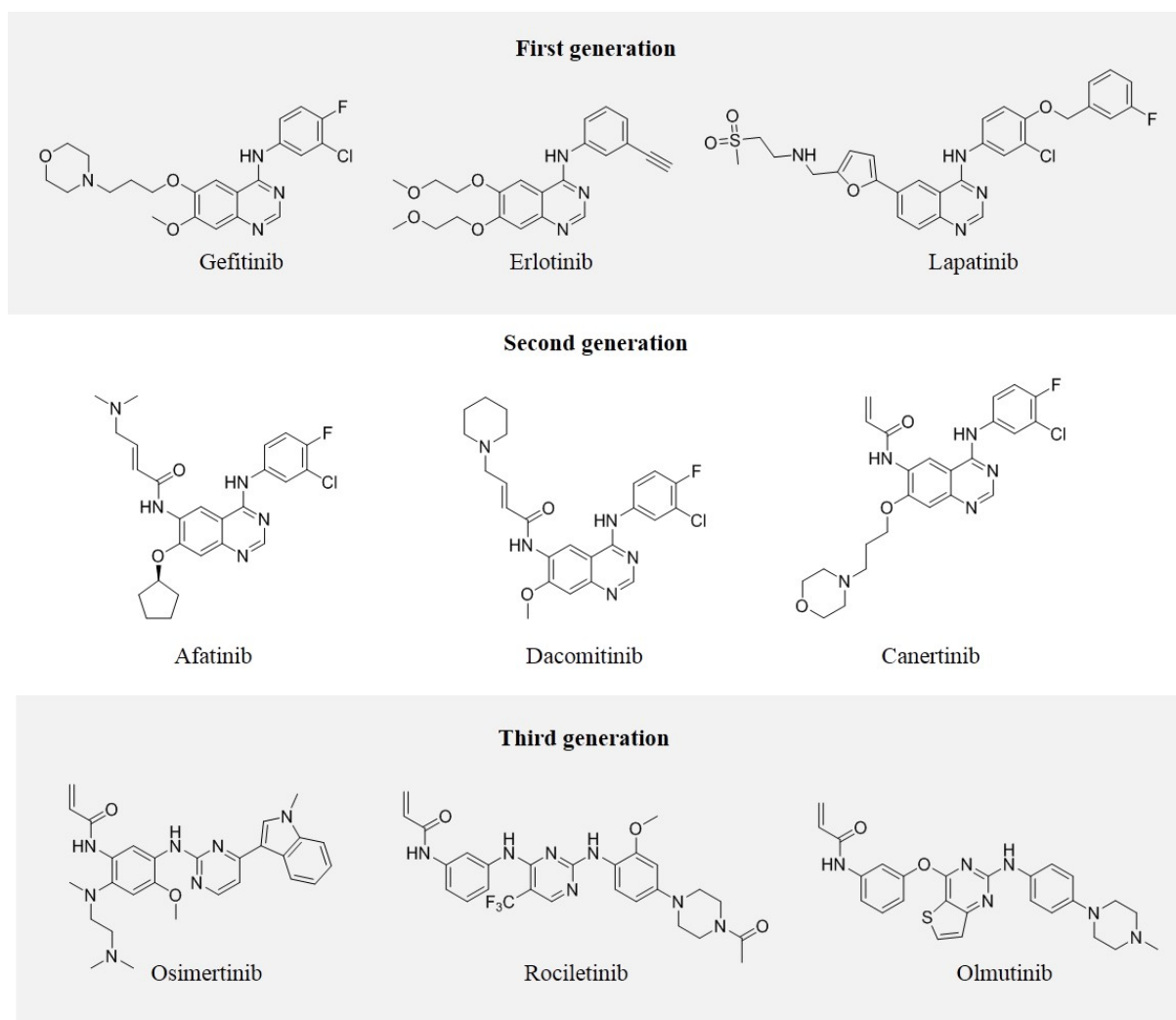
Fig. 1.5 Structure of some EGFR small-molecule TKI

While small-molecule TKI are used almost exclusively in the treatment of NSCLC, EGFR-targeting antibodies have been found more effective in the treatment of CRC, which is dependent on ligand-activated WT EGFR. Currently only two antibodies, Cetuximab and Panitumumab, have received FDA-approval. Just as the efficacy of small-molecule TKI is limited to specific mutant forms of EGFR in NSCLC, the efficacy of these antibodies in CRC is limited to the presence of WT KRAS.[80-82] Since EGFR inhibition leads mainly to a decrease in ERK signalling, KRAS mutations result in intrinsic activation of this signalling pathway and primary resistance to EGFR inhibitors.

1.2.3 Mechanisms of drug resistance to EGFR-targeted therapies

Despite the initial success in solid tumours, EGFR inhibitors eventually become resistant to targeted therapies. Several mechanisms of drug resistance have been reported,[83-85] especially in the context of NSCLC,[86-87] CRC[88-89] and HNSCC[90-91]. Drug resistance mechanisms can be primary (intrinsic) or secondary (acquired). Both mechanisms can originate from changes in EGFR (e.g. mutations, PTMs, altered localization) or from EGFR-independent activation of signalling pathways (bypass mechanisms). Some of these mechanisms have been elucidated after identifying subpopulations sensitive to mono or combinatorial therapies. The major mechanisms of resistance to EGFR-targeted therapies are described next.

1.2.3.1 EGFR mutations

One of the most common mechanisms of drug resistance is the T790M mutation at the gatekeeper position.[76-77] This accounts for 50% of all the cases of resistance in NSCLC. This secondary mutation restores the affinity of EGFR for ATP, thus reversing the sensitivity of the primary activating mutation towards Gefitinib and Erlotinib.[92-93] Third generation inhibitors target this T790M resistant mutant form while sparing WT EGFR. However, resistance to third-generation inhibitors was reported even before the approval of Osimertinib.[94-95] Resistance arises from mutations including C797S (which abrogates irreversible inhibition) and L718Q. Fourth generation EGFR allosteric inhibitors that avoid irreversible ATP-competitive inhibition of T790M EGFR have been reported recently as a way to overcome resistance.[96] S492R mutation in the extracellular domain has also been reported as a mechanisms of drug resistance to Cetuximab.[97]

1.2.3.2 Bypass mechanisms

It is generally accepted that the therapeutic effect of EGFR inhibitors occurs through the simultaneous suppression of several of its downstream signalling pathways (e.g. PI3K/Akt, Ras/Raf/MEK/ERK) that ultimately result in proliferation arrest and apoptosis. If any of these

signalling pathways is activated by mechanisms independent of EGFR, EGFR inhibition is compensated and thus resistance develops. This is known as "bypass track" resistance.[98] Several conditions might lead to bypass resistance, including activation of signalling pathways by other RTKs and constitutive activation due to intrinsic mutations or downregulation of regulatory proteins.

1.2.3.2.1 Activation by other RTKs

One of the first mechanisms of bypass resistance was identified in some cases of NSCLC with amplification of the hepatocyte growth factor receptor (MET).[99-100] MET amplification resulted in the EGFR-independent activation of the PI3K/Akt pathway through its association with HER3. Similarly, later studies demonstrated that MET activation by its ligand HGF also results in resistance to EGFR inhibitors through persistent activation of the PI3K/Akt pathway.[101] However, unlike MET amplification, activation of this pathway was independent of HER3 and mediated by the adaptor protein GAB1.[102] In both cases, the combination of EGFR and MET inhibitors was effective in supressing downstream signalling. Since then, several other RTKs such as AXL,[103-104] IGF-1,[105-106] ErbB2[107] and ErbB3[108-109] have been reported as mediators of resistance to EGFR inhibitors. In most of these cases the activity of RTK does not involve genetic mutations.

1.2.3.2.2 Constitutive activation of signalling pathways

Aberrant activation of signalling pathways downstream of EGFR, mainly due to genetic alterations, may also result in bypass resistance to EGFR inhibitors. Alterations in both the PI3K/Akt and Ras/Raf/MEK/ERK pathways have been reported as mechanisms of drug resistance. Major alterations in the PI3K/Akt pathway include PI3K mutations[110] and loss of PTEN.[111] For the Ras/Raf/MEK/ERK pathway, mutations in KRAS,[112] BRAF[113] as well as aberrant ERK activation[114] by amplification of MAPK1 or downregulation of negative regulators have been identified as causing resistance. Presence of mutant KRAS in CRC is an

indicator of intrinsic resistance to Cetuximab.[112] In all above cases, inhibition of both EGFR and the persistently activated pathway (PI3K/Akt or ERK) is necessary to stop tumour progression.

1.2.3.3 Altered cellular localization

Although primarily localized to the plasma membrane, EGFR has been localized in other cellular compartments such the endosome, the mitochondria and the nucleus.[115] Many cellular processes have been associated with EGFR resistance, including intrinsic defects in apoptosis[116], histological transformation,[117-118] stress-induced EGFR trafficking,[119] as well as EGFR roles in autophagy and metabolism.[120] How altered EGFR localisation contributes to these defects is being intensely investigated, and some of the EGFR localisation effects are described below.

1.2.3.3.1 Nuclear EGFR

Nuclear localization of EGFR has been associated with resistance to EGFR inhibitors and poor prognosis.[121-123] In the nucleus, EGFR can act as a co-transcriptional factor of oncogenic genes or promote DNA replication and repair through phosphorylation and association with effector molecules.

One of the first associations of nuclear EGFR with resistance to targeted therapies was reported by Li et al in 2009.[124] In this work, they identified that NSCLC cells resistant to Cetuximab had increased levels of nuclear EGFR and Src family kinase (SFK) activity. SFK was proven to mediate nuclear trafficking of EGFR, as treatment with the pan SKF inhibitor Dasatinib reduced the levels of EGFR in the nucleus and restored sensitivity to Cetuximab. Following this, overexpression of specific SFK proteins Yes and Lyn was identified in Cetuximab-resistant cells.[125] In this case, Yes and Lyn promote EGFR nuclear localization by direct phosphorylation of Y1101, which is necessary but not sufficient for nuclear transport.

Interestingly, current EGFR inhibitors do not have an effect on pY1101 level in EGFR.[122] Later, it was demonstrated that Yes and Lyn expression is mediated by AXL, suggesting that AXL inhibition might block EGFR nuclear trafficking and resensitize the cells to Cetuximab.[126]

In addition to SFK phosphorylation at Y1101, EGFR nuclear localization has also been reported to be mediated by Akt phosphorylation of EGFR at S229, which promotes resistance to Gefitinib,[127] and PKC phosphorylation at Y654, which promotes resistance to radiotherapy.[128]

1.2.3.3.2   Mitochondrial EGFR

Compared to the role of EGFR in the nucleus, the functions of mitochondrial EGFR are less clear. Several stimulus, including treatment with EGF[129], apoptotic inducers or Gefitinib,[130] have been reported to induce EGFR translocation to the mitochondria. Mitochondrial EGFR interacts with COXII, which is dependent on but not mediated by pY845 in EGFR (in turn dependent on phosphorylation by c-Src).[129] In the mitochondria, COXII is phosphorylated by both EGFR and c-Src, which results in a decrease in Cox activity and ATP levels and thus prevents apoptosis.

1.2.4 Role of EGFR in specific cancer types

1.2.4.1 Triple Negative Breast Cancer

Breast cancer (BC) is the most frequently diagnosed cancer and a leading cause of death in women.[131] Certain markers such as estrogen receptor (ER), progesterone receptor (PR) and HER2 are commonly used to classify BC in the clinic, helping to stratify patients to specific treatments. According to the expression of these markers, BC is classified as hormone receptor (ER and PR) positive, HER2 amplified or triple negative (TNBC) if none of the three markers are expressed. Although TNBC represents only 10-20% of all breast cancer cases, it

is the most aggressive type and presents the poorest clinical outcome.[132] TNBC itself comprises a series of highly heterogeneous subtypes with varied molecular characteristics, requiring a particular therapeutic approach.[133] These factors combined with a lack of validated targeted therapies make TNBC treatment challenging.

Compared to hormone receptor positive and HER2 amplified types, EGFR is more frequently overexpressed in TNBC, and EGFR expression is correlated with a poor prognosis.[134-135] In virtually all TNBC aberrant EGFR signaling originates from protein overexpression rather than activating mutations. *EGFR* gene amplification is observed in only a small fraction of TNBC, suggesting that other mechanism are involved in EGFR dysregulation.[135]

EGFR is one of several proteins currently under investigation as a potential target for TNBC.[136-138] Unlike the success of EGFR inhibitors in NSCLC and CRC, clinical trials of EGFR-targeted therapies in TNBC have been disappointing.[139] This lack of benefit suggests that other players besides EGFR are involved in maintaining the malignant phenotype. Thus, combinatorial therapies targeting EGFR and related supporting molecules may help overcome intrinsic resistance to EGFR inhibitors in TNBC. Several examples of such combinations are reported in the literature, and commonly include targeting other RTKs or components of the signaling pathways activated by EGFR (**Table 1.2**, entries 1-7).

In addition to these typical combinations, other approaches based on non-canonical EGFR functions (e.g. subcellular localization, metabolic control) which had been previously underexplored are gaining more attention (**Table 1.2**, entries 8-12). For example, some types of TNBC present high levels of nuclear EGFR, which is related with poor prognosis and drug resistance. Targeting proteins that promote EGFR nuclear translocation or nuclear EGFR effector proteins sensitize cells to EGFR inhibition.[140-141] Similarly, EGFR localization in lipid rafts promotes EGFR kinase-independent Akt activation, and inhibitors that promote

18

EGFR dissociation from lipids rafts or impede EGFR interactions in lipid rafts sensitize cells to EGFR inhibition.[142-143] In addition to aberrant cell signaling, cancerous cells also rely on specific metabolic characteristics. EGFR has been implicated in the regulation of these metabolic characteristics, and the combination of metabolic and EGFR inhibitors has shown to have a synergistic effect in TNBC.[144-145] Other molecules tested in combination with EGFR inhibitors include regulators of apoptosis[146] and autophagy.[147]

Table 1.2 EGFR combination therapies reported in TNBC

Entry	Target	Mechanism of action	Reference
1	HER3	RTK bypass activation	[148]
2	MET	RTK bypass activation	[149-150]
3	AXL	RTK bypass activation	[104]
4	PI3K	EGFR downstream signaling pathway	[151]
5	mTOR	EGFR downstream signaling pathway	[152-153]
6	MEK	EGFR downstream signaling pathway	[154]
7	PYK2/FAK	EGFR downstream signaling pathway	[155-156]
8	SFK	Subcellular EGFR (nuclear, lipid rafts)	[140, 143]
9	PCNA	Subcellular EGFR (nuclear)	[141]
10	NOTCH3	Subcellular EGFR (lipid rafts)	[142]
11	FASN	Metabolism regulation	[157]
12	MCL-1	Apoptosis regulation	[146]

1.2.4.2 Lung Cancer

Lung cancer is the leading cause of cancer-related death worldwide.[158] Among the different histological types of lung cancer, non-small cell lung cancer (NSCLC) is the most common, with approximately 75% of all cancer types.[159] Aberrant EGFR signaling is commonly observed in NSCLC and associated with pathogenesis and prognosis. Common EGFR abnormalities include EGFR overexpression, mutations and gene amplification.[160] Among these mechanisms, EGFR mutations have resulted more useful in EGFR-targeted treatments,

since mutant-activating EGFR is more sensitive to small-molecule EGFR inhibitors.[57-59] EGFR is currently a validated drug target for NSCLC. Although secondary EGFR mutations are a very common cause of acquired resistance to EGFR inhibitors, several other mechanism have also been reported, and multiple combinatorial therapies are currently undergoing clinical trials.[161]

1.2.4.3 Colorectal Cancer

Colorectal cancer (CRC) represents the third most common and second deadliest cancer among men and women.[158] Although EGFR mutations in CRC are rare, EGFR is estimated to be overexpressed in 60% to 80% of all CRC cases, and EGFR is currently a validated drug target for CRC.[162] The EGFR-targeted antibody cetuximab is more effective in CRC that do not harbor *KRAS* mutations[163] and in those with high expression of amphiregulin and epiregulin,[164] suggesting that ligand-dependent activation of EGFR is a leading cause of oncogenesis in CRC.

1.3   CHEMICAL BIOLOGY APPROACHES FOR STUDYING EGFR SIGNALING NETWORKS

1.3.1 Chemical probes

Chemical probes are small molecules designed to target and perturb a biomolecule, usually a protein, in order to study the properties and functions of such biomolecule in a particular system. This chemical approach is complementary to biological techniques such as RNAi or CRISPR, allowing modulation of certain protein functions (e.g. enzymatic activity) without altering protein levels or completely abrogating the presence of a protein.[165-167] Chemical probes have been extensively used in drug discovery and target validation, elucidation of signaling pathways and bioimaging, which have allowed a better understanding of several biological areas.[167-168] Several articles have recently addressed the importance of high quality

chemical probes for validating biological mechanisms with confidence.[169-171] In this project, EGFR-directed chemical probes were used to identify potential EGFR interactors in breast cancer cells.
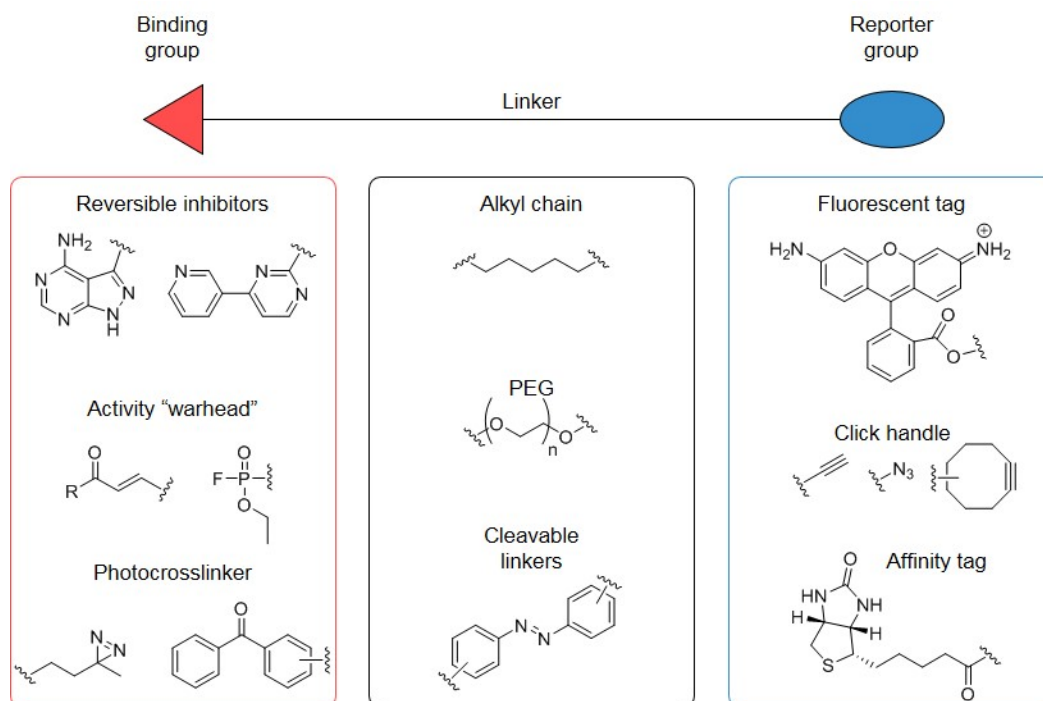
Fig. 1.6 General structure of a chemical probe and some typical components

The structure of a chemical probe generally comprises three different elements: a binding group, a reporter tag and a linker (**Fig 1.6**).[172] The binding group mediates the probe-target interaction. If this interaction requires the participation of a catalytic residue, the probe is called "activity based" (ABP); otherwise, it is an affinity based probe (A*f*BP). A covalent A*f*BP allows the irreversible labelling of a target of interest by incorporating a reactive group, usually an electrophile or photocrosslinker. Once the probe-target interaction takes place, the reporter tag allows further manipulation of the protein of interest. Common reporter tags include fluorophores for visualization, biotin for affinity purification and bioorthogonal handles for click chemistry.[173] The linker provides enough space between the target and the reporter to minimize disruptions of probe-target interaction. The linker usually consist of an
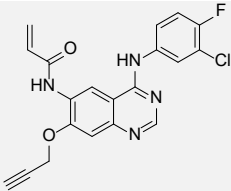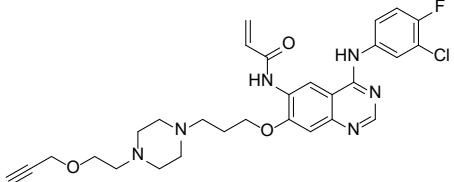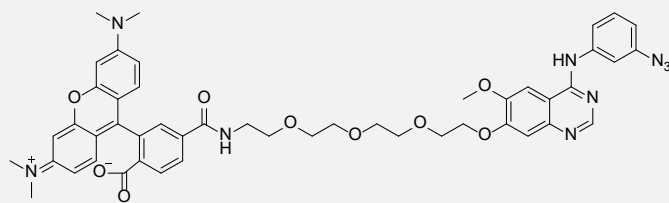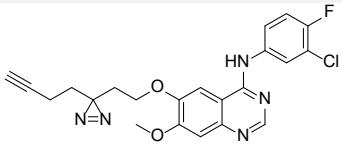
alkyl or polyethylene glycol (PEG) chain, and it may also include some cleavable groups to facilitate dissociation from the targeted protein after affinity purification.[172, 174]

1.3.1.1 EGFR-targeting chemical probes

Several A*f*BP targeting EGFR have been reported in the literature (**Table 1.3**). These probes are usually obtained by modification of 4-anilinoquinazoline drugs at positions C6 and C7 in the quinazoline ring without significant reduction of binding affinity. Depending on their interaction with EGFR, probes are classified as reversible or covalent.

Table 1.3 Summary of selected EGFR probes reported in the literature.

Entry	Structure	Type	Ref.
1		Reversible, kinobeads	175-180
2		Reversible, biotin	181
3		Reversible, biotin	182
4		Covalent (cysteine), biotin	183
5		Covalent (cysteine), fluorophore	184
6		Covalent (cysteine), fluorophore	185-186

7		Covalent (cysteine), click handle	[187-188]
8		Covalent (cysteine), click handle	[183]
9		Covalent (photocrosslinker), fluorophore	[189]
10		Covalent (photocrosslinker), click handle	[190]

Reversible probes are often used after immobilization into "kinobeads" in chemical proteomics analysis. Immobilization can be done reacting probes with a terminal amino group on beads bearing epoxy groups, or incubating biotinylated probes on streptavidin beads. Immobilized EGFR probes have been used in the identification of cellular targets of Gefitinib[175] and Erlotinib[191] (inhibitor selectivity), differentiation in the mode of action of these two drugs,[178] and quantitative analysis of perturbations to the (phosphor)proteome in response to inhibitor treatment.[176-177] By their nature, immobilized reversible probes are limited to *in vitro* studies with cell lysates.

Unlike reversible EGFR A*f*BP, covalent probes are rarely immobilized on beads, allowing their use for studies *in situ* and *in vivo*. When using reversible probes, probe-target interactions might be disrupted if harsh conditions are use. Covalent probes overcome this problem through the irreversible labelling of the interacting protein. Most covalent EGFR probes are based on irreversible inhibitors and make use of an acrylamide derivative to label

Cys-797 in EGFR. The highly reactive nature of this electrophile defines a limited window of selectivity, after which indiscriminate labelling of other cysteine-containing proteins is observed.[187] A few probes containing a photocrosslinker as the reactive group have also been reported.[189-190]

1.3.1.2 Lysine-targeting chemical probes

Compared with cysteine-targeting probes, the covalent modification of lysine residues by chemical probes has not been extensively explored. This might be in part due to poorer nucleophilic character of lysine under physiological conditions compared with cysteine. However, molecules containing two terminal activated esters (most common NHS) are commercially available and used as crosslinkers reagents for studying protein complexes in combination with mass spectrometry.[192-193] These NHS crosslinkers have been used in studies of EGFR dimer formation.[194-196]

The interest in the covalent modification of lysines by covalent inhibitor or chemical probes has been increasing recently.[197-200] Several groups have reported the global proteome reactivity of activated esters towards lysine modification.[201-204] The modification of known drugs with activated esters have shown that is possible to label a specific lysine residue with high selectivity (**Fig. 1.7**). For example, Kelly et al reported a covalent inhibitor for transthyretin capable of reacting chemoselectively with one of eight lysine residues present in the protein.[205] This selectivity was achieved by introducing a fluorophenyl ester in a position of the ligand that favoured interaction with the target lysine (proximity effect) based in crystal structures. The labelling kinetics was also fine-tuned by using different substituents in the phenyl ring. Similarly Campos et al reported the modification of a PI3K inhibitor with a fluorophenyl ester, achieving selective labelling of the protein of interest with minimal off-target labelling.[198] The Hamachi group has also reported several probes based on tosylates or phenyl esters showing high selectivity for the target of interest in living cells (**Fig. 1.7**).[206-208]

Using this methodology, they were able to identify protein disulphide isomerase (PDI) as an off-target of Lapatinib in the gastric cancer cell line NCI-N87.[207] Except for this Lapatinib-derived probe, lysine-reactive EGFR probes have not been well explored, and represent a novel research opportunity.
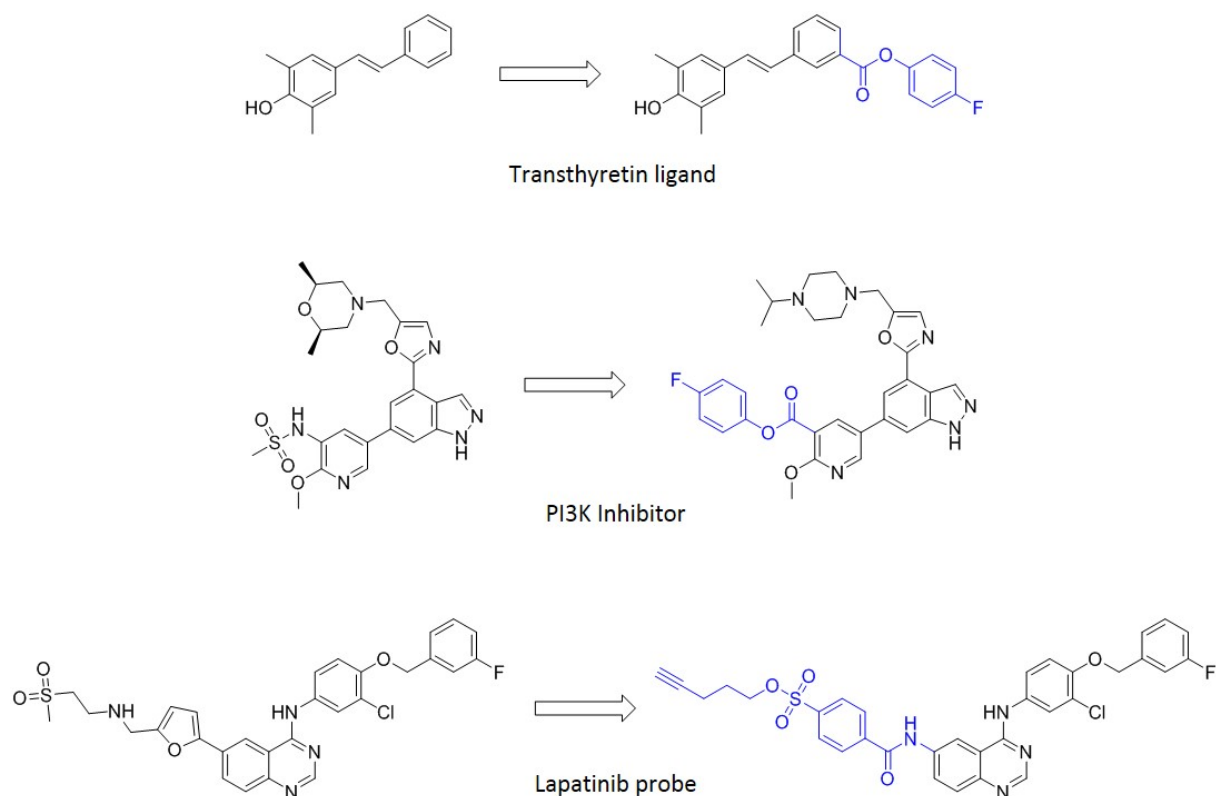
Fig. 1.7 Modification of drugs into lysine-targeting drugs or probes

1.3.2 Chemical proteomics: forward and reverse

Gene function is diversified at the protein level through numerous functions originating from molecular interactions specific to a spatiotemporal context and often mediated by PTMs. Therefore, proteomics studies are indispensable and complementary to genomic studies in the elucidation of molecular mechanisms of cancer.[209]

With its capacity to elucidate information on PTMs and PPI, global proteomics analysis have shed light on EGFR signaling networks[210-211] and this topic has been extensively reviewed.[28] Dynamic information on EGFR PTMs and corresponding interactome,[212-213] linking of

phenotypic characteristics to molecular events and therapeutic manipulation of such events,[214]elucidation of mechanisms of drug resistance[215] and identification of therapeutical vulnerabilities[216-218] as well as markers of therapeutic response[219] have been possible with global proteomics analysis.

Chemical proteomics, or forward chemical proteomics,[220] is an approach of analysis that further expands the global proteomic potential in order to provide information on small-molecule/protein interactions. This is achieved through the combined use of chemical probes, which target proteins of interest, with proteomics analysis. Depending on the nature of the probe, chemical proteomics analysis is usually classified in two major groups: Activity-Based Proteome Profiling (ABPP, for ABP), which focuses on the enzymatic activity of a particular protein family, and Compound Centric Chemical proteomics (CCCP, for A*f*BP), which focuses on characterizing the molecular mechanism of action of bioactive small molecules.[221-223] Chemical proteomics has been extensively used in small-molecule drug profiling and discovery.

For EGFR, most chemical proteomics studies are based on CCCP using kinobeads. This strategy has been used to profile EGFR inhibitors off-targets, allowing the elucidation and expansion of drug mechanisms of action.[175, 177-178, 191] For example, despite their similar structure, only Erlotinib but not Gefitinib is approved for the treatment of pancreatic cancer. Since in this case EGFR expression does not correlate with therapeutic response, the mechanism of action is thought to be mediated by an off-target effect. Using chemical proteomics, Conradt et al identified six kinases binding to Erlotinib with higher affinity than EGFR, suggesting that their inhibition may be responsible for the results observed.[191] Similarly, both Gefitinib and Erlotinib have antineoplastic effects on acute myeloid leukaemia (AML) despite the absence of measurable EGFR expression, and chemical proteomics has identified SFKs and Btk as direct drug interactors and possible mediators of drug

sensitivity.[177] Chemical proteomics has also allowed the elucidation of drug resistance mechanisms. Using a combination of kinobeads and label-free quantitative proteomics, Koch et al identified the upregulation of ephrin receptor A2 (EPHA2) as a novel mediator of bypass acquired resistance to Gefitinib in HCC127 NSCLC cell;[224] EPHA2 inhibition resensitized cells to EGFR inhibition.

1.3.2.1 Reverse chemical proteomics

Despite its many strengths, chemical proteomics also presents technical limitations. Abundant proteins and "sticky" proteins prone to non-specific binding can create a high background, concealing low abundant targets. Proteins with low stability or solubility in cell extracts are also difficult to identify. Reverse chemical proteomics is an alternative approach to traditional or forward chemical proteomics that can help overcome these limitations.[174, 225] Rather than starting from the endogenous cell proteome, reverse chemical proteomics makes use of the cell transcriptome and expression cloning techniques that provide a physical link between each protein and its encoding gene. Although the small-molecule/protein interaction takes place directly, target identification is done indirectly at the gene rather than protein level. This gene-protein link is used to amplify any protein interacting with the probe in an iterative process, facilitating the enrichment and identification of avid but low abundant targets. Common expression cloning techniques that have been used in this approach include three-hybrid systems[226] and display technologies.[227-228] A detailed description of reverse chemical proteomics based on phage display techniques and its value as a complementary approach to forward chemical proteomics can be found in section 3.1.1.

1.4 PROJECT OUTLINE

Numerous studies on the prototypical RTK EGFR have shed light on our current understanding of kinase cell signaling, with important implications in cancer targeted

therapies. This is reflected in the eight EGFR inhibitors currently in clinical use and extensive clinical trials trying to validate this target beyond NSCLC and CRC. Intrinsic or acquired drug resistance is a complex problem originating from heterogeneous molecular mechanisms that keeps impeding the long term success of EGFR-targeted therapies. Among these resistance mechanisms, secondary EGFR mutations have been most extensively studied, reflected in four generations of drugs. Better understanding of EGFR signaling networks robustness and kinome reprogramming, made possible in part by chemical proteomics analysis, have also allowed us to identify several potential combinatorial therapies for restoring sensitivity to enzymatic inhibition.

With increasing evidence recently linking non-canonical kinase independent functions of EGFR with drug resistance, new design principles can be explored to expand the current therapeutic paradigm. Elucidation of the molecular mechanisms behind these emerging functions is necessary to develop more efficacious therapies and requires a multidirectional approach. Chemical proteomics can provide information on direct physical interactions of EGFR and it is therefore an important tool in this quest.

Supporting these efforts, we have developed novel EGFR-directed chemical probes and tested their performance in forward and reverse chemical proteomics analysis. Chapter two describes the design and synthesis of covalent, lysine-targeting EGFR probes based on first and second generation EGFR inhibitors with potential to improve labelling selectivity and capture endogenous EGFR interactors. Chapter three reports the identification of EGFR drug targets potentially associated with drug resistance using reverse chemical proteomics with non-covalent affinity probes. Finally, chapter four describes the identification of potential EGFR interactors and enrichment of EGFR in subcellular compartments with various chemical probes in MDA-MB-468 TNBC cells using forward chemical proteomics analysis. Chapter five summarises these findings and discusses future directions from this work.

1.5   REFERENCES

1.   Roskoski, R., Jr., The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacol. Res.* **2014,** *79*, 34-74.
2.   Roskoski, R., Jr., ErbB/HER protein-tyrosine kinases: Structures and small molecule inhibitors. *Pharmacol. Res.* **2014,** *87*, 42-59.
3.   Harris, R. C.; Chung, E.; Coffey, R. J., EGF receptor ligands. *Exp. Cell. Res.* **2003,** *284* (1), 2-13.
4.   Ullrich, A.; Schlessinger, J., Signal transduction by receptors with tyrosine kinase activity. *Cell* **1990,** *61*, 203-212.
5.   Schlessinger, J., Cell signaling by receptor tyrosine kinases. *Cell* **2000,** *103*, 211-225.
6.   Lemmon, M. A.; Schlessinger, J., Cell signaling by receptor tyrosine kinases. *Cell* **2010,** *141* (7), 1117-1134.
7.   Schlessinger, J., Receptor tyrosine kinases: legacy of the first two decades. *Cold Spring Harbor Perspect. Biol.* **2014,** *6* (3), a008912.
8.   Endres, N. F.; Barros, T.; Cantor, A. J.; Kuriyan, J., Emerging concepts in the regulation of the EGF receptor and other receptor tyrosine kinases. *Trends Biochem. Sci.* **2014,** *39* (10), 437-446.
9.   Lemmon, M. A.; Schlessinger, J.; Ferguson, K. M., The EGFR family: not so prototypical receptor tyrosine kinases. *Cold Spring Harbor Perspect. Biol.* **2014,** *6* (4), a020768.
10.   Bessman, N. J.; Freed, D. M.; Lemmon, M. A., Putting together structures of epidermal growth factor receptors. *Curr. Opin. Struct. Biol.* **2014,** *29*, 95-101.
11.   Kovacs, E.; Zorn, J. A.; Huang, Y.; Barros, T.; Kuriyan, J., A structural perspective on the regulation of the epidermal growth factor receptor. *Annu. Rev. Biochem.* **2015,** *84*, 739-764.
12.   Ferguson, K. M.; Berger, M. B.; Mendrola, J. M.; Cho, H.-S.; Leahy, D. J.; Lemmon, M. A., EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol. Cel.* **2003,** *11*, 507-517.
13.   Burgess, A. W.; Cho, H.-S.; Eigenbrot, C.; Ferguson, K. M.; Garrett, T. P. J.; Leahy, D. J., . . . Yokoyama, S., An open-and-shut case? Recent insights into the activation of EGF/ErbB Receptors. *Mol. Cel.* **2003,** *12*, 541-552.
14.   Zhang, X.; Gureasko, J.; Shen, K.; Cole, P. A.; Kuriyan, J., An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* **2006,** *125* (6), 1137-1149.
15.   Endres, N. F.; Das, R.; Smith, A. W.; Arkhipov, A.; Kovacs, E.; Huang, Y., . . . Kuriyan, J., Conformational coupling across the plasma membrane in activation of the EGF receptor. *Cell* **2013,** *152* (3), 543-556.
16.   Arkhipov, A.; Shan, Y.; Das, R.; Endres, N. F.; Eastwood, M. P.; Wemmer, D. E., . . . Shaw, D. E., Architecture and membrane interactions of the EGF receptor. *Cell* **2013,** *152* (3), 557-569.
17.   Lemmon, M. A., Ligand-induced ErbB receptor dimerization. *Exp. Cell Res.* **2009,** *315* (4), 638-648.
18.   Matsushita, C.; Tamagaki, H.; Miyazawa, Y.; Aimoto, S.; Smith, S. O.; Sato, T., Transmembrane helix orientation influences membrane binding of the intracellular juxtamembrane domain in Neu receptor peptides. *Proc. Natl. Acad. Sci. USA* **2012,** *110* (5), 1646-1651.
19.   Citri, A.; Skaria, K. B.; Yarden, Y., The deaf and the dumb: the biology of ErbB-2 and ErbB-3. *Exp. Cell. Res.* **2003,** *284* (1), 54-65.
20.   Schlessinger, J.; Lemmon, M. A., SH2 and PTB Domains in Tyrosine Kinase Signaling. *Sci. Signaling* **2003,** *2003* (191), Re12.
21.   Pawson, T., Specificity in Signal Transduction: From Phosphotyrosine-SH2 Domain Interactions to Complex Cellular Systems. *Cell* **2004,** *116*, 191-203.
22.   Yarden, Y.; Sliwkowski, M. X., Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell. Biol.* **2001,** *2*, 127-137.
23.   Citri, A.; Yarden, Y., EGF-ERBB signalling: towards the systems level. *Nat. Rev. Mol. Cell. Biol.* **2006,** *7* (7), 505-516.
24.   Wee, P.; Wang, Z., Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. *Cancers* **2017,** *9*, 52.
25.   Barberan, S.; Cebria, F., The role of the EGFR signaling pathway in stem cell differentiation during planarian regeneration and homeostasis. *Semin. Cell Dev. Biol.* **2019,** *87*, 45-57.
26.   Ali, R.; Brown, W.; Purdy, S. C.; Davisson, V. J.; Wendt, M. K., Biased signaling downstream of epidermal growth factor receptor regulates proliferative versus apoptotic response to ligand. *Cell Death Dis.* **2018,** *9*, 976.
27.   Lu, Z.; Jiang, G.; Blume-Jensen, P.; Hunter, T., Epidermal Growth Factor-Induced Tumor Cell Invasion and Metastasis Initiated by Dephosphorylation and Downregulation of Focal Adhesion Kinase. *Mol. Cell Biol.* **2001,** *21* (12), 4016-4031.
28.   Morandell, S.; Stasyk, T.; Skvortsov, S.; Ascher, S.; Huber, L. A., Quantitative proteomics and phosphoproteomics reveal novel insights into complexity and dynamics of the EGFR signaling network. *Proteomics* **2008,** *8* (21), 4383-4401.

29. Wilson, K. J.; Gilmore, J. L.; Foley, J.; Lemmon, M. A.; Riese, D. J., 2nd, Functional selectivity of EGF family peptide growth factors: implications for cancer. *Pharmacol. & Ther.* **2009,** *122* (1), 1-8.

30. Olayioye, M.; Neve, R. M.; Lane, H. A.; Hynes, N. E., The ErbB signaling network: receptor heterodimerization in development and cancer. *EMBO J.* **2000,** *19* (13), 3159-3167.

31. Deribe, Y. L.; Pawson, T.; Dikic, I., Post-translational modifications in signal integration. *Nat. Struct. & Mol. Biol.* **2010,** *17* (6), 666-672.

32. Kholodenko, B. N., Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.* **2006,** *7* (3), 165-176.

33. Kholodenko, B. N.; Hancock, J. F.; Kolch, W., Signalling ballet in space and time. *Nat. Rev. Mol. Cell Biol.* **2010,** *11* (6), 414-426.

34. Mason, J. M.; Morrison, D. J.; Basson, M. A.; Licht, J. D., Sprouty proteins: multifaceted negative-feedback regulators of receptor tyrosine kinase signaling. *Trends Cell Biol.* **2006,** *16* (1), 45-54.

35. Gur, G.; Rubin, C.; Katz, M.; Amit, I.; Citri, A.; Nilsson, J., . . . Yarden, Y., LRIG1 restricts growth factor signaling by enhancing receptor ubiquitylation and degradation. *EMBO J.* **2004,** *23* (16), 3270-3281.

36. Anastasi, S.; Fiorentino, L.; Fiorini, M.; Fraioli, R.; Sala, G.; Castellani, L., . . . Segatto, O., Feedback inhibition by RALT controls signal output by the ErbB network. *Oncogene* **2003,** *22* (27), 4221-4234.

37. Gotoh, N., Feedback inhibitors of the epidermal growth factor receptor signaling pathways. *Int. J. Biochem. Cell Biol.* **2009,** *41* (3), 511-515.

38. Bakker, J.; Spits, M.; Neefjes, J.; Berlin, I., The EGFR odyssey – from activation to destruction in space and time. *J. Cell. Sci.* **2017,** *130*, 4087-4096.

39. Tomas, A.; Futter, C. E.; Eden, E. R., EGF receptor trafficking: consequences for signaling and cancer. *Trends Cell Biol.* **2014,** *24* (1), 26-34.

40. Goh, L. K.; Sorkin, A., Endocytosis of receptor tyrosine kinases. *Cold Spring Harb. Perspect. Biol.* **2013,** *5* (5), a017459.

41. Sorkin, A.; Goh, L. K., Endocytosis and intracellular trafficking of ErbBs. *Exp. Cell. Res.* **2009,** *315* (4), 683-696.

42. Downward, J.; Yarden, Y.; Mayes, E.; Scrace, G.; Totty, N.; Stockwell, P., . . . Waterfield, M. D., Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences. *Nature* **1984,** *307* (9), 521–527.

43. Arteaga, C. L.; Engelman, J. A., ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell* **2014,** *25* (3), 282-303.

44. Zandi, R.; Larsen, A. B.; Andersen, P.; Stockhausen, M. T.; Poulsen, H. S., Mechanisms for oncogenic activation of the epidermal growth factor receptor. *Cell. Signal.* **2007,** *19* (10), 2013-2023.

45. Nicholson, R. I.; Gee, J. M. W.; Harper, M. E., EGFR and cancer prognosis. *Eur. J. Cancer* **2001,** *37* (Suppl. 4), S9–S15.

46. Bhargava, R.; Gerald, W. L.; Li, A. R.; Pan, Q.; Lal, P.; Ladanyi, M.; Chen, B., EGFR gene amplification in breast cancer: correlation with epidermal growth factor receptor mRNA and protein expression and HER-2 status and absence of EGFR-activating mutations. *Mod. Pathol.* **2005,** *18* (8), 1027-1033.

47. Suzuki, S.; Dobashi, Y.; Sakurai, H.; Nishikawa, K.; Hanawa, M.; Ooi, A., Protein overexpression and gene amplification of epidermal growth factor receptor in nonsmall cell lung carcinomas. *Cancer* **2005,** *103* (6), 1265-1273.

48. Wong, A. J.; Bigner, S. H.; Bigner, D. D.; Kinzler, K. W.; Hamilton, S. R.; Vogelstein, B., Increased expression of the epidermal growth factor receptor gene in malignant gliomas is invariably associated with gene amplification. *Proc. Nat.l Acad. Sci. USA* **1987,** *84*, 6899–6903.

49. Brandt, B.; Meyer-Staeckling, S.; Schmidt, H.; Agelopoulos, K.; Buerger, H., Mechanisms of egfr gene transcription modulation: relationship to cancer risk and therapy response. *Clin. Cancer Res.* **2006,** *12* (24), 7252-7560.

50. Johnston, D.; Hall, H.; DiLorenzo, T. P.; Steinberg, B. M., Elevation of the Epidermal Growth Factor Receptor and Dependent Signaling in Human Papillomavirus-infected Laryngeal Papillomas. *Cancer Res.* **1999,** *59*, 968–974.

51. Chakraborty, S.; Li, L.; Puliyappadamba, V. T.; Guo, G.; Hatanpaa, K. J.; Mickey, B., . . . Habib, A. A., Constitutive and ligand-induced EGFR signalling triggers distinct and mutually exclusive downstream signalling networks. *Nat. Commun.* **2014,** *5*, 5811.

52. Guo, G.; Gong, K.; Wohlfeld, B.; Hatanpaa, K. J.; Zhao, D.; Habib, A. A., Ligand-Independent EGFR Signaling. *Cancer Res.* **2015,** *75* (17), 3436-3441.

53. Huang, P. H.; Xu, A. M.; White, F. M., Oncogenic EGFR Signaling Networks in Glioma. *Sci. Signaling* **2009,** *2* (87), re6.

54. Zhang, Z.; Stiegler, A. L.; Boggon, T. J.; Kobayashi, S.; Halmos, B., EGFR-mutated lung cancer: a paradigm of molecular oncology. *Oncotarget* **2010,** *1* (7), 497–514.

55. Gan, H. K.; Cvrljevic, A. N.; Johns, T. G., The epidermal growth factor receptor variant III (EGFRvIII): where wild things are altered. *FEBS J* **2013,** *280* (21), 5350-5370.

56. Gazdar, A. F., Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene* **2009,** *28 Suppl 1,* S24-31.

57. Lynch, T. J.; Bell, D. W.; Sordella, R.; Gurubhagavatula, S.; Okimoto, R. A.; Brannigan, B. W., . . . Haber, D. A., Activating mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non-Small-Cell Lung Cancer to Gefitinib. *N. Engl. J. Med.* **2004,** *350* (21), 2129-2139.

58. Paez, J. G.; Jane, P. A.; Lee, J. C.; Tracy, S.; Greulich, H.; Gabriel, S., . . . Meyerson, M., EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy. *Science* **2004,** *304,* 1497–1500.

59. Pao, W.; Miller, V.; Zakowski, M.; Doherty, J.; Politi, K.; Sarkaria, I., . . . Varmus, H., EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. USA* **2004,** *101* (36), 13306-13311.

60. Decker, S. J., Epidermal Growth Factor and Tansforming Growth Factor-alpha Induce Differential Processing of the Epidermal Growth Factor Receptor. *Biochem. Biophys. Res. Commun.* **1990,** *166* (2), 615-621.

61. Ebner, R.; Derynck, R., Epidermal growth factor and transforming growth factor-a: differential intracellular routing and processing of ligand-receptor complexes. *Cell Regulation* **1991,** *2,* 599-612.

62. Roepstorff, K.; Grandal, M. V.; Henriksen, L.; Knudsen, S. L. J.; Lerdrup, M.; Grovdal, L., . . . Deurs, B. v., Differential Effects of EGFR Ligands on Endocytic Sorting of the Receptor. *Traffic* **2009,** *10* (1115-1127).

63. Salomon, D. S.; Brandt, R.; Ciardiello, F.; Normanno, N., Epidermal growth factor-related peptides and their receptors in human malignancies. *Crit. Rev. Oncol. Hematol.* **1995,** *19,* 183-232.

64. Tørring, N.; Jørgensen, P. E.; Sørensen, B. S.; Nexø, E., Increased expression of heparin binding EGF (HB-EGF), amphiregulin, TGF alpha and epiregulin in androgen-independent prostate cancer cell lines. *Anticancer Res.* **2000,** *20* (1A), 91-95.

65. Singh, B.; Coffey, R. J., From wavy hair to naked proteins: the role of transforming growth factor alpha in health and disease. *Semin. Cell. Dev. Biol.* **2014,** *28,* 12-21.

66. Singh, B.; Coffey, R. J., Trafficking of epidermal growth factor receptor ligands in polarized epithelial cells. *Annu. Rev. Physiol.* **2014,** *76,* 275-300.

67. Singh, B.; Bogatcheva, G.; Washington, M. K.; Coffey, R. J., Transformation of polarized epithelial cells by apical mistrafficking of epiregulin. *Proc. Natl. Acad. Sci. USA* **2013,** *110* (22), 8960–8965.

68. Johnson, L. N., Protein kinase inhibitors: contributions from structure to clinical compounds. *Q. Rev. Biophys.* **2009,** *42* (1), 1-40.

69. Eck, M. J.; Yun, C. H., Structural and mechanistic underpinnings of the differential drug sensitivity of EGFR mutations in non-small cell lung cancer. *Biochim. Biophys. Acta* **2010,** *1804* (3), 559-566.

70. Yun, C. H.; Boggon, T. J.; Li, Y.; Woo, M. S.; Greulich, H.; Meyerson, M.; Eck, M. J., Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* **2007,** *11* (3), 217-27.

71. Perez, R.; Crombet, T.; de Leon, J.; Moreno, E., A view on EGFR-targeted therapies from the oncogene-addiction perspective. *Frontiers Pharmacol.* **2013,** *4,* 53.

72. Wood, E. R.; Truesdale, A. T.; McDonald, O. B.; Yuan, D.; Hassell, A.; Dickerson, S. H., . . . Shewchuk, L., A Unique Structure for Epidermal Growth Factor Receptor Bound to GW572016 (Lapatinib): Relationships among Protein Conformation, Inhibitor Off-Rate, and Receptor Activity in Tumor Cells. *Cancer Res.* **2004,** *64* (18), 6652-6659.

73. Keating, G. M., Afatinib: a review of its use in the treatment of advanced non-small cell lung cancer. *Drugs* **2014,** *74* (2), 207-221.

74. Nelson, V.; Ziehr, J.; Agulnik, M.; Johnson, M., Afatinib: emerging next-generation tyrosine kinase inhibitor for NSCLC. *OncoTargets Ther.* **2013,** *6,* 135-43.

75. Yu, H. A.; Riely, G. J., Second-generation epidermal growth factor receptor tyrosine kinase inhibitors in lung cancer. *J. Natl. Compr. Canc. Netw.* **2013,** *11.*

76. Pao, W.; Miller, V. A.; Politi, K. A.; Riely, G. J.; Somwar, R.; Zakowski, M. F., . . . Varmus, H., Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS medicine* **2005,** *2* (3), e73.

77. Kobayashi, S.; Boggon, T. J.; Dayaram, T.; Janne, P. A.; Kocher, O.; Meyerson, M., . . . Halmos, B., EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **2005,** *352,* 786-792.

78. Steuer, C. E.; Khuri, F. R.; Ramalingam, S. S., The next generation of epidermal growth factor receptor tyrosine kinase inhibitors in the treatment of lung cancer. *Cancer* **2015,** *121* (8), E1-6.

79. Barnes, T. A.; O'Kane, G. M.; Vincent, M. D.; Leighl, N. B., Third-Generation Tyrosine Kinase inhibitors Targeting Epidermal Growth Factor Receptor Mutations in non-Small Cell Lung Cance. *Front. Oncol.* **2017,** *7,* 113.

80. Lievre, A.; Bachet, J. B.; Le Corre, D.; Boige, V.; Landi, B.; Emile, J. F., . . . Laurent, P., P. , KRAS Mutation Status Is Predictive of Response to Cetuximab Therapy in Colorectal Cancer. *Cancer Res.* **2006,** *66* (8), 3992-3995.

81. Benvenuti, S.; Sartore-Bianchi, A.; Di Nicolantonio, F.; Zanon, C.; Moroni, M.; Veronese, S., . . . Bardelli, A., Oncogenic Activation of the RAS/RAF Signaling Pathway Impairs the Response of Metastatic Colorectal Cancers to Anti– Epidermal Growth Factor Receptor Antibody Therapies. *Cancer Res.* **2007,** *67* (6), 2643-2648.

82. Amado, R. G.; Wolf, M.; Peeters, M.; Van Cutsem, E.; Siena, S.; Freeman, D. J., . . . Chang, D. D., Wild-Type KRAS Is Required for Panitumumab Efficacy in Patients With Metastatic Colorectal Cancer. *J. Clin. Oncol.* **2008,** *26* (10), 1626-1434.

83. Brand, T. M.; Iida, M.; Wheeler, D. L., Molecular mechanisms of resistance to the EGFR monoclonal antibody cetuximab. *Cancer Biol. Ther.* **2011,** *11* (9), 777-792.

84. Yang, S. H., Molecular basis of drug resistance: epidermal growth factor receptor tyrosine kinase inhibitors and anaplastic lymphoma kinase inhibitors. *Tuberc. Respir. Dis.* **2013,** *75* (5), 188-98.

85. Chong, C. R.; Janne, P. A., The quest to overcome resistance to EGFR-targeted therapies in cancer. *Nat. Med.* **2013,** *19* (11), 1389-1400.

86. Huang, L.; Fu, L., Mechanisms of resistance to EGFR tyrosine kinase inhibitors. *Acta Pharm Sin B D***2015,** *5* (5), 390-401.

87. Nakata, A.; Gotoh, N., Recent understanding of the molecular mechanisms for the efficacy and resistance of EGF receptor-specific tyrosine kinase inhibitors in non-small cell lung cancer. *Expert Opin Ther Targets* **2012,** *16* (8), 771-81.

88. Zhao, B.; Wang, L.; Qiu, H.; Zhang, M.; Sun, L.; Peng, P., . . . Yuan, X., Mechanisms of resistance to anti-EGFR therapy in colorectal cancer. *Oncotarget* **2017,** *8* (3), 3980-4000.

89. Bardelli, A.; Siena, S., Molecular Mechanisms of Resistance to Cetuximab and Panitumumab in Colorectal Cancer. *J. Clin. Oncol.* **2010,** *28* (7), 1254-1261.

90. Ratushny, V.; Astsaturov, I.; Burtness, B. A.; Golemis, E. A.; Silverman, J. S., Targeting EGFR resistance networks in head and neck cancer. *Cell. Signal.* **2009,** *21*, 1255-1268.

91. Campbell, N. P.; Hensing, T. A.; Bhayani, M. K.; Shaikh, A. Y.; Brockstein, B. E., Targeting pathways mediating resistance to anti-EGFR therapy in squamous cell carcinoma of the head and neck. *Expert Rev. Anticancer Ther.* **2016,** *16* (8), 847-858.

92. Yun, C. H.; Mengwasser, K. E.; Toms, A. V.; Woo, M. S.; Greulich, H.; Wong, K. K., . . . Eck, M. J., The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci U S A* **2008,** *105* (6), 2070-5.

93. Yoshikawa, S.; Kukimoto-Niino, M.; Parker, L.; Handa, N.; Terada, T.; Fujimoto, T., . . . Yokoyama, S., Structural basis for the altered drug sensitivities of non-small cell lung cancer-associated mutants of human epidermal growth factor receptor. *Oncogene* **2013,** *32* (1), 27-38.

94. Wang, S.; Song, Y.; Yan, F.; Liu, D., Mechanisms of resistance to third-generation EGFR tyrosine kinase inhibitors. *Front. Med.* **2016,** *10* (4), 383-388.

95. Patel, H.; Pawara, R.; Ansari, A.; Surana, S., Recent updates on third generation EGFR inhibitors and emergence of fourth generation EGFR inhibitors to combat C797S resistance. *Eur. J. Med. Chem.* **2017,** *142*, 32-47.

96. Jia, Y.; Yun, C. H.; Park, E.; Ercan, D.; Manuia, M.; Juarez, J., . . . Eck, M. J., Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors. *Nature* **2016,** *534* (7605), 129-132.

97. Montagut, C.; Dalmases, A.; Bellosillo, B.; Crespo, M.; Pairet, S.; Iglesias, M., . . . Albaneil, J., Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer. *Nat. Med.* **2012,** *18* (2), 221-223.

98. Niederst, M. J.; Engelman, J. A., Bypass Mechanisms of Resistance to Receptor Tyrosine Kinase Inhibition in Lung Cancer. *Sci. Signaling* **2013,** *6*, re6.

99. Engelman, J. A.; Zejnullahu, K.; Mitsudomi, T.; Song, Y.; Hyland, C.; Park, J. O., . . . Janne, P. A., MET amplification leads to gefitinib resistance in lung cancer by activating ErbB3 signaling. *Science* **2007,** *316*, 1039-1043.

100. Bean, J.; Brennan, C.; Shih, J. Y.; Riely, G.; Viale, A.; Wang, L., . . . Pao, W., MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib. *Proc. Natl. Acad. Sci. USA* **2007,** *104* (52), 20932-20937.

101. Yano, S.; Wang, W.; Li, Q.; Matsumoto, K.; Sakurama, H.; Nakamura, T., . . . Sone, S., Hepatocyte growth factor induces gefitinib resistance of lung adenocarcinoma with epidermal growth factor receptor-activating mutations. *Cancer Res.* **2008,** *68* (22), 9479-9487.

102. Turke, A. B.; Zejnullahu, K.; Wu, Y. L.; Song, Y.; Dias-Santagata, D.; Lifshits, E., . . . Janne, P. A., Preexistence and clonal selection of MET amplification in EGFR mutant NSCLC. *Cancer Cell* **2010,** *17* (1), 77-88.

103. Zhang, Z.; Lee, J. C.; Lin, L.; Olivas, V.; Au, V.; LaFramboise, T., . . . Bivona, T. G., Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. *Nature Genetics* **2012,** *44* (8), 852-860.

104. Meyer, A. S.; Miller, M. A.; Gertler, F. B.; Lauffenburger, D. A., The Receptor AXL Diversifies EGFR Signaling and Limits the Response to EGFR-Targeted Inhibitors in Triple-Negative Breast Cancer Cells. *Sci. Signal.* **2013,** *6*, ra66.

105. Guix, M.; Faber, A. C.; Wang, S. E.; Olivares, M. G.; Song, Y.; Qu, S., . . . Engelman, J. A., Acquired resistance to EGFR tyrosine kinase inhibitors in cancer cells is mediated by loss of IGF-binding proteins. *J. Clin. Inv.* **2008,** *118* (7), 2609-2619.

106. Cortot, A. B.; Repellin, C. E.; Shimamura, T.; Capelletti, M.; Zejnullahu, K.; Ercan, D., . . . Janne, P. A., Resistance to irreversible EGF receptor tyrosine kinase inhibitors through a multistep mechanism involving the IGF1R pathway. *Cancer Res.* **2013,** *73* (2), 834-843.

107. Yonesaka, K.; Zejnullahu, K.; Okamoto, I.; Satoh, T.; Cappuzzo, F.; Souglakos, J., . . . Janne, P. A., Activation of ErbB2 signaling causes resistance to EGF directed therapeutic antibody cetuximab. *Sci. Transl. Med.* **2011,** *3* (99), 99ra86.

108. Zhang, L.; Castanaro, C.; Luan, B.; Yang, K.; Fan, L.; Fairhurst, J. L., . . . Daly, C., ERBB3/HER2 Signaling Promotes Resistance to EGFR Blockade in Head and Neck and Colorectal Cancer Models. *Mol. Cancer Ther.* **2014,** *13* (5), 1345-1355.

109. Mancini, M.; Gaborit, N.; Lindzen, M.; Salame, T.M.; Dall'Ora, M.; Sevilla-Sharon, M., . . . Yarden, Y., Combining three antibodies nulifies feedback-mediated resistance to erlotinib in lung cancer. *Sci. Signaling* **2015,** *8* (379), ra53.

110. Engelman, J. A.; Mukohara, T.; Zejnullahu, K.; Lifshits, E.; Borras, A. M.; Gale, C. M., . . . Janne, P. A., Allelic dilution obscures detection of a biologically significant resistance mutation in EGFR-amplified lung cancer. *J. Clin. Inv.* **2006,** *116*, 2695-2706.

111. Sos, M. L.; Koker, M.; Weir, B. A.; Heynck, S.; Rabinovsky, R.; Zander, T., . . . Thomas, R. K., PTEN loss contributes to erlotinib resistance in EGFR-mutant lung cancer by activation of Akt and EGFR. *Cancer Res.* **2009,** *69* (8), 3256-3261.

112. Misale, S.; Yaeger, R.; Hobor, S.; Scala, E.; Janakiraman, M.; Liska, D., . . . Bardelli, A., Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **2012,** *486* (7404), 532-536.

113. Ohashi, K.; Sequist, L. V.; Arcila, M. E.; Moran, T.; Chmielecki, J.; Lin, Y. L., . . . Pao, W., Lung cancers with acquired resistance to EGFR inhibitors occasionally harbor BRAF gene mutations but lack mutations in KRAS, NRAS, or MEK1. *Proc. Natl. Acad. Sci. USA* **2012,** *109* (31), E2127-33.

114. Ercan, D.; Xu, C.; Yanagita, M.; Monast, C. S.; Pratilas, C. A.; Montero, J., . . . Janne, P. A., Reactivation of ERK signaling causes resistance to EGFR kinase inhibitors. *Cancer Discov.* **2012,** *2* (10), 934-947.

115. Han, W.; Lo, H. W., Landscape of EGFR signaling network in human cancers: Biology and therapeutic response in relation to receptor subcellular locations. *Cancer Letters* **2012,** *318*, 124-134.

116. Faber, A. C.; Corcoran, R. B.; Ebi, H.; Sequist, L.; Waltman, B. A.; Chung, E., . . . Engelman, J. A., BIM Expression in Treatment-Naïve Cancers Predicts Responsiveness to Kinase Inhibitors. *Cancer Discov.* **2011,** *1* (4), 353-365.

117. Thomson, S.; Buck, E.; Petti, F.; Griffin, G.; Brown, E.; Ramnarine, N., . . . Haley, J. D., Epithelial to Mesenchymal Transition Is a Determinant of Sensitivity of Non–Small-Cell Lung Carcinoma Cell Lines and Xenografts to Epidermal Growth Factor Receptor Inhibition. *Cancer Res.* **2005,** *65*, 9455-9462.

118. Li, L.; Gu, X.; Yue, J.; Zhao, Q.; Lv, D.; Chen, H.; Xu, L., Acquisition of EGFR TKI resistance and EMT phenotype is linked with activation of IGF1R/NF-κB pathway in EGFR-mutant NSCLC. *Oncotarget* **2017,** *8* (54), 92240-92253.

119. Tan, X.; Lambert, P. F.; Rapraeger, A. C.; Anderson, R. A., Stress-Induced EGFR Trafficking: Mechanisms, Functions, and Therapeutic Implications. *Trends Cell Biology* **2016,** *26* (5), 352-366.

120. Sigismund, S.; Avanzato, D.; Lanzetti, L., Emerging functions of the EGFR in cancer. *Mol. Oncol.* **2018,** *12*, 3-20.

121. Brand, T. M.; Iida, M.; Li, C.; Wheeler, D. L., The Nuclear Epidermal Growth Factor Receptor Signaling Network and its Role in Cancer. *Discov. Med.* **2011,** *12* (66), 419-432.

122. Brand, T. M.; Iida, M.; Luthar, N.; Starr, M. M.; Huppert, E. J.; Wheeler, D. L., Nuclear EGFR as a molecular target in cancer. *Radiother Oncol* **2013,** *108* (3), 370-7.

123. Ali, R.; Wendt, M. K., The paradoxical functions of EGFR during breast cancer progression. *Signal Transduct Target Ther* **2017,** *2*.

124. Li, C.; Iida, M.; Dunn, E. F.; Ghia, A. J.; Wheeler, D. L., Nuclear EGFR contributes to acquired resistance to cetuximab. *Oncogene* **2009,** *28* (43), 3801-13.

125. Iida, M.; Brand, T. M.; Campbell, D. A.; Wheeler, D. L., Yes and Lyn play a role in nuclear translocation of the epidermal growth factor receptor. *Oncogene* **2013,** *32*, 759-767.

126. Brand, T. M.; Iida, M.; Corrigan, K. L.; Braverman, C. M.; Coan, J. P.; Flanigan, B. G., . . . Wheeler, D. L., The receptor tyrosine kinase AXL mediates nuclear translocation of the epidermal growth factor receptor. *Sci. Signal.* **2017,** *10*, eaag1064.

127. Huang, W. C.; Chen, Y. J.; Li, L. Y.; Wei, Y. L.; Hsu, S. C.; Tsai, S. L., . . . Hung, M. C., Nuclear translocation of epidermal growth factor receptor by Akt-dependent phosphorylation enhances breast cancer-resistant protein expression in gefitinib-resistant cells. *J Biol Chem* **2011,** *286* (23), 20558-68.

128. Wanner, G.; Mayer, C.; Kehlbach, R.; Rodemann, P., H.; Dittmann, K., Activation of protein kinase Cε stimulates DNA-repair via epidermal growth factor receptor nuclear accumulation. *Radiother. Oncol.* **2008,** *86*, 383-390.

129. Boerner, J. L.; Demory, M. L.; Silva, C.; Parsons, S. J., Phosphorylation of Y845 on the Epidermal Growth Factor Receptor Mediates Binding to the Mitochondrial Protein Cytochrome c Oxidase Subunit II. *Mol. Cell. Biol.* **2004,** *24* (16), 7059-7071.

130. Cao, X.; Zhu, H.; Ali-Osman, F.; Lo, H. W., EGFR and EGFRvIII undergo stress- and EGFR kinase inhibitor-induced mitochondrial translocation: A potential mechanism of EGFR-driven antagonism of apoptosis. *Molecular Cancer* **2011,** *10*, 26.

131. Chavez, K. J.; Garimella, S. V.; Lipkowitz, S., Triple Negative Breast Cancer Cell Lines: One Tool in the Search for Better Treatment of Triple Negative Breast Cancer. *Breast Dis.* **2010,** *31*, 35-48.

132. R, D.; Trudaeu, M.; Pritchard, K. I.; Hanna, W. M.; Kahn, H. K.; Sawka, C. A., . . . Narod, S. A., Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin. Cancer Res.* **2007,** *13*, 4429-4434.

133. Garrido-Castro, A. C.; Lin, N. U.; Polyak, K., Insights into Molecular Classifi cations of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment. *Cancer Discov.* **2019,** *9*, 176-198.

134. Changavi, A. A.; Shashikala, A.; Ramji, A. S., Epidermal Growth Factor Receptor Expression in Triple Negative and Nontriple Negative Breast Carcinomas. *J. Lab Physicians* **2015,** *7*, 79-83.

135. Park, H. S.; Jang, M. H.; Kim, E. J.; Kim, H. J.; Lee, H. J.; Kim, Y. J., . . . Park, S. Y., High EGFR gene copy number predicts poor outcome in triple-negative breast cancer. *Mod. Pathol.* **2014,** *27*, 1212-1222.

136. Nakai, K.; Hung, M. C.; Yamaguchi, H., A perspective on anti-EGFR therapies targeting triple-negative breast cancer. *Am. J. Transl. Res.* **2016,** *6*, 1609-1623.

137. Costa, R.; Shah, A. N.; Santa-Maria, C. A.; Cruz, M. R.; Mahalingam, D.; Carneiro, B. A., . . . Giles, F. J., Targeting Epidermal Growth Factor Receptor in triple negative breast cancer: New discoveries and practical insights for drug development. *Cancer Treat. Rev.* **2017,** *53*, 111-119.

138. Crown, J.; O'Shaughnessy, J.; Gullo, G., Emerging targeted therapies in triple-negative breast cancer. *Annals Oncol.* **2012,** *23 Suppl 6*, vi56-vi65.

139. Hachem, G. E. H., EGFR inhibition in metastatic triple negative breast cancer: a losing target. *Int. Clin. Pathol. J.* **2018,** *6*, 134-135.

140. Brand, T. M.; Iida, M.; Dunn, E. F.; Luthar, N.; Kostopoulos, K. T.; Corrigan, K. L., . . . Wheeler, D. L., Nuclear epidermal growth factor receptor is a functional molecular target in triple-negative breast cancer. *Mol. Cancer Ther.* **2014,** *13* (5), 1356-68.

141. Yu, Y. L.; Chou, R. H.; Liang, J. H.; Chang, W. J.; Su, K. J.; Tseng, Y. J., . . . Hung, M. C., Targeting the EGFR/PCNA signaling suppresses tumor growth of triple-negative breast cancer cells with cell-penetrating PCNA peptides. *PloS one* **2013,** *8* (4), e61362.

142. Diluvio, G.; Del Gaudio, F.; Giuli, M. V.; Franciosa, G.; Giuliani, E.; Palermo, R., . . . Checquolo, S., NOTCH3 inactivation increases triple negative breast cancer sensitivity to gefitinib by promoting EGFR tyrosine dephosphorylation and its intracellular arrest. *Oncogenesis* **2018,** *7*, 42.

143. Irwin, M. E.; Bohin, N.; Boerner, J. L., Src family kinases mediate epidermal growth factor receptor signaling from lipid rafts in breast cancer cells. *Cancer Biol. Ther.* **2011,** *12* (8), 718-726.

144. Lim, S.-O.; Li, C.-W.; Xia, W.; Lee, H.-H.; Chang, S.-S.; Shen, J., . . . Hung, M.-C., EGFR Signaling Enhances Aerobic Glycolysis in Triple-Negative Breast Cancer Cells to Promote Tumor Growth and Immune Escape. *Cancer Res.* **2016,** *76* (5), 1284-1296.

145. Lanning, N. J.; Castle, J. P.; Singh, S. J.; Leon, A. N.; Tovar, E. A.; Sanghera, A., . . . Graveel, C. R., Metabolic profiling of triple-negative breast cancer cells reveals metabolic vulnerabilities. *Cancer & Metabolism* **2017,** *5:6*.

146. Merino, D.; Whittle, J. R.; Vaillant, F.; Serrano, A.; Gong, J.-N.; Giner, G., . . . Lindeman, G. L., ynergistic action of the MCL-1 inhibitor S63845 with current therapies in preclinical models of triple-negative and HER2-amplified breast cancer. *Sci. Transl. Med.* **2017,** *9*, eaam7049.

147. Liu, Z.; He, K.; Ma, Q.; Yu, Q.; Liu, C.; Ndege, I., . . . Yu, Z., Autophagy inhibitor facilitates gefitinib sensitivity in vitro and in vivo by activating mitochondrial apoptosis in triple negative breast cancer. *PloS one* **2017,** *12* (5), e0177694.

148. Tao, J. J.; Castel, P.; Radosevic-Robin, N.; Elkabets, M.; Auricchio, N.; Aceto, N., . . . Scaltriti, M., Antagonism of EGFR and HER3 Enhances the Response to Inhibitors of the PI3K-Akt Pathway in Triple-Negative Breast Cancer. *Sci. Signal.* **2014,** *7*, ra29.

149. Kim, Y. J.; Choi, J.-S.; Seo, J.; Song, J.-Y.; Lee, S. E.; Kwon, M. J., . . . Choi, Y. L., MET is a potential target for use in combination therapy with EGFR inhibition in triple-negative/basal-like breast cancer. *Int. J. Cancer* **2014,** *134*, 2424-2436.

150. Linklater, E. S.; Tovar, E. A.; Essenburg, C. J.; Turner, L.; Madaj, Z.; Winn, M. E., . . . Graveel, C. R., Targeting MET and EGFR crosstalk signaling in triple-negative breast cancers. *Oncotarget* **2016,** *7*, 69903-69915.

151. Yi, Y. W.; Hong, W.; Kang, H. J.; Kim, H. J.; Zhao, W.; Wang, A., . . . Bae, I., Inhibition of the PI3K/AKT pathway potentiates cytotoxicity of EGFR kinase inhibitors in triple-negative breast cancer cells. *J. Cell Mol. Med.* **2013,** *17* (5), 648-656.

152. Liu, T.; Yacoub, R.; Taliaferro-Smith, L. D.; Sun, S.-Y.; Graham, T. R.; Dolan, R., . . . O'Regan, R. M., Combinatorial Effects of Lapatinib and Rapamycin in Triple-Negative Breast Cancer Cells. *Mol. Cancer Ther.* **2011,** *10,* 1460-1469.

153. You, K. S.; Yi, Y. W.; Kwak, S.-J.; Seong, Y.-S., Inhibition of RPTOR overcomes resistance to EGFR inhibition in triple-negative breast cancer cells. *Int. J. Oncol.* **2018,** *52,* 828-840.

154. Maiello, M. R.; D'Alessio, A.; Bevilacqua, S.; Gallo, M.; Normanno, N.; De Luca, A., EGFR and MEK Blockade in Triple Negative Breast Cancer Cells. *J. Cell Biochem.* **2015,** *116,* 2778-2785.

155. Verma, N.; Muller, A.-K.; Kothari, C.; Panayotopolou, E.; Kedan, A.; Selitrennik, M., . . . Lev, S., Targeting of PYK2 Synergizes with EGFR Antagonists in Basal-like TNBC and Circumvents HER3-Associated Resistance via the NEDD4–NDRG1 Axis. *Cancer Res.* **2017,** *77* (1), 86-99.

156. Tai, Y. L.; Chu, P. Y.; Lai, I. R.; Wang, M. Y.; Tseng, H. Y.; Guan, J. L., . . . Shen, T. L., An EGFR/Src-dependent beta4 integrin/FAK complex contributes to malignancy of breast cancer. *Sci Rep* **2015,** *5,* 16408.

157. Giro-Perafita, A.; Palomeras, S.; Lum, D. H.; Blancafort, A.; Viñas, G.; Oliveras, G., . . . Puig, T., Preclinical Evaluation of Fatty Acid Synthase and EGFRInhibition in Triple-NegativeBreastCancer. *Clin. Cancer Res.* **2016,** *22* (18), 4687-4697.

158. Siegel, R. L.; Miller, K. D.; Jemal, A., Cancer Statistics, 2019. *CA Cancer J. Clin.* **2019,** *69,* 7-34.

159. Meza, R.; Meernik, C.; Jeon, J.; Cote, M. L., Lung Cancer Incidence Trends by Gender, Race and Histology in the United States, 1973–2010. *PloS one* **2015,** *10* (3), e0121323.

160. Bethune, G.; Bethune, D.; Ridgway, N.; Xu, Z., Epidermal growth factor receptor (EGFR) in lung cancer: an overview and update. *J. Thorac. Dis.* **2010,** *2,* 48-51.

161. Rotow, J.; Bivona, T. G., Understanding and targeting resistance mechanisms in NSCLC. *Nat. Rev. Cancer* **2017,** *17,* 637-658.

162. Pabla, B.; Bissonnette, M.; Konda, V. J., Colon cancer and the epidermal growth factor receptor: Current treatment paradigms, the importance of diet, and the role of chemoprevention. *World J. Clin. Oncol.* **2015,** *6* (5), 133-141.

163. Cunningham, D.; Humblet, Y.; Siena, S.; Khayat, D.; Bleiberg, H.; Santoro, A., . . . Van Cutsem, E., Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N. Engl. J. Med.* **2004,** *351* (4), 337-345.

164. Khambata-Ford, S.; Garrett, C. R.; Meropol, N. J.; Basik, M.; Harbison, C. T.; Wu, S. H., . . . Mauro, D. J., Expression of epiregulin and amphiregulin and K-ras mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab. *J. Clin. Oncol.* **2007,** *25* (22), 3230-3237.

165. Alaimo, P. J.; Shogren-Knaak, M. A.; Shokat, K. M., Chemical genetic approaches for the elucidation of signaling pathways. *Curr. Opin. Cell Biol.* **2001,** *5,* 360-367.

166. Weiss, W. A.; Taylor, S. S.; Shokat, K. M., Recognizing and exploiting differences between RNAi and small-molecule inhibitors. *Nat. Chem. Biol.* **2007,** *3* (12), 739-744.

167. Blagg, J.; Workman, P., Chemical biology approaches to target validation in cancer. *Curr. Opin. Pharmacol.* **2014,** *17,* 87-100.

168. Schreiber, S. L., The small-molecule approach to biology. *Chem. Eng. News* **2003,** *81,* 51-61.

169. Arrowsmith, C. H.; Audia, J. E.; Austin, C.; Baell, J.; Bennett, J.; Blagg, J., . . . Zuercher, W. J., The promise and peril of chemical probes. *Nat. Chem. Biol.* **2015,** *11* (8), 536-541.

170. Blagg, J.; Workman, P., Choose and Use Your Chemical Probe Wisely to Explore Cancer Biology. *Cancer Cell Perspect.* **2017,** *32* (1), 9-25.

171. Bunnage, M. E.; Chekler, E. L.; Jones, L. H., Target validation using chemical probes. *Nat. Chem. Biol.* **2013,** *9* (4), 195-199.

172. Jeffery, D. A.; Bogyo, M., Chemical proteomics and its application to drug discovery. *Curr. Opin. Biotechnol.* **2003,** *14,* 87-95.

173. Sadaghiani, A. M.; Verhelst, S. H.; Bogyo, M., Tagging and detection strategies for activity-based proteomics. *Curr. Opin. Chem. Biol.* **2007,** *11* (1), 20-28.

174. Ziegler, S.; Pries, V.; Hedberg, C.; Waldmann, H., Target identification for small bioactive molecules: finding the needle in the haystack. *Angewandte Chemie* **2013,** *52* (10), 2744-92.

175. Brehmer, D.; Greff, Z.; Godl, K.; Blencke, S.; Kurtenbach, A.; Weber, M., . . . Daub, H., Cellular Targets of Gefitinib. *Cancer Res.* **2005,** *65* (2), 379-382.

176. Sharma, K.; Weber, C.; Bairlein, M.; Greff, Z.; Keri, G.; Cox, J., . . . Daub, H., Proteomics strategy for quantitative protein interaction profiling in cell extracts. *Nat. Methods* **2009,** *6* (10), 741-744.

177. Weber, C.; Schreiber, T. B.; Daub, H., Dual phosphoproteomics and chemical proteomics analysis of erlotinib and gefitinib interference in acute myeloid leukemia cells. *J.Proteomics* **2012,** *75,* 1343-1356.

178. Augustin, A.; Lamerz, J.; Meistermann, H.; Golling, S.; Scheiblich, S.; Hermann, J. C., . . . Klughammer, B., Quantitative chemical proteomics profiling differentiates erlotinib from gefitinib in EGFR wild-type non-small cell lung carcinoma cell lines. *Mol. Cancer Ther.* **2013,** *12* (4), 520-9.

179. Wissing, J.; Jansch, L.; Nimtz, M.; Dieterich, G.; Hornberger, R.; Keri, G., . . . Daub, H., Proteomics Analysis of Protein Kinases by Target Class-selective Prefractionation and Tandem Mass Spectrometry. *Mol. Cel. Proteomics* **2007,** *6,* 537-547.

180. Daub, H.; Olsen, J. V.; Bairlein, M.; Gnad, F.; Oppermann, F. S.; Korner, R., . . . Mann, M., Kinase-Selective Enrichment Enables Quantitative Phosphoproteomics of the Kinome across the Cell Cycle. *Mol. Cell* **2008,** *31*, 438-448.

181. Grotzfeld, R. M.; Milanov, Z.; Patel, H., K; Lai, A. G.; Mehta, S.; Lockhart, D. J. Conjugated small molecules. US 2005/0153371 A1, Jul 14, 2005, 2005.

182. Salazar Estrada, I. J. Molecular tools for elucidating signaling networks of the protein kinase EGFR. Macquarie University, Sydney, Australia, 2015.

183. Sun, H.; Ren, Y.; Hou, W.; Li, L.; Zeng, F.; Li, S., . . . Zhang, Z., Focusing on probe-modified peptides: a quick and effective method for target identification. *Chem. Commun.* **2016,** *52*, 10225-10228.

184. Boyce, J. P.; Brown, M. E.; Fitzner, J. N.; Kowski, T. J. Kinase-directed, activity-based probes. US 2007/0009977 A1, Jan 11 2007, 2007.

185. Blair, J. A.; Rauh, D.; Kung, C.; Yun, C. H.; Fan, Q. W.; Rode, H., . . . Shokat, K. M., Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. *Nat. Chem. Biol.* **2007,** *3* (4), 229-38.

186. Barkovich, K. J.; Hariono, S.; Garske, A. L.; Zhang, J.; Blair, J. A.; Fan, Q. W., . . . Weiss, W. A., Kinetics of inhibitor cycling underlie therapeutic disparities between EGFR-driven lung and brain cancers. *Cancer Discov.* **2012,** *2* (5), 450-457.

187. Lanning, B. R.; Whitby, L. R.; Dix, M. M.; Douhan, J.; Gilbert, A. M.; Hett, E. C., . . . Cravatt, B. F., A road map to evaluate the proteome-wide selectivity of covalent kinase inhibitors. *Nat. Chem. Biol.* **2014,** *10* (9), 760-7.

188. Niessen, S.; Dix, M. M.; Barbas, S.; Potter, Z. E.; Lu, S.; Brodsky, O., . . . Cravatt, B. F., Proteome-wide Map of Targets of T790M-EGFR-Directed Covalent Inhibitors. *Cell Chem. Biol.* **2017,** *24*, 1388-1400.

189. Shreder, K.; Wong, M. S.; Nomanbhoy, T.; Leventhal, P. S.; Fuller, S. R., Synthesis of AX7593, a quinazoline-derived photoaffinity probe for EGFR. *Org. Lett.* **2004,** *6* (21), 3715-3718.

190. Li, Z.; Hao, P.; Li, L.; Tan, C. Y.; Cheng, X.; Chen, G. Y., . . . Yao, S. Q., Design and synthesis of minimalist terminal alkyne-containing diazirine photo-crosslinkers and their incorporation into kinase inhibitors for cell- and tissue-based proteome profiling. *Angew. Chem. Int. Ed.* **2013,** *52* (33), 8551-8556.

191. Conradt, L.; Godl, K.; Schaab, C.; Tebbe, A.; Eser, S.; Diersch, S., . . . Schneider, G., Disclosure of Erlotinib as a Multikinase Inhibitor in Pancreatic Ductal Adenocarcinoma. *Neoplasia* **2011,** *13* (11), 1026-1034.

192. Sinz, A., Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J. Mass Spectrom. : JMS* **2003,** *38*, 1225-1237.

193. Merkley, E. D.; Cort, J. R.; Adkins, J. N., Cross-linking and mass spectrometry methodologies to facilitate structural biology: finding a path through the maze. *J. Struct. Funct. Genomics* **2013,** *14*, 77-90.

194. Turk, H. F.; Chapkin, R. S., Analysis of epidermal growth factor receptor dimerization by BS(3) cross-linking. In *Receptor Tyrosine Kinases: Methods and Protocols*, Germano, S., Ed. Springer: 2015; Vol. 1233, pp 25-34.

195. He, L.; Hristova, K., Consequences of replacing EGFR juxtamembrane domain with an unstructured sequence. *Sci. Rep.* **2012,** *2:854*, srep00854.

196. Brown, P. M.; Debanne, M. T.; Grothe, S.; Bergsma, D.; Caron, M.; Kay, C.; O'Connor-McCourt, M. D., The extracellular domain of the epidermal growth factor receptor. Studies on the affinity and stoichiometry of binding, receptor dimerization and a binding-domain mutant. *Eur. J. Biochem.* **1994,** *225*, 223-233.

197. Pettinger, J.; Jones, K.; Cheeseman, M. D., Lysine-Targeting Covalent Inhibitors. *Angew. Chem. Int. Ed.* **2017,** *56*, 15200-15209.

198. Dalton, S. E.; Dittus, L.; Thomas, D. A.; Convery, M. A.; Nunes, J.; Bush, J. T., . . . Campos, S., Selectively Targeting the Kinome-Conserved Lysine of PI3Kδ as a General Approach to Covalent Kinase Inhibition. *J. Am. Chem. Soc.* **2018,** *140*, 932-939.

199. Pettinger, J.; Le Bihan, Y.-V.; Widya, M.; van Montfort, R. L. M.; Jones, K.; Cheeseman, M. D., An Irreversible Inhibitor of HSP72 that Unexpectedly Targets Lysine-56. *Angew. Chem. Int. Ed.* **2017,** *2017*, 3536-3540.

200. Akcay, G.; Belmonte, M. A.; Aquila, B.; Chuaqui, C.; Hird, A. W.; Lamb, M. L., . . . Su, Q., Inhibition of Mcl-1 through covalent modification of a noncatalytic lysine side chain. *Nat. Chem. Biol.* **2016,** *12*, 931-936.

201. Shannon, D. A.; Banerjee, R.; Webster, E. R.; Bak, D. W.; Wang, C.; Weerapana, E., Investigating the proteome reactivity and selectivity of aryl halides. *J. Am. Chem. Soc.* **2014,** *136* (9), 3330-3333.

202. Hacker, S. M.; Backus, K. M.; Lazear, M. R.; Forli, S.; Correia, B. E.; Cravatt, B. F., Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **2017,** *9*, 1181-1190.

203. Ward, C. C.; Kleinman, J. I.; Nomura, D. K., NHS-Esters As Versatile Reactivity-Based Probes for Mapping Proteome-Wide Ligandable Hotspots. *ACS Chem. Biol.* **2017,** *12*, 1478-1483.

204. Zhao, Q.; Ouyang, X.; Wan, X.; Gajiwala, K. S.; Kath, J. C.; Jones, L. H., . . . Taunton, J., Broad-Spectrum Kinase Profiling in Live Cells with Lysine-Targeted Sulfonyl Fluoride Probes. *J. Am. Chem. Soc.* **2017,** *139*, 680-685.

205. Choi, S.; Connelly, S.; Reixach, N.; Wilson, I. A.; Kelly, J. W., Chemoselective small molecules that covalently modify one lysine in a non-enzyme protein in plasma. *Nat. Chem. Biol.* **2010,** *6* (2), 133-139.

206. Tamura, T.; Tsukiji, S.; Hamachi, I., Native FKBP12 engineering by ligand-directed tosyl chemistry: labeling properties and application to photo-cross-linking of protein complexes in vitro and in living cells. *J. Am. Chem. Soc.* **2012,** *134* (4), 2216-2226.

207. Yamaura, K.; Kuwata, K.; Tamura, T.; Kioi, Y.; Takaoka, Y.; Kiyonaka, S.; Hamachi, I., Live cell off-target identification of lapatinib using ligand-directed tosyl chemistry. *Chem. Commun.* **2014,** *50* (91), 14097-14100.

208. Takaoka, Y.; Nishikawa, Y.; Hashimoto, Y.; Sasaki, K.; Hamachi, I., Ligand-directed dibromophenyl benzoate chemistry for rapid and selective acylation of intracellular natural proteins. *Chem. Sci.* **2015,** *6* (5), 3217-3224.

209. Shruthi, B. S.; Vinodhkumar, P.; Selvamani, Proteomics: A new perspective for cancer. *Adv. Biomed. Res.* **2016,** *19*, 5:67.

210. Kolch, W.; Pitt, A., Functional proteomics to dissect tyrosine kinase signalling pathways in cancer. *Nat. Rev. Cancer* **2010,** *10* (9), 618-629.

211. Choudhary, C.; Mann, M., Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell. Biol.* **2010,** *11*, 427-439.

212. Wu, S.-L.; Kim, J.; Bandle, R. W.; Liotta, L.; Petricoin, E.; Karger, B. L., Dynamic Profiling of the Post-translational Modifications and Interaction Partners of Epidermal Growth Factor Receptor Signaling after Stimulation by Epidermal Growth Factor Using Extended Range Proteomic Analysis (ERPA). *Mol. Cell Proteomics* **2006,** *5*, 1610-1627.

213. Tong, J.; Taylor, P.; Morant, M. F., Proteomic Analysis of the Epidermal Growth Factor Receptor (EGFR) Interactome and Post-translational Modifications Associated with Receptor Endocytosis in Response to EGF and Stress. *Mol. Cell Proteomics* **2014,** *13*, 1644-1658.

214. Francavilla, C.; Papetti, M.; Rigbolt, K. T.; Pedersen, A. K.; Sigurdsson, J. O.; Cazzamali, G., . . . Olsen, J. V., Multilayered proteomics reveals molecular switches dictating ligand-dependent EGFR trafficking. *Nat. Struct. Mol. Biol.* **2016,** *23* (6), 608-18.

215. Koch, H.; Wilhelm, M.; Ruprecht, B.; Beck, S.; Frejno, M.; Klaeger, S.; Kuster, B., Phosphoproteome Profiling Reveals Molecular Mechanisms of Growth-Factor-Mediated Kinase Inhibitor Resistance in EGFROverexpressing Cancer Cells. *J. Prot. Res.* **2016,** *15*, 4490-4504.

216. Li, J.; Bennet, K.; Stukalov, A.; Fang, B.; Zhang, G.; Yoshida, T., . . . Haura, E. B., Perturbation of the mutated EGFR interactome identifies vulnerabilities and resistance mechanisms. *Mol. Systems Biol.* **2013,** *9*, 705.

217. Mulder, C.; Prust, N.; van Doorn, S.; Reinecke, M.; Kuster, B.; van Bergen En Henegouwen, P.; S., L., Adaptive Resistance to EGFR-Targeted Therapy by Calcium Signaling in NSCLC Cells. *Mol. Cancer Res.* **2018,** *16* (11), 1773-1784.

218. Li, W.; Wang, H.; Yang, Y.; Zhao, T.; Zhang, Z.; Tian, Y., . . . Zhang, F., Integrative Analysis ofProteome and Ubiquitylome Reveals Unique Features ofLysosomal and Endocytic Pathways in Gefitinib-Resistant Non-Small Cell Lung Cancer Cells. *Proteomics* **2018,** *18*, 1700388.

219. Kani, K.; Faca, V. M.; Hughes, L. D.; Zhang, W.; Fang, Q.; Shahbaba, B., . . . Mallick, P., Quantitative Proteomic Profiling Identifies Protein Correlates to EGFR Kinase Inhibition. *Mol. Cancer Ther.* **2012,** *11* (5), 1071-1081.

220. Palcy, S.; Chevet, E., Integrating forward and reverse proteomics to unravel protein function. *Proteomics* **2006,** *6*, 5467-5480.

221. Rix, U.; Superti-Furga, G., Target profiling of small molecules by chemical proteomics. *Nat. Chem. Biol.* **2009,** *5* (9), 616-624.

222. Saghatelian, A.; Jessani, N.; Joseph, A.; Humphrey, M.; Cravatt, B. F., Activity-based probes for the proteomic profiling of metalloproteases. *Proc. Nat. Acad. Sci. USA* **2004,** *101* (27), 10000-10005.

223. Miao, Q.; Zhang, C.-C.; Kast, J., Chemical proteomics and its impact on the drug discovery process. *Expert. Rev. Proteomics* **2012,** *9* (3), 281-291.

224. Koch, H.; Del Castillo Busto, M. E.; Kramer, K.; Medard, G.; Kuster, B., Chemical Proteomics Uncovers EPHA2 as a Mechanism of Acquired Resistance to Small Molecule EGFR Kinase Inhibition. *J. Prot. Res.* **2015,** *14*, 2617-2625.

225. Kawatani, M.; Osada, H., Affinity-based target identification for bioactive small molecules. *Med. Chem. Commun.* **2014,** *5*, 277.

226. Becker, F.; Murthi, K.; Smith, C.; Come, J.; Costa-Roldan, N.; Kaufmann, C., . . . Kley, N., A Three-Hybrid Approach to Scanning the Proteome for Targets of Small Molecule Kinase Inhibitors. *Chem. Biol.* **2004,** *11*, 211-233.

227. Takakusagi, Y.; Takakusagi, K.; Sugawara, F.; Sakaguchi, K., Use of phage display technology for the determination of the targets for small-molecule therapeutics. *Exp. Opin. Drug Discov.* **2010,** *5* (4), 361-389.

228. Van Dorst, B.; Mehta, J.; Rouah-Martin, E.; Blust, R.; Robbens, J., Phage display as a method for discovering cellular targets of small molecules. *Methods* **2012,** *58*, 56-61.

CHAPTER 2

Design and Synthesis of Chemical

Probes Targeting EGFR

2.1 INTRODUCTION

2.1.1 Structure of EGFR inhibitors and probes

Small-molecule EGFR inhibitors comprise three generations. First (Gefitinib, Erlotinib) and second (Afatinib, Dacomitinib) generation inhibitors target WT EGFR, while third generation inhibitors are specific for the T790M EGFR resistant mutant form. Both first and second generation inhibitors are based on the 4-anilinoquinazoline pharmacophore, which was first identified as selective EGFR selective inhibitor in 1994.[1-3] These drugs bind to the EGFR kinase domain in its active conformation (**Fig. 2.1**).[4-7]

The inhibitors bind with the N1 atom of quinazoline oriented at the back of the ATP-binding pocket, with the N1- and C7-motifs of the quinazoline directed to the hinge region connecting the N and C lobes of EGFR (**Fig 2.1**). Atom N1 of quinazoline accepts a hydrogen bond from the NH backbone of M793 residue, while atom N3 contacts the side chain of the gatekeeper residue T790 through a water molecule. The quinazoline and aniline group form an interplanar angle of approximately 42-45°. Gefitinib, Afatinib and Dacomitinib share the same 3-chloro-4-fluoroaniline moiety, which binds in the gatekeeper hydrophobic pocket (**Fig. 2.1a,c**). For Erlotinib, the acetylene moiety in the aniline ring is located in a pocket behind the ATP binding site, which can only be accessed if the gatekeeper residue is small (**Fig. 2.1b**). Substituents of atoms C6 and C7 of quinazoline are solvent exposed and do not significantly affect binding affinity, but are generally derivatized to improve pharmacokinetic properties. These positions are commonly used to modify EGFR inhibitors into chemical probes. In the second generation inhibitors Afatinib and Dacomitinib, atom C6 is substituted with a crotonamide derivative that functions as a Michael acceptor, allowing the formation of a covalent bond with C797 located at the edge of the active site.
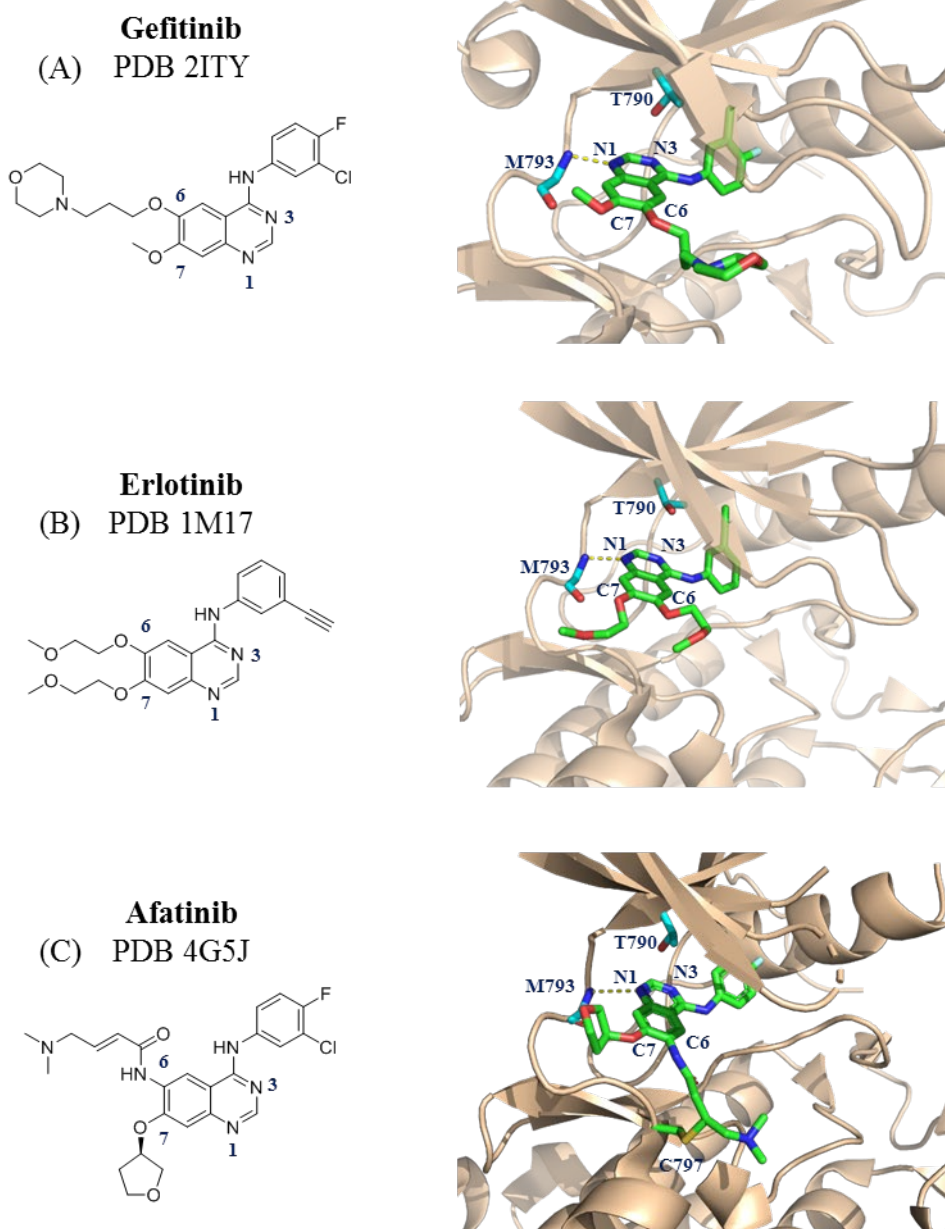
Fig. 2.1 Crystal structure of EGFR inhibitors bound to the tyrosine kinase domain (TKD)[4-6]

2.1.2 Synthetic approaches to EGFR probes

EGFR-targeting chemical probes can be classified in two major groups: reversible and covalent. As observed in the crystal structures of the parent inhibitors, the solvent-exposed atoms C6 and C7 of the quinazoline core are usually the starting point for linker and reporter attachment. Most reversible probes are obtained from inhibitor modifications at the C6 position. In covalent probes, this position is reserved for the electrophilic warhead mediating

the covalent labelling of the C797 residue, therefore the linker and reporter groups are typically attached through the C7 atom.
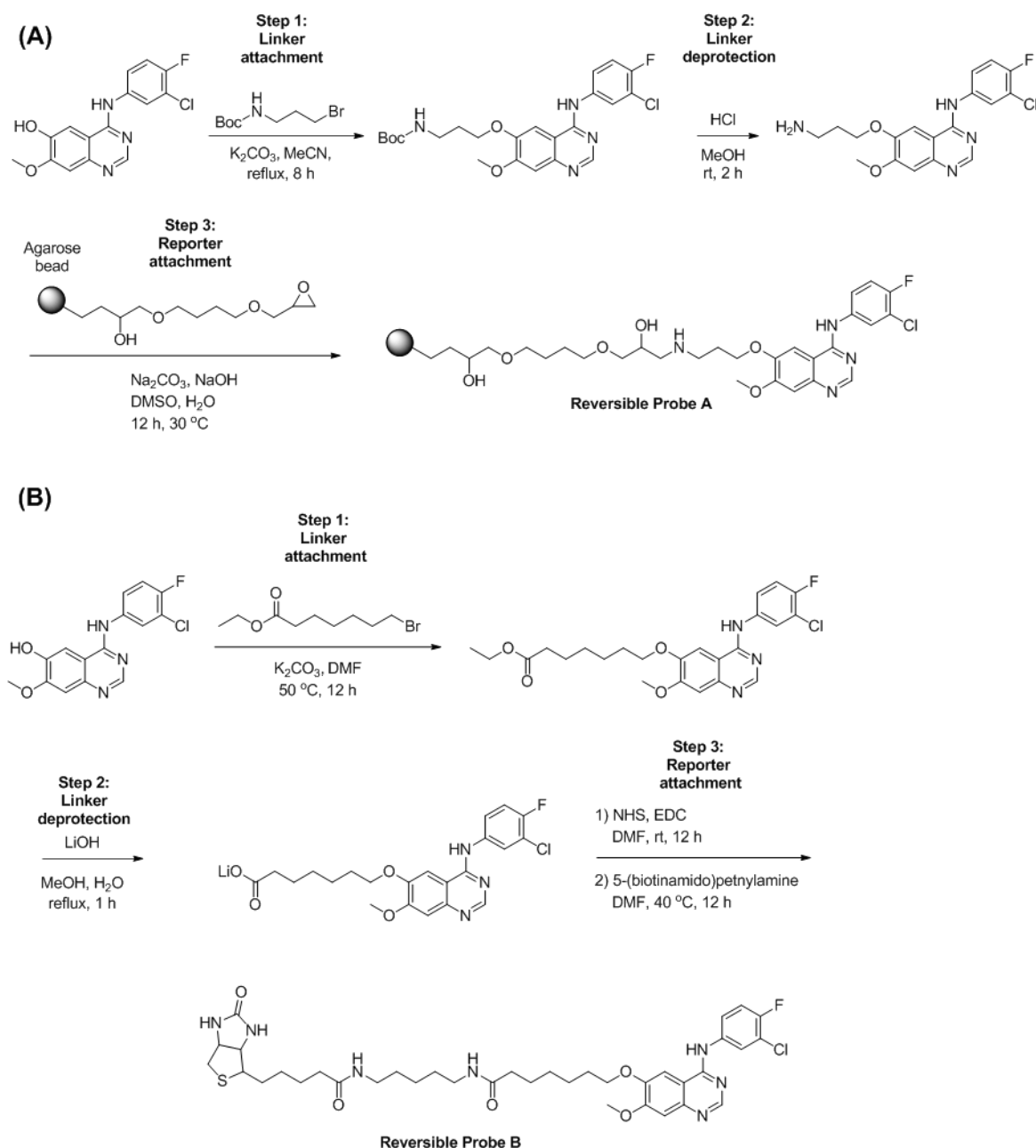
Scheme 2.1 Reported synthetic approach for representative reversible EGFR probes.[8-9]

A typical EGFR probe synthesis follows a modular approach that allows the integration of the required probe modules into the quinazoline core (**Scheme 2.1** and **2.2**).[8-11] Probe synthesis usually starts with the attachment of a linker to the quinazoline core. These linkers are based on bifunctional molecules containing two terminal reactive groups: one of them mediates the

covalent bond formation with quinazoline, while the other is protected and used in a following reaction for attaching the reporter. In reversible probes, this first step usually involves the aliphatic substitution of a halide-linker with a 6-hydroxy-quinazoline derivative (**Scheme 2.1a**).[8-9] Since reversible probes are mostly limited to the formation of kinobeads, they usually present a biotin group (**Scheme 2.1b**)[9] or are directly attached to agarose beads through a covalent bond (**Scheme 2.1a**).[8]
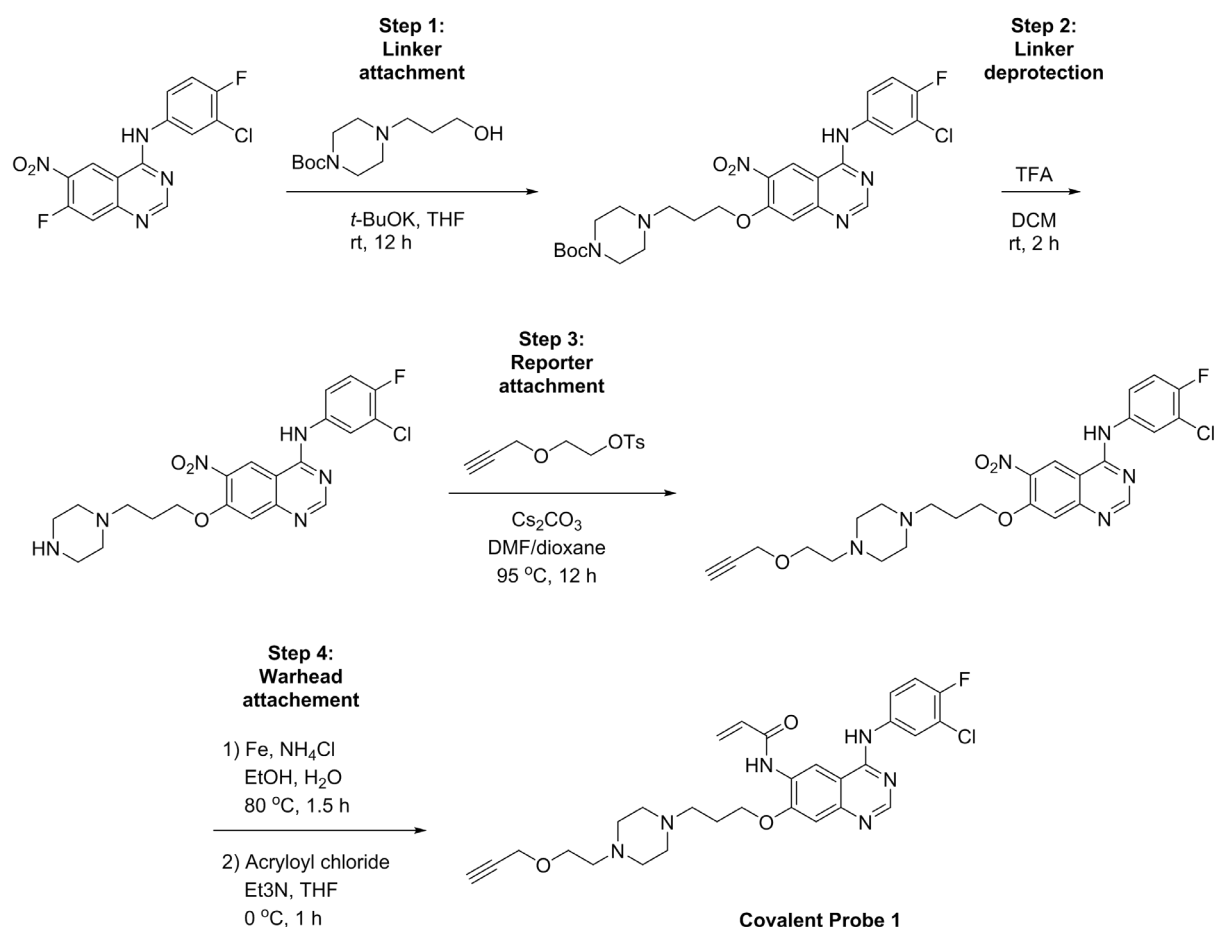
Scheme 2.2 Reported synthetic approach for covalent EGFR Probes.[10-11]

For covalent probes (**Scheme 2.2**), linker attachment usually involves the nucleophilic aromatic substation of a 7-fluro-6-nitroquinazoline derivative.[10-11] This is followed by attachment of the reporter group after deprotection of the linker's second reactive group. The second reactive group is usually an amine or carboxylic acid which allows to attach the reporter group through a nucleophilic substitution or amidation reaction. In most covalent

probes the reporter group contains a terminal alkyne for further click chemistry. The small size of the click handle facilitates probe permeability into the cell, allowing its use for *in vivo* or *in situ* studies. A biotin or fluorophore group can then be attached in a bioorthogonal click reaction via the copper(I)-catalyzed azide alkyne cicloadition (CuAAC). Finally, in covalent probes the Michael acceptor required for C797 labelling is formed via reduction of the nitrate group at the C6 position to the corresponding amine followed by amidation to afford the acrylamide.

2.1.3 Limitations and opportunities in EGFR-targeting probes

Reversible probes have been used for off-target profiling of EGFR inhibitors, but are limited to in vitro studies.[8, 12] Cysteine-targeting covalent probes have expanded this scope to *in vivo* and *in situ* studies. However, the nature of the electrophilic warhead creates off-target labelling in a concentration and time-dependent manner.[13-14] This defines a restricted window of selectivity above which indiscriminate labelling becomes dominant. Recently, Shindo et al demonstrated that it is possible to chemically tune the reactivity of the electrophile to increase this window of selectivity.[15] Using a chloroacetamide warhead instead of the common Michael acceptors, they obtained a probe with improved selectivity for EGFR at a wider range of concentrations.

Beyond cysteine labelling, lysine-targeting covalent inhibitors with reduced warhead-mediated off reactivity have recently been reported.[16-17] This improved target selectivity allowed the selective labelling of the conserved catalytic lysine in PI3K[17] by combining highly specific affinity drugs with chemically tuned electrophiles. Despite the promise of improved selectivity, the design of lysine-targeting inhibitors is challenging due to the lower nucleophilicity and more ubiquitous presence of lysine compared with cysteine.[18-19] Recent efforts have been devoted to proteome-wide identification of targetable lysines with chemical probes.[20-21]

Covalent modification of non-cysteine residues in EGFR has been previously reported. Using an N-hydroxyphthalamide (NHP) ester bifunctional crosslinker in live HeLa cells, Chavez et al identified a direct intramolecular crosslinking interaction between lysines K739 and K728, both located in the N-lobe of the kinase domain, by mass spectrometry (MS).[22] In another case, a bioconjugate, consisting of a DMAP-based catalyst covalently attached to an EGFR-directed antibody, was used to covalently attach a fluorophore in the extracellular domain of EFGFR in live A431 cells.[23] Peptide mass fingerprinting allowed the identification of residues S282, S474 and K476 in in subdomains II and III as the sites of protein labelling by this catalyst bioconjugate. These cases support the hypothesis that it might be possible to develop lysine-targeting EGFR-directed probes as a complementary approach to the probes currently in use. The design and synthesis of such probes are discussed next.

2.2 RESULTS AND DISCUSSION

2.2.1 Design of EGFR-targeting probe

To determine the viability of the covalent modification of a specific lysine residue in EGFR through the modification of current quinazoline-based chemical probes, the different lysine residues were first identified in reported crystal structures of second-generation inhibitors covalently bound to WT and T790M EGFR (**Fig. 2.2** and **Table 2.1**). Five lysine residues, K716, K728, K745, K846 and K852, were identified within a radius of 15 Å from the C7 atom of the quinazoline core. Residues K846 and K852 are located in a direction opposite to the ATP binding site, and therefore are not suitable targets. K745 is the conserved catalytic lysine, analogous to the residue previously labeled by the selective lysine-targeting PI3K inhibitor.[17] This residue is within a close distance (4 Å) of the aniline ring. While the attachment of an electrophile at the aniline ring might afford a K745-targeting EGFR inhibitor, this strategy is not suitable for a probe due the localization of the aniline in a back pocket. The remaining two residues, K716 and K728, are located in the vicinity of the

quinazoline core. Of these last two residues, K728 is closest to the quinazoline, with an averaged distance of 9 Å to the C7 position. This is the same residue previously labeled by Chavez et al.[22] Therefore, K728 was selected as a potential target residue for the new probe.
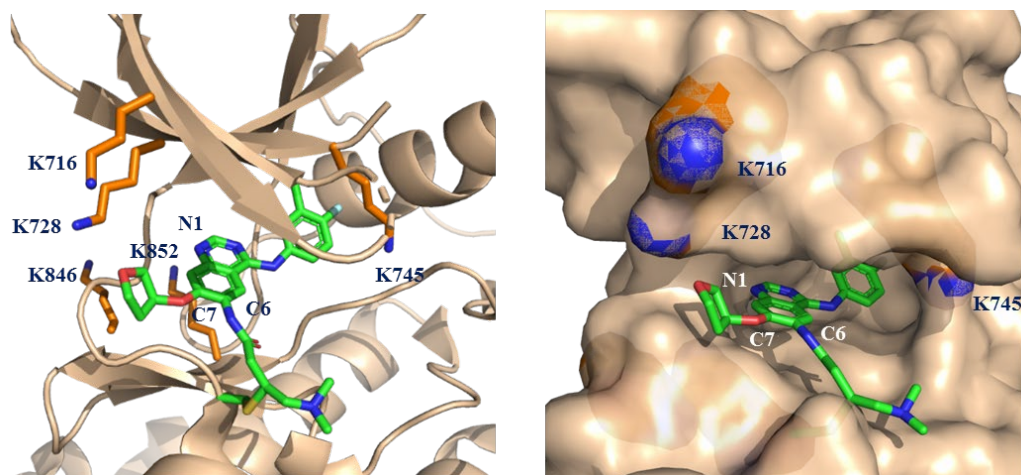
Fig. 2.2 Mapping of lysine residues in close proximity to 4-anilinoquinazoline inhibitors covalently bound to EGFR (PDB 4G5J).[6]

Table 2.1 Distances of lysine residues to C7 position in 4-anilinoquinazoline inhibitors covalently bound to EGFR through C797.

UniProt ID	K716	K728	K745	K846	K852
2J5E	15.2	9.8	10.8	14.3	13.0
2J5F	13.5	9.5	10.9	14.3	12.8
4G5J	11.5	8.1	9.6	15.1	13.7
4G5P	13.1	8.6	13.8	12.6	12.9
Average	13.3	9.0	11.3	14.1	13.1

The distance between the target lysine K728 and the quinazoline core is too large for the direct attachment of a reactive electrophile, but could be introduced through a linker of appropriate length (**Fig. 2.3**). Similar probes in which a non-cysteine-targeting warhead is attached through a linker to the drug of interest have achieved good selectivity mediated by a proximity effect that emerges from the combined features of the drug, linker and warhead.[24-26]

The solvent-exposed localization of K728 at the rim of the ATP binding site, along with the conformational flexibility of EGFR, suggest that such probes might be able to form a covalent bond not only with K728 but potentially with EGFR interacting proteins. If lysine targeting does allow the capture of EGFR interactors through protein protein interaction (PPI), a dual cysteine-lysine targeting probe could be envisaged next that would provide structural evidence of PPI. In a dual cysteine-labelling probe, both solvent exposed positions C6 and C7 of the quinazoline would be substituted, requiring appendage of the click handle to the linker. With this information, the blueprint for two potential probes was constructed: a monofunctional lysine-targeting (**Fig. 2.3, P1**) or a bifunctional cysteine-lysine targeting (**Fig 2.3, P2**) click probe.

Probe P1 **Probe P2**
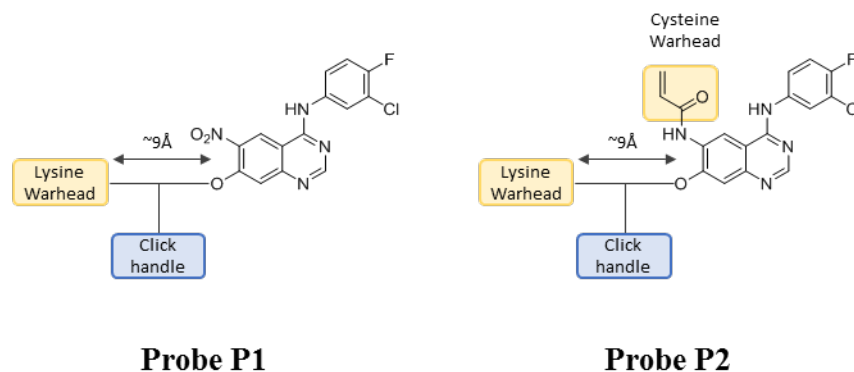
Fig. 2.3 Blueprints of potential lysine-targeting EGFR probes

For proposing actual structures according to the blueprint, an adequate linker with the right length was designed next. The use of branched aminoacids (e.g. lysine, glutamic acid) and peptide coupling reactions is a common strategy for the synthesis of trifunctional linkers for biological probes.[27] This approach is useful for the construction of linkers with higher length, facilitating the connection of large building blocks. Modelling of potential linkers based on these amino acids resulted in sterically hindered and rigid linkers (data not shown). To overcome this problem, smaller trifunctional chiral building blocks were considered for potential etherification reactions to resemble the PEG linkers normally used in chemical probes (**Figure 2.4**). A glycerol-type precursor, (*S*)-(+)-1,2-isopropylideneglycerol, was

identified as a suitable starting material for this purpose. For the warhead, an N-hydroxysuccinimide (NHS) ester was used as a starting point, since NHS esters are commonly used as crosslinking reagents and have known proteome-wide reactivity.[20] A typical terminal alkyne was also chosen as the click handle for the tandem attachment of a reporter through CuAAC-type click chemistry. With a chemical structure for all the different modules, the structures of probes **P1** and **P2** was finally determined (**Fig. 2.4**).
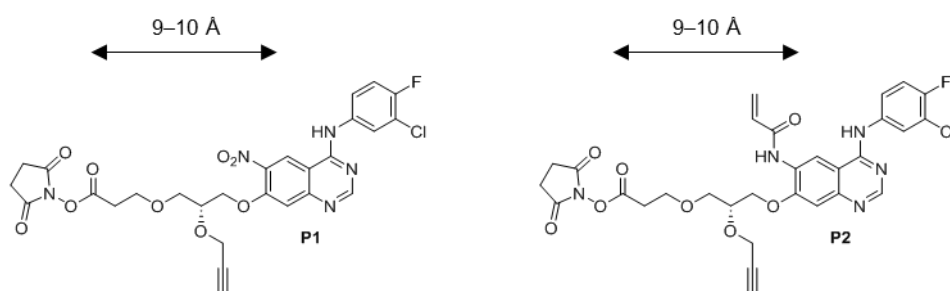
Fig. 2.4 Structure of proposed probes **P1** and **P2**

2.2.2 Synthesis of EGFR-targeting probes

2.2.2.1 Retrosynthetic analysis

The introduction of two reactive groups (acrylamide for cysteine labelling and NHS ester for lysine labelling) into the same molecule is very challenging, and a careful analysis of the stability of both reactive groups indicated that the NHS ester, while susceptible to hydrolysis, should be introduced first as the acrylamide group is reactive toward nucleophiles. This leads to the first target probe, **P1**, from which an acrylamide group could be introduced to access **P2** in the future (**Scheme 2.3**). The introduction of the activated NHS ester can be performed with carbodiimide coupling between NHS and the corresponding carboxylic acid **S1**. Following the synthetic methods reported for EGFR covalent probes,[10-11] **S1** could be obtained from the nucleophilic aromatic substitution of 4-[(3-chloro-4-fluorophenyl)amino]-7-fluoro-6-nitroquinazoline (hereafter **Q1**), the core structure of Gefitinib, Afatinib and Dacomitinib, with linker **S2**. The basic conditions required for aromatic substitution with

linker **S2** also demand a protecting group for the carboxylic acid. Protecting groups requiring cleavage under basic or reducing conditions are incompatible with the synthetic route; thus, the acid labile *tert*-butyl ester protecting group was proposed for linker **S2**.
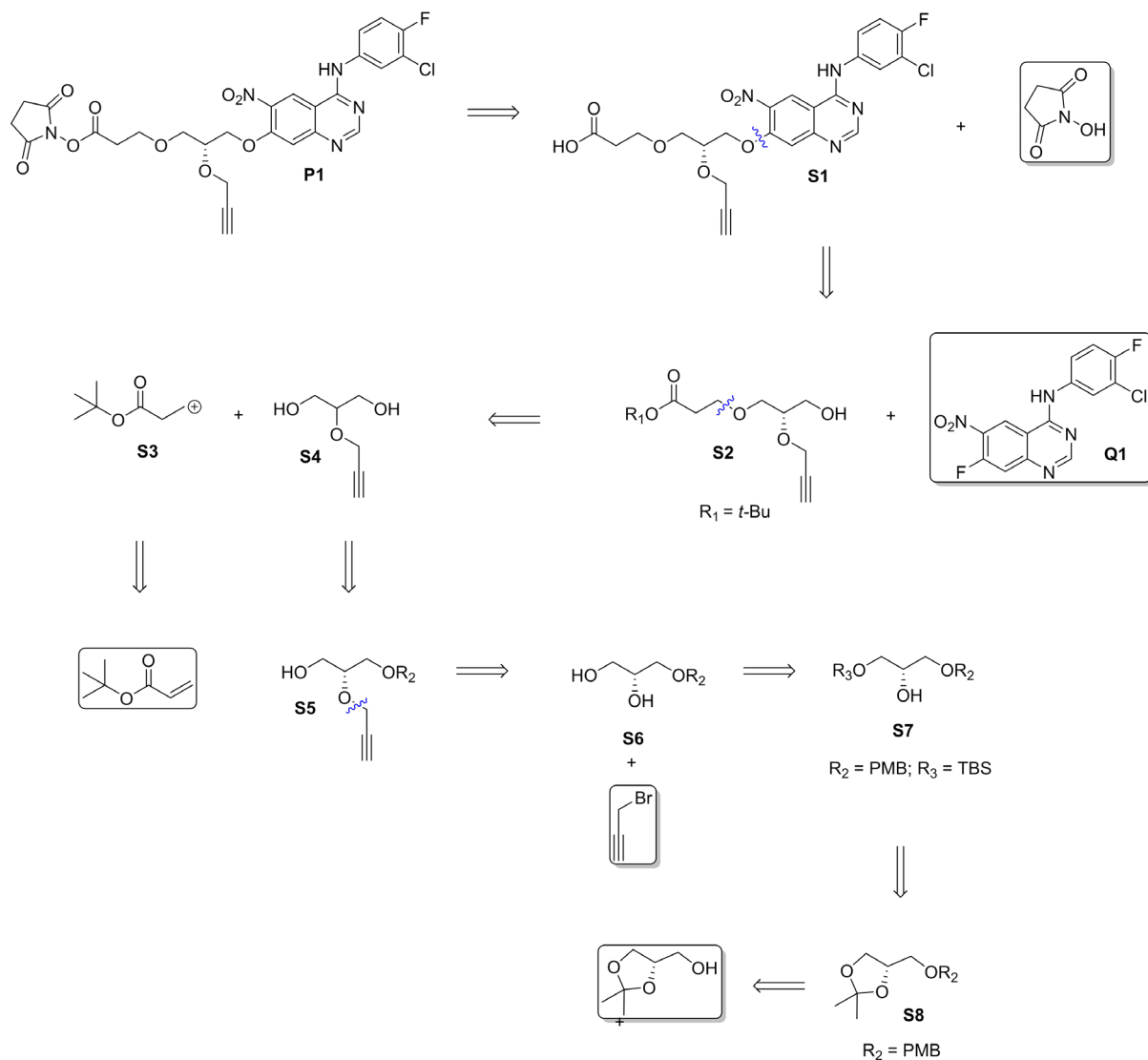
Scheme 2.3 Retrosynthetic analysis of EGFR Probe **P1.**

Retrosynthetic analysis of **S2** led to the synthons **S3** and **S4**. The electrophilic centre β to the carbonyl in **S6** led us to propose an *oxa*-Michael addition. Here, the stability of the *tert*-butyl ester to basic conditions is advantageous to avoid hydrolysis that would otherwise result in a mixture of by-products. A protecting group (R_2) is also required for one of the primary alcohols of the diol **S4** in order to maintain the chirality of the linker, leading to synthons **S5**. The propargyloxy group in **S5** can be introduced through the Williamson etherification

between propargyl bromide and the corresponding secondary alcohol in synthons **S6**, provided the primary alcohol is first masked with a protecting group (R_3) orthogonal to R_2 as in the synthons **S7**. The R_2 protecting group for the primary alcohol must be stable to the basic conditions of the Williamson etherification; after etherification, R_2 and cannot be cleaved under reducing (incompatible with alkyne) or acidic (incompatible with *tert*-butyl ester) conditions. For R_3, only stability to bases and non-reducing conditions for cleavage are required. Both PMB and TBS were identified as suitable protecting groups for R_2 and R_3, since they can be removed under non-reducing neutral conditions with DDQ and TBAF respectively. TBS-PMB protected glycerol **S7** can in turn be obtained from the key building block (*S*)-(+)-1,2-isopropylideneglycerol after acidic hydrolysis of the R_2-protected acetonide **S8**, which requires R_2 to be stable to acidic conditions. Thus, PMB was assigned for R_2 and TBS for R_3. This completed the retrosynthetic analysis leading us to the identification of suitable starting materials.

As outline in the next sections in the forward sense, the synthesis of the monofunctional lysine-targeting probe **P1** required the synthetic investigations on several linkers, **L1A-B**, **L2A-B**; and **L3A-B** with varying degree of complexity (**Fig. 2.5**).
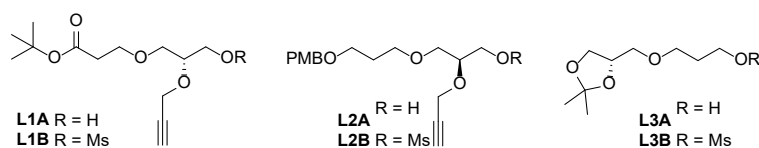
Fig. 2.5 Structure of linkers obtained during synthesis of probe **P1**

2.2.2.2 Synthesis of linker L1A

The first attempt towards probe **P1** focused on the synthesis of the fully functional linker **L1A**, which contains both the protected ester for lysine-labelling and alkyne click handle reporter (**Fig. 2.5** and **Scheme 2.4**). Compounds **1.2** and **1.3** had been previously reported in

the literature,[28-29] and here were obtained in high yield using slightly modified procedures. PMB protection of (S)-(+)-1,2-isopropylideneglycerol (**1.1**) was possible both using NaH in THF and under PTC conditions, and the latter resulted in a cleaner product and higher yield. Acidic hydrolysis of the corresponding acetal **1.2** was completed in a couple of hours using either 80% AcOH or TsOH in MeOH. Selective TBS protection of the primary alcohol in diol **1.3** was possible using TBSTOTf at −80 °C. The Williamson etherification of alcohol **1.4** with propargyl bromide afforded the corresponding compound **1.5**, which was then subjected to TBS deprotection with TBAF leading to the primary alcohol **1.6** with good yields.
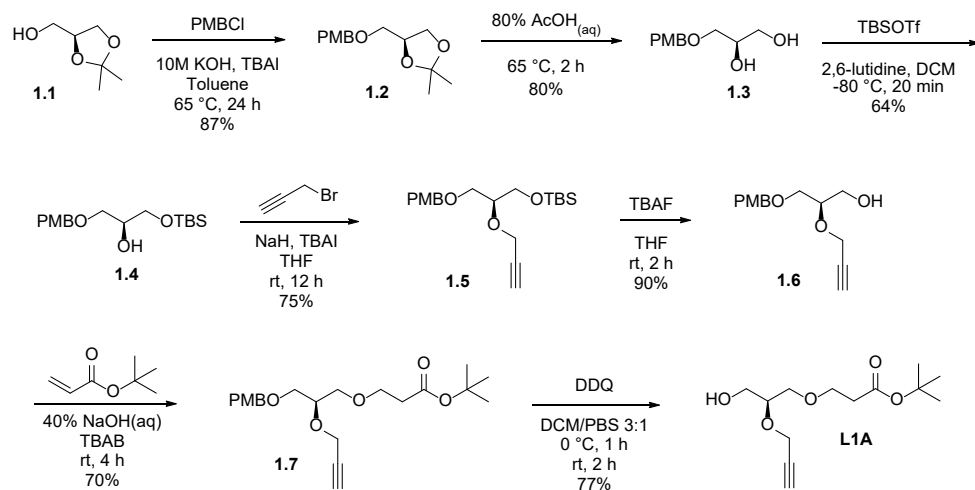
Scheme 2.4 Synthesis of linker **L1.A**

To continue extending the linker with the introduction of the protected ester fragment to alcohol **1.6**, the Williamson etherification was attempted with *tert*-butyl 3-bromopropanoate under several conditions (various solvents, bases, temperatures and concentrations), however without success (**Fig. 2.6**). Using the primary alcohol **1.1** as a control substrate, *O*-alkylation was observed with *tert*-butyl 2-bromoacetate, but not *tert*-butyl 3-bromopropionate or *tert*-butyl 4-bromobutyrate. This suggested that in *tert*-butyl 3-bromopropionate and analogous esters of higher length, the halogenated carbon is not active enough towards nucleophilic substitution with a primary alcohol. Since the carbon required for the *O*-alkylation is in the β position with respect to the carbonyl, a different strategy was attempted using electrophilic

addition instead of nucleophilic substitution. *Oxa*-Michael addition of alcohol **1.6** to *tert*-butyl acrylate afforded the corresponding product **1.7** with good yield. Finally, PMB deprotection of **1.7** with DDQ led to linker **L1A** with 16% yield over seven steps.
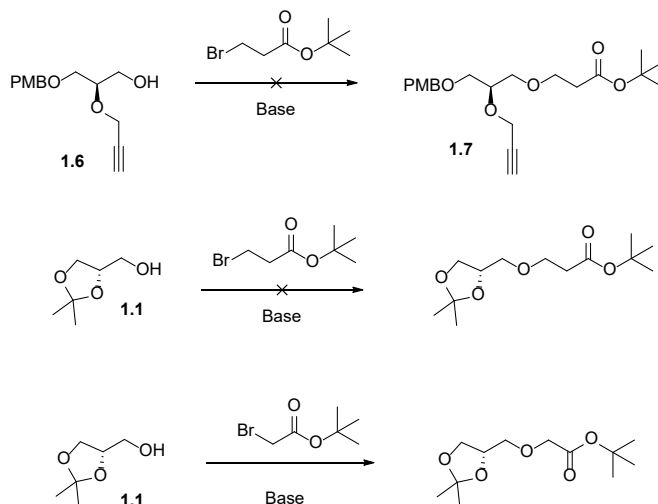
Fig. 2.6 Attempts to *O*-alkylation of primary alcohols via nucleophilic substitution of *tert*-butyl halo esters

2.2.2.3 First approach to probe P1: S$_N$Ar

Although the S$_N$Ar substitution of 6-nitro-7-fluoroquinazolines with alcohols of small molecular weight is reported in the literature,[10-11] the initial attempts to attach linker **L1A** to the 7-fluoroquinazoline **Q1** using NaH in THF at room temperature were unsuccessful (**Table 2.2, entry 1**). Increasing the reaction temperature to reflux in THF resulted in the degradation of both the linker and quinazoline. Given the low reactivity of *tert*-butyl esters towards substitution, it was hypothesized that the presence if this group in the linker might be impeding the S$_N$Ar reaction. To test this possibility, nucleophilic aromatic substitution of **Q1** was attempted with other linkers that lack the *tert*-butyl ester group: the starting alcohol **1.1**, the reported linker **L3A** and the linker intermediate **1.6** (**Table 2.2, entries 2-4**). In the absence of the *tert*-butyl ester group, the nucleophilic aromatic substitution of **Q1** proceeded at room temperature with good to moderate yields. As the structural complexity of the linker

increased, its nucleophilicity decreased resulting in lower yields. Therefore, the synthesis of probe **P1** switched to alternative routes based on the tested linkers (**Scheme 2.3**).

Table 2.2 Nucleophilic aromatic substitution of quinazoline **Q1** with linkers of varying complexity

Entry	Linker	Conditions	Yield
1	**L1A**	60 °C, 16 h	No reaction
2	**1.1**	rt, 4 h	78%
3	**L3A**	rt, 12 h	37%
4	**1.6**	rt, 4 h	32%

Compounds **1.1** and **L3A** were subjected to S_NAr reactions before the introduction of the alkyne reporter and activated ester (**Scheme 2.5**). The corresponding S_NAr products **1.10** and **1.14** were subjected to acetal hydrolysis followed by TBS monoprotection of the resulting diols using procedures analogous to those used for linker **L1A**. Initial attempts to the *O*-propargylation of compounds **1.12** and **1.16** using an excess of NaH and propargyl bromide in THF resulted in a complex mixture of several products. LC-MS analysis of these products suggested the presence of *N*-alkylated, *O*-alkylated and bi-alkylated compounds. The regioselectivity of the alkylation was then controlled by limiting the equivalents of NaH and propargyl bromide. Under these conditions with substrate **1.12**, *O*-alkylation was not observed by ^{1}H NMR (**Fig. 2.7a, b**).
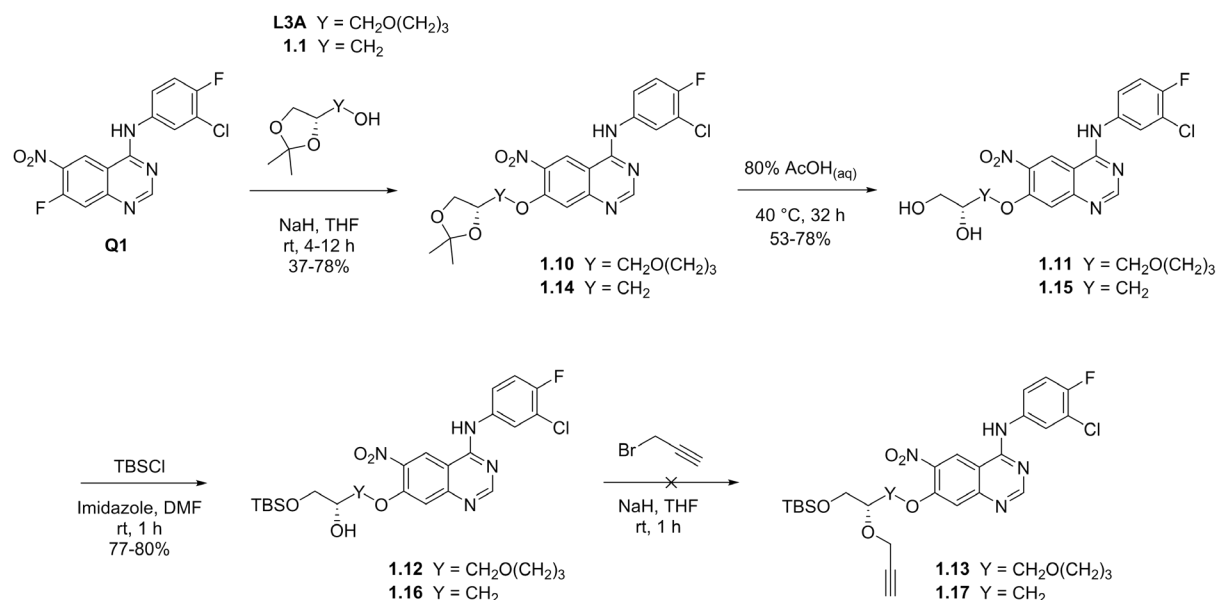
Scheme 2.5 Alternative approach to probe **P1** with **1.1** and **L3A**

Due to the higher nucleophilicity of nitrogen compared with oxygen, *N*-propargylation was expected. However, ^{1}H and HSQC NMR suggested the aromatic substitution at the C2 position of the quinazoline core with a vinyl group had taken place (**1.12X**), as indicated by the loss of signal corresponding to H-2 in the quinazoline core and the characteristic signals of the vinyl group (**Fig. 2.7b and 2.8**). A similar pattern of vinyl substitution was also observed when compound **1.14** was treated under similar conditions to provide only *N*-alkylation (**1.14X, Fig. 2.7c**). A mechanistic investigation of this unexpected result is outside the scope for this project, but represents an interesting topic for further research.

After the problems encountered in the *O*-propargylation of quinazoline-alcohol substrates, a second alternative was investigated using the alkyne-containing linker **1.6** (**Scheme 2.4** and **Table 2.2, entry 4**). Despite the moderate yield obtained after S$_N$Ar of **Q1**, the use of linker **1.6** represented a plausible solution to the problem of introducing the alkyne reporter.
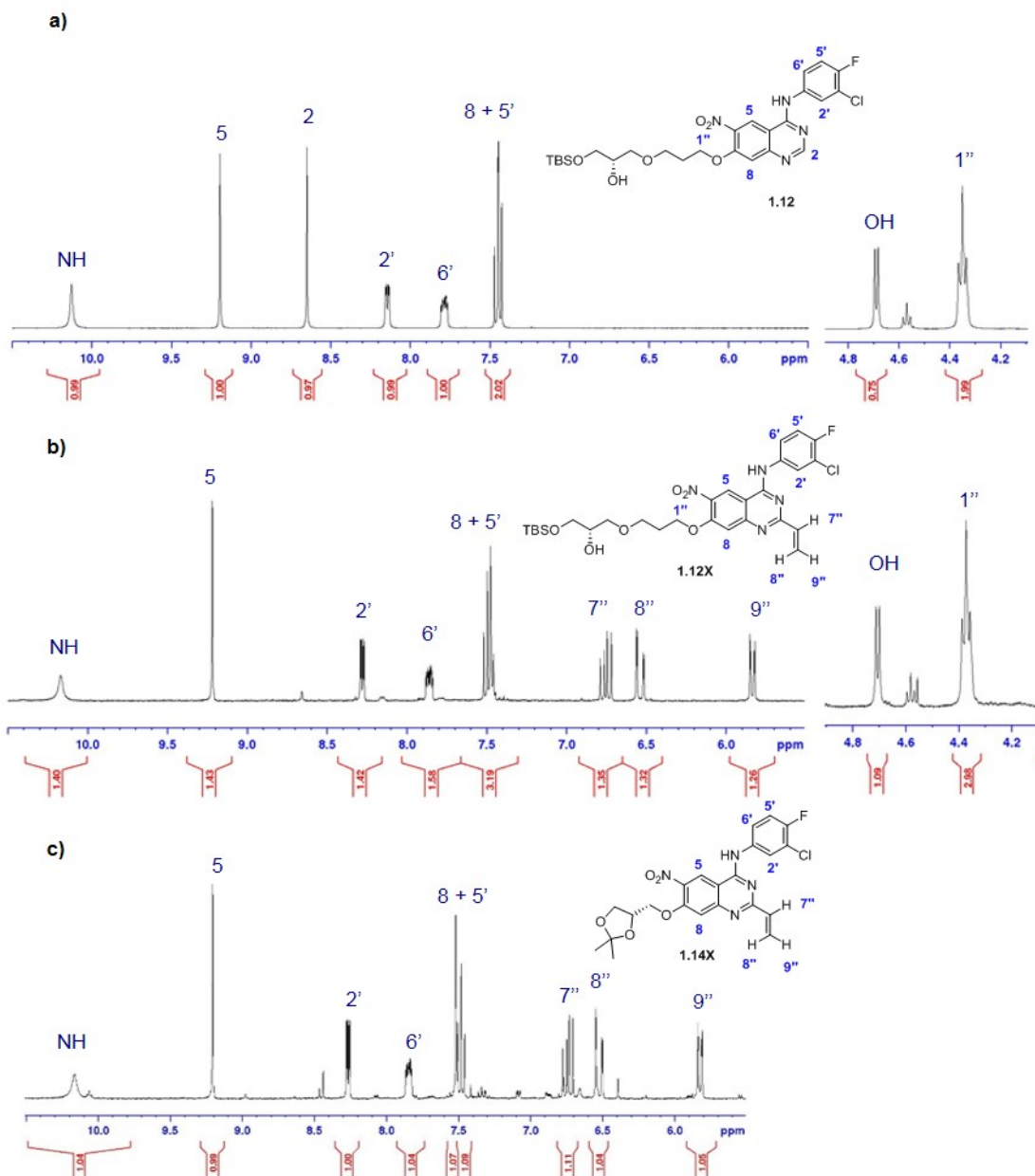
Fig. 2.7 ^{1}H NMR spectra of **1.12** and products obtained from **1.12** and **1.14** after treatment with NaH and propargyl bromide

The synthesis of the probe was then continued with the PMB deprotection of compound **1.18** with DDQ, which afforded the corresponding quinazoline-alcohol **1.19** with low yield (**Scheme 2.6**). The next step required the introduction of the activated ester and extension of linker length. Unlike its analogous substrate **1.6**, several attempts to the *oxa*-Michel addition of 4.6 with *tert*-butyl acrylate under various solvent/base conditions were unsuccessful. In general, the *O*-alkylation of an aliphatic alcohol precursor attached to the quinazoline core was very difficult and mostly ineffective.
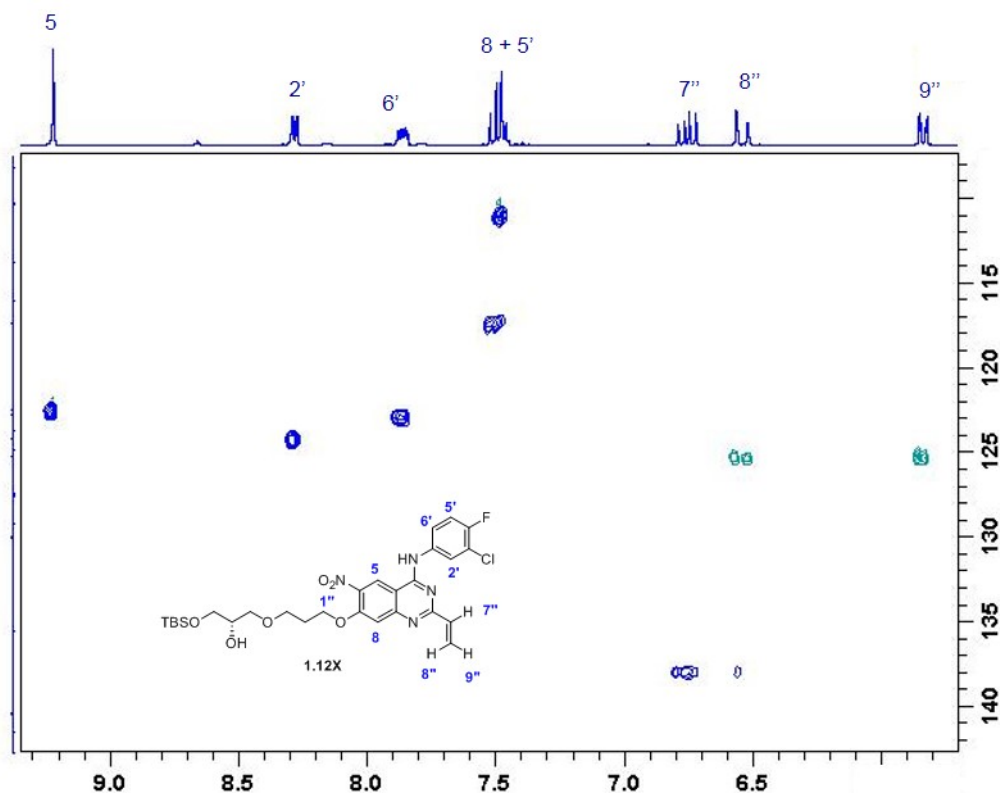
Fig. 2.8 HSQC NMR spectrum of product obtained from **1.12** after treatment with NaH and propargyl bromide
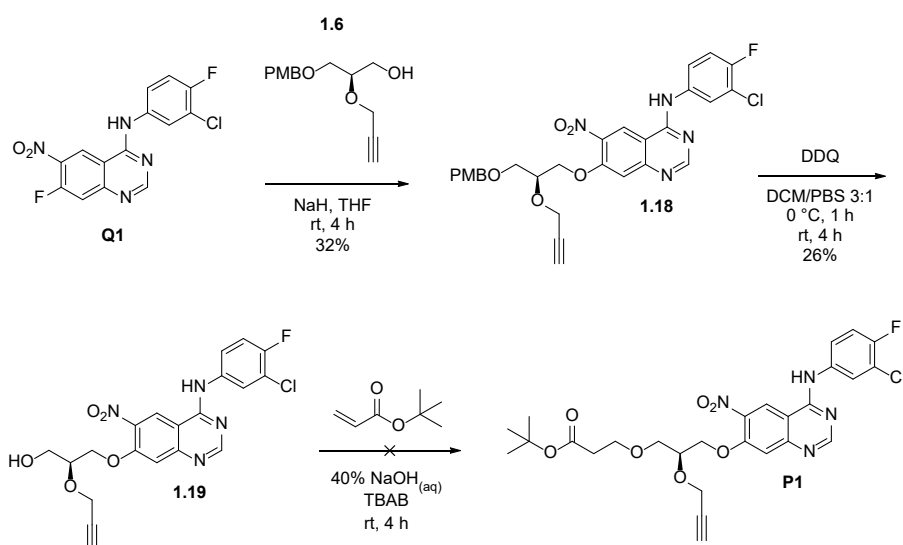
Scheme 2.6 Second alternative approach to the synthesis of probe **P1**

2.2.2.4 Synthesis of linker L2A

After various failed attempts to complete the probe synthesis through aliphatic *O*-alkylation of a quinazoline precursor, the advantage of having a linker with the required final length and the alkyne reporter group became obvious. Although the *ter*-butyl ester group was

incompatible with the S_NAr conditions required for attachment of the linker, the synthesis of **1.18** suggested that it might be possible to use a full-length linker in which the *tert*-butyl ester is replaced by a protected alcohol. This would facilitate the aromatic substitution, after which the alcohol could then be deprotected and oxidized to the required activated ester. Thus, linker **L2A** was proposed as another alternative to the synthesis of the probe.

Scheme 2.7 Synthesis of linker **L2A**
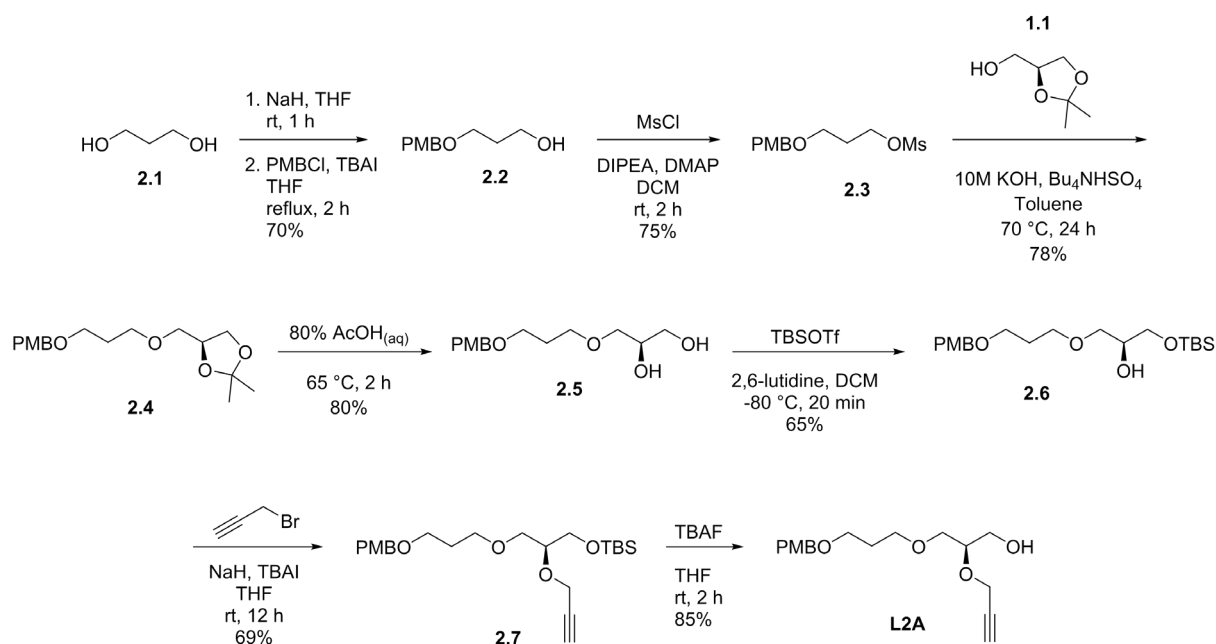
The synthesis of linker **L2A** started with the protection and activation of 1,3-propanodiol (**2.1**, **Scheme 2.7**). Compounds **2.2** and **2.3** had been previously reported in the literature,[30] and their synthesis was reproduced without problems with good yields. The mesylate **2.3** was then used to alkylate the starting alcohol **1.1** under PTC conditions, leading to compound **2.4** with good yield. Intermediate **2.4** was then subjected to a series of protection/deprotection and alkylation steps analogous to those previously used for the synthesis of linker **L1A**. Linker **L2A** was eventually obtained with 12% yield over seven steps.

As expected, the reactivity of linker **L2A** towards the aromatic substation of quinazoline **Q1** was intermediate between that of linker **L1A** and compound **1.6** (**Table 2.3**, **entry 2**).
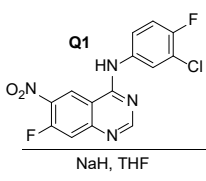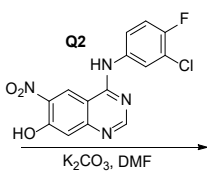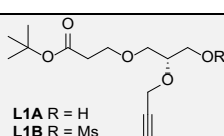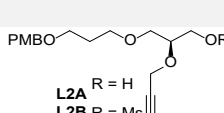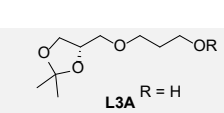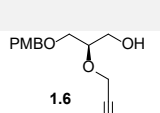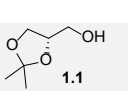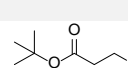
Although the reaction did not proceed at room temperature as with **1.6**, no degradation of starting materials was observed after refluxing in THF for a couple of hours. Under the different reaction conditions screened (solvent and concentration), the S_NAr reaction always reached an equilibrium at about 20% conversion, and the corresponding product was isolated in 9% yield. As such another alternative to the synthesis of probe **P1** was investigated.

2.2.2.5 Second approach to probe P1: S_N2

After the various failed attempts to complete the synthesis of the probe through S_NAr of 7-fluoroquinazoline **Q1**, an alternative S_N2 approach was identified using the 7-hydroxyquinazoline derivative **Q2** and the mesylate linkers **L1–3B** to shift the linker attachment from aromatic to aliphatic substitution (**Table 2.3**). To test the stability of the *tert*-butyl ester to the milder S_N2 conditions compared with S_NAr, 7-hydroxyquinazoline **Q2** was reacted with *tert*-butyl 3-bromopropionate using K_2CO_3 as a base in DMF. The corresponding product was obtained in good yield after heating for 24 hours at °C (**Table 2.3, entry 6**), supporting the use of aliphatic substation as a better alternative. Although the yields for S_N2 also decreased with increased linker complexity, it was possible to attach all the previously synthesized linkers with better yields compared with nucleophilic aromatic substitution.

The synthesis of probe **P1** was eventually completed continuing from linker **L1B** (**Scheme 2.8**). After mesylation of **L1A** and substitution with **Q2**, the *tert*-butyl ester in the corresponding product **1.8** was hydrolysed with TFA quantitatively. Finally, EDC coupling of NHS with the carboxylic acid **1.9** afforded the probe **P1** with a yield of 1% over 11 steps (starting from **1.1**). The control inhibitor **I1**, in which the propargyl reporter tag is replaced by a methyl group, was obtained in a similar way with an overall yield of 2% over 11 steps (**Fig. 2.9**).

Table 2.3 Comparison of aliphatic and aromatic substitution of
6-nitro-4-anilinoquinazolines with different linkers

Entry	Linker	S_NAr		S_N2	
		Conditions	Yield	Conditions	Yield
		NaH, THF		K₂CO₃, DMF	
1	**L1A** R = H **L1B** R = Ms	60 °C, 16 h	-	90 °C, 24 h	22%
2	**L2A** R = H **L2B** R = Ms	60 °C, 16 h	9%	90 °C, 24 h	50%
3	**L3A** R = H **L3B** R = Ms	rt, 12 h	37%	90 °C, 4 h	61%
4	**1.6**	rt, 4 h	32%	-	-
5	**1.1**	rt, 4 h	78%	-	-
6		-	-	90 °C, 24 h	84%
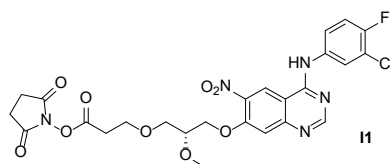
Fig. 2.9 Structure of the control NHS inhibitor **I1**
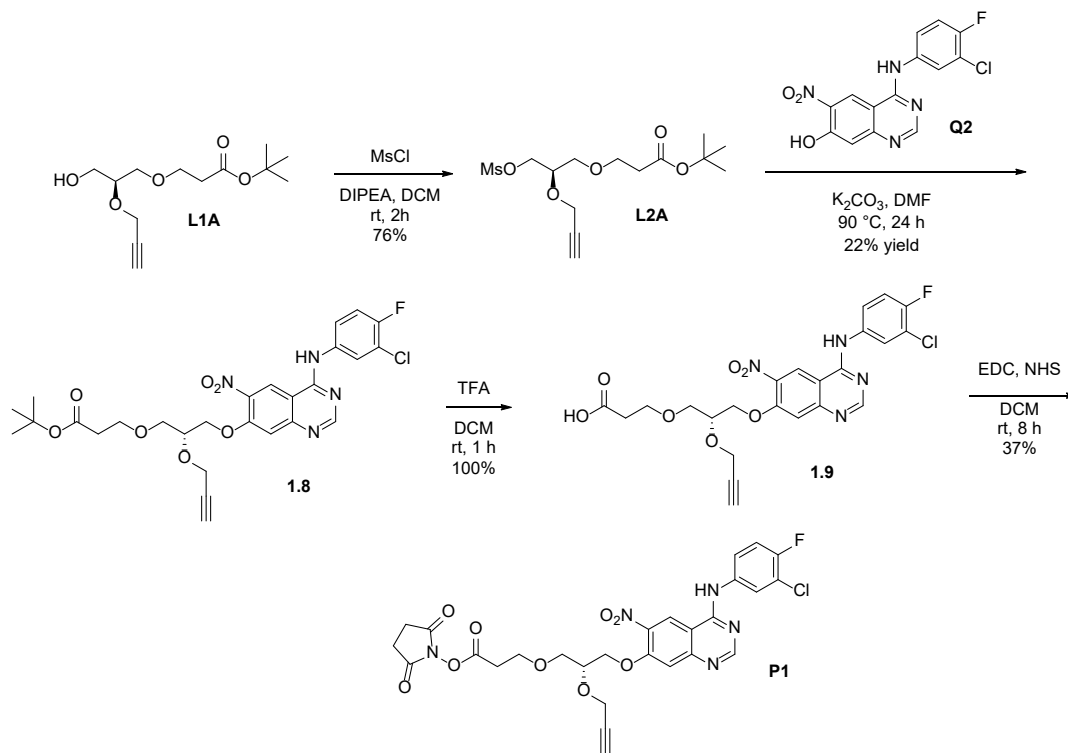
Scheme 2.8 Synthesis of probe **P1** continuing from linker **L1A**

2.2.3 Conclusions

The synthesis and structure-activity relationship of EGFR inhibitors and probes based on the 4-anilinoquinazoline is extensively reported in the literature. Chemical probes are usually obtained through modifications at the C6 and C7 solvent-exposed positions. In this work, the novel lysine-targeting EGFR probe **P1** was designed and synthesized. In the quest for the synthesis of these probes, several linkers of various complexities were also obtained, and these can be used for the modification of other drug targets. In general, linkers with greater complexity required higher energy and afforded lower yields for the attachment to the corresponding 4-anilinoquinazoline core. The nucleophilic aliphatic substitution of quinazoline **Q2** represented a better route compared with the nucleophilic aromatic substitution of quinazoline **Q1.** The modular strategy here reported can be used in the future to access novel EGFR-directed probes (such as **P2**) rapidly. A summary of the whole synthetic route for probe **P1**, inhibitor **I1**, as well as the latest intermediate towards probe synthesis with alternative routes is shown in the following schemes (**Scheme 2.9–2.12**).
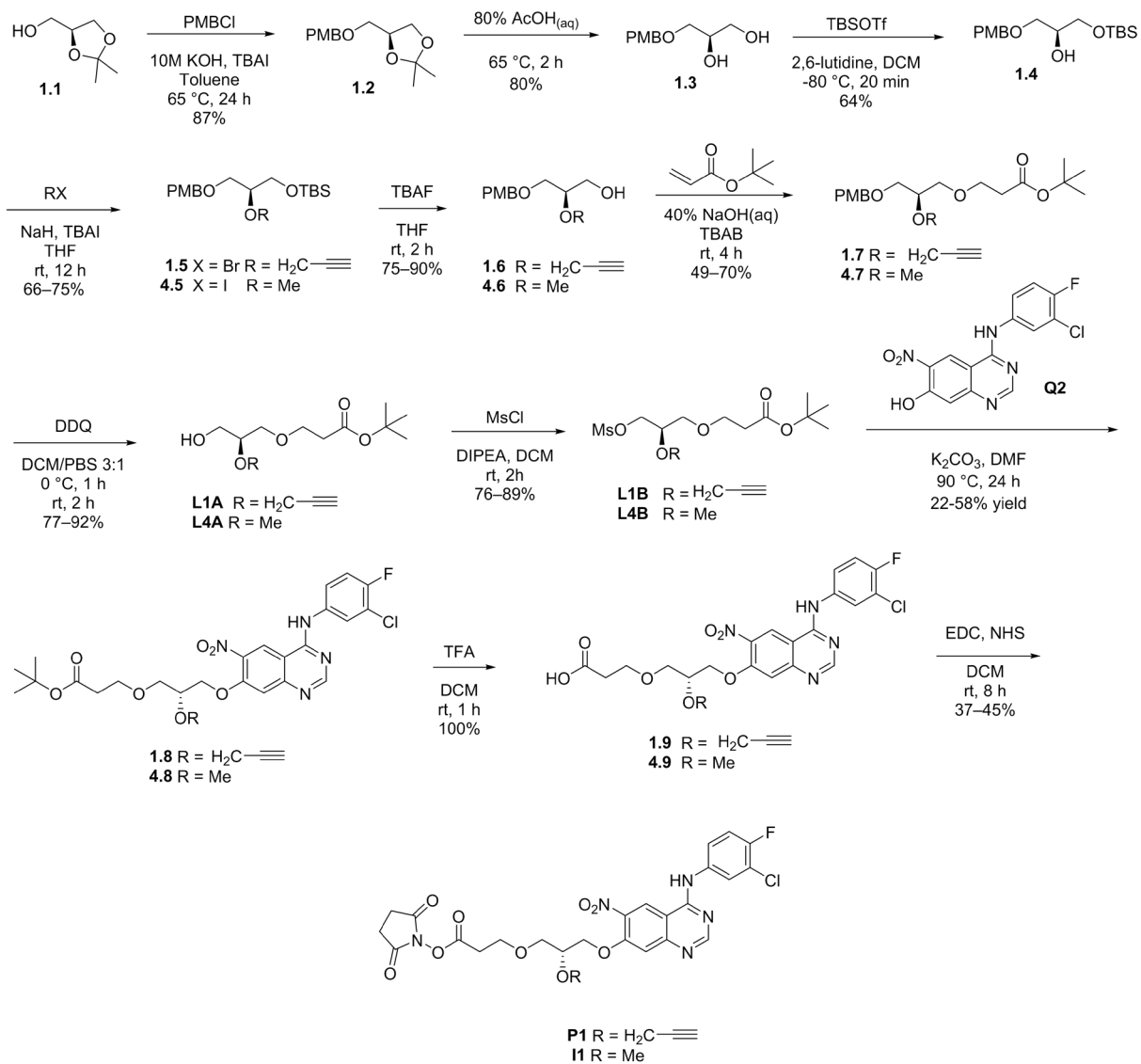
Scheme 2.9 Complete synthesis of probe **P1** and inhibitor **I1** via **L1B** linker
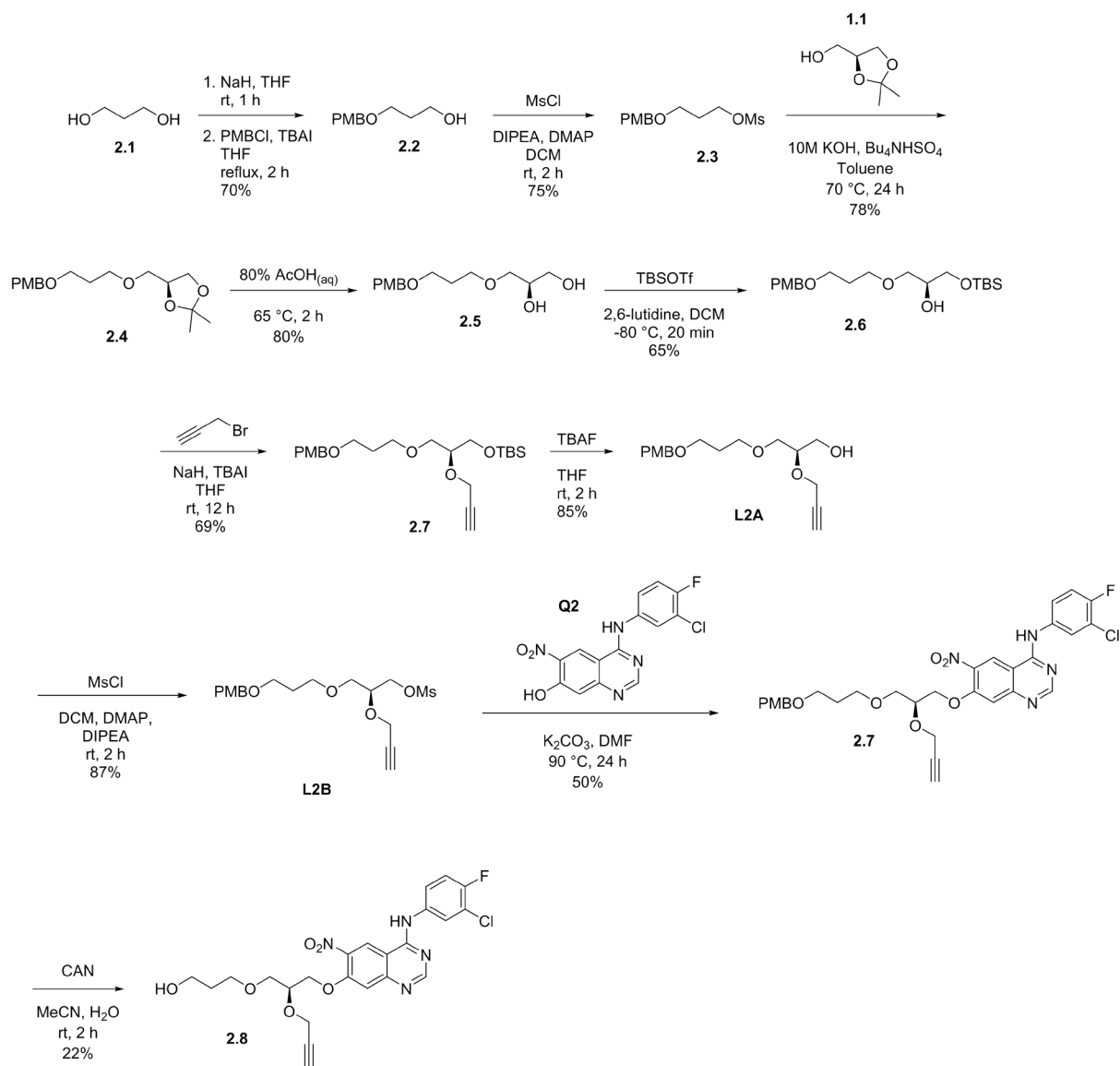
Scheme 2.10 Approach to **P1** via **L2B** linker
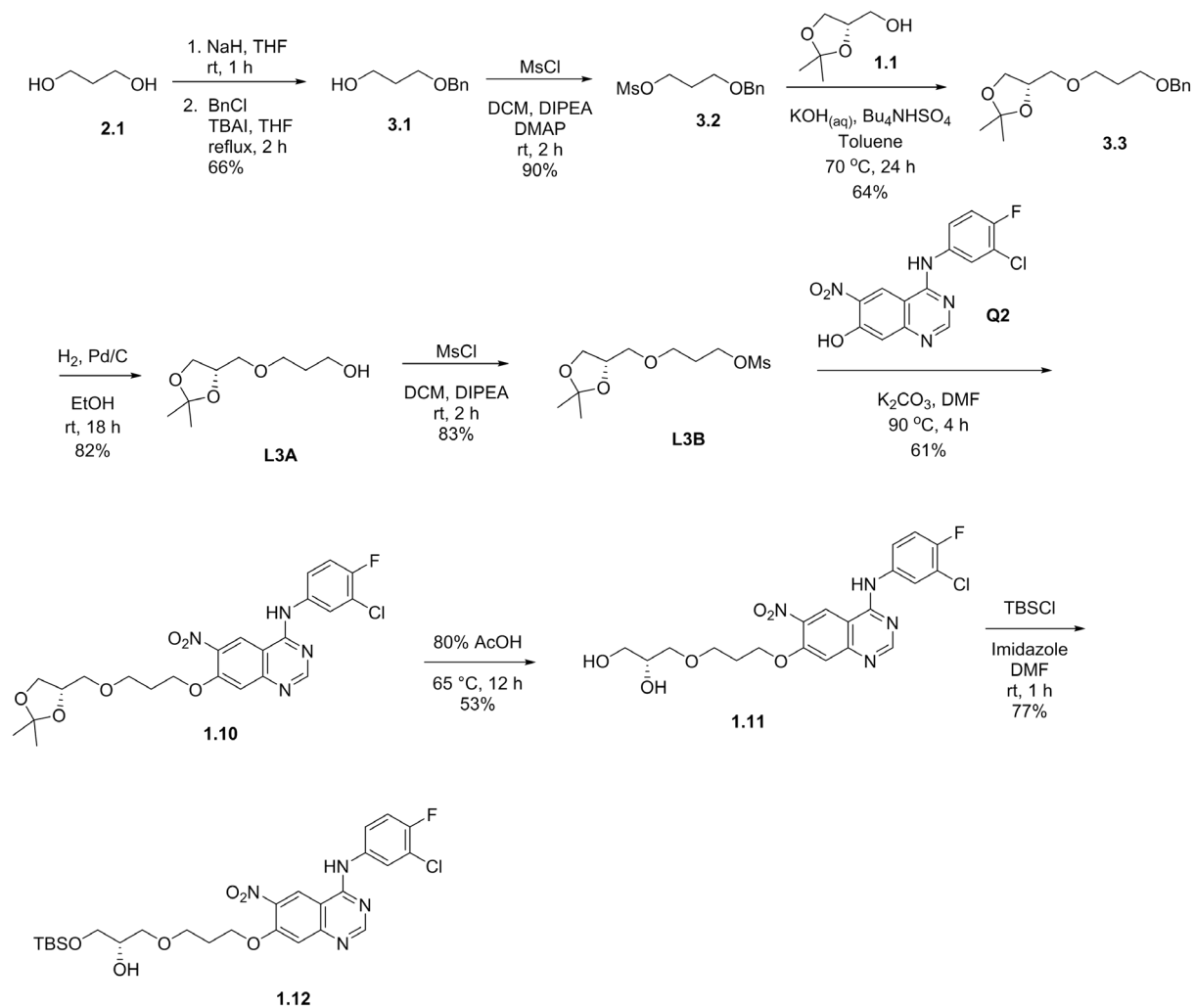
Scheme 2.11 Approach to **P1** via **L3B** linker
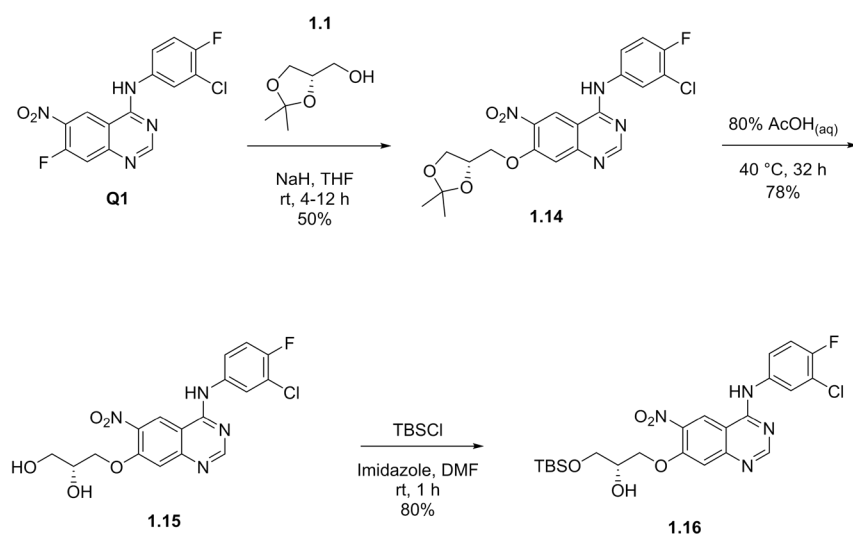
Scheme 2.12 Approach to **P1** via direct attachment of **1.1**

2.3 EXPERIMENTAL

2.3.1 General information

Unless otherwise specified, all reagents were used as received from common commercial suppliers. All solvents were either anhydrous or HPLC grade and used without further purification unless otherwise specified. DCM was distilled from CaH_2 under nitrogen and stored over 4Å-MS. THF was freshly distilled from sodium/benzophenone under nitrogen. All reactions were magnetically stirred and monitored by thin-layer chromatography using Merck silica gel 60 F_{254} pre-coated plates (0.25 mm) revealed by UV or $KMnO_4$ solution. Flash column chromatography was performed on Merck silica gel 60 (0.015–0.040 mm). ^1H and ^{13}C NMR spectra were recorded on a Bruker Avance DPX 400 MHz at 25 °C. Chemical shifts were reported in ppm using the solvent residual signal (DMSO-d_6 δ_H = 2.49, δ_C = 39.5; CDCl$_3$ δ_H = 7.24, δ_C = 77.0; Acetone-d_6 δ_H = 2.05, δ_C = 29.8) as an internal reference. Low-resolution mass-spectrometry was conducted on an Agilent 6130 quadrupole LC-MS system with an electrospray ionization source (ESI) using a Phenomenex Gemini C18 column (2.0 x 150 mm, particle size 3 µm); the mobile phase consisted of a gradient of 40–90% acetonitrile in water with 0.05% formic acid over 6 minutes (flow rate 0.2 mL/min, column temperature 25 °C) and the spectra were acquire in positive mode, scanning over the *m/z* range of 100–1000.

2.3.2 Synthesis and characterization

2.3.2.1 Synthesis of linker L1A, probe P1 and inhibitor I1

*(S)-4-[((4-methoxybenzyl)oxy)methyl]-2,2-dimethyl-1,3-dioxolane (**1.2**).*[28]
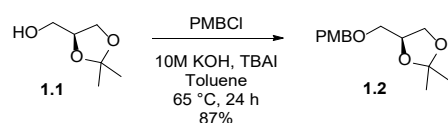
To a stirred solution of **1.1** (2.0 g, 15.1 mmol, 1 equiv) in toluene (150 ml) was added TBAI (1.4 g, 3.8 mmol, 0.25 equiv), 10 M KOH (150 mL) and PMBCl (2.6 g, 16.6 mmol, 1.1

equiv). The mixture was heated at 65 °C for 24 h. The mixture was then diluted with water; the organic and aqueous layers were separated, and the aqueous layer was extracted with EtOAc. The combined organic layers were washed with saturated NH_4Cl until neutral, dried over Na_2SO_4, filtered and rotary-evaporated to afford **1.2** as a yellow liquid without further purification as previously reported (3.3 g, 87%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.19 (m, 2H), 6.80 (m, 2H), 4.43 (m, 2H), 4.21 (quin, $J = 6.0$ Hz, 1H), 3.97, (dd, $J = 8.2, 6.4$ Hz, 1H), 3.73 (s, 3H), 3.65 (dd, $J = 8.2, 6.4$ Hz, 1H), 3.45 (dd, $J = 9.8, 5.7$ Hz, 1H), 3.36 (dd, $J = 9.8, 5.7$ Hz, 1H), 1.34 (s, 3H), 1.29 (S, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.4, 130.6, 130.1, 113.9, 109.5, 74.9, 73.3, 70.9, 67.1, 55.4, 26.9, 25.5.

*(R)-3-[(4-methoxybenzyl)oxy]propane-1,2-diol (**1.3**).[29]*
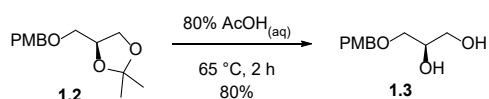
A solution of **1.2** (3.8 g, 15.0 mmol, 1 equiv) in 80% AcOH$_{(aq)}$ (29 mL) was heated at 65 °C for 2 hours. The solvent was then rotary-evaporated at 50 °C. The crude mixture was diluted with toluene, and rotary-evaporated three times to remove any residual AcOH. The isolated crude mixture was purified by flash column chromatography (60–80% EtOAc/hexanes) to afford **1.3** as a white solid as previously reported (2.6 g, 80%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.27 (m, 2H), 6.90 (m, 2H), 4.50 (s, 2H), 3.89 (m, 1H), 3.82 (s, 3H), 3.70 (m, 1H), 3.64 (m, 1H), 3.54 (m, 1H), 2.85 (br, 1H), 2.41 (br, 1H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.5, 129.9, 129.6, 114.0, 73.4, 71.6, 70.7, 64.2, 55.4.

*(S)-1-[(tert-butyldimethylsilyl)oxy]-3-[(4-methoxybenzyl)oxy]propan-2-ol (**1.4**).[31]*
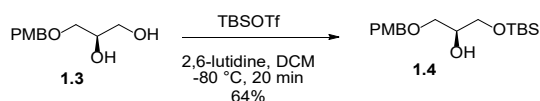
To a solution of **1.3** (1.6 g, 7.3 mmol, 1 equiv) in DCM (73 mL) at −78°C (dry ice/*i*PrOH bath) was added 2,6-lutidine (1.7 mL, 14.6 mmol, 2 equiv) and TBSOTf (1.6 mL, 7.0 mmol,

0.95 equiv) dropwise. The mixture was stirred at −78 °C for 20 min, and then allowed to warm up to room temperature. The solvent was rotary-evaporated and the crude mixture purified by flash column chromatography (15% EtOAc/Hexanes) to afford **1.4** as a colourless liquid (1.5 g, 64%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.19 (m, 2H), 6.82 (m, 2H), 4.42 (s, 2H), 3.77 (m, 1H), 3.74 (s, 3H), 3.58 (m, 2H), 3.43 (m, 2H), 0.83 (s, 9H), 0.00 (s, 6H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.4, 130.3, 129.5, 113.9, 73.2, 70.8, 64.1, 55.4, 26.0, 18.4, −5.3.

*(S)-tert-butyl[3-[(4-methoxybenzyl)oxy]-2-(prop-2-yn-1-yloxy)propoxy]dimethylsilane (**1.5**)*
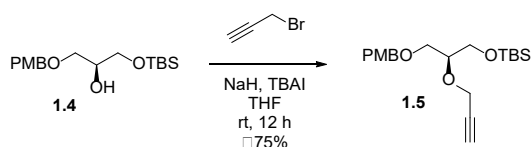
To a suspension of NaH (60% w/w, 0.76 g, 19.0 mmol, 5 equiv) in THF (19 mL) at 0 °C was added **1.4** (1.24 g, 3.8 mmol, 1 equiv) dropwise. After stirring at room temperature for 1 h, TBAI (0.12 g, 0.38 mmol, 0.1 equiv) was added followed by propargyl bromide (0.90 g, 7.6 mmol, 2 equiv) and the mixture was stirred at room temperature overnight. The reaction was quenched by dropwise addition of saturated NH$_4$Cl. THF was rotary-evaporated and the obtained crude mixture diluted with DCM and washed with brine. The organic layer was dried over Na$_2$SO$_4$, filtered and rotary-evaporated. The isolated crude mixture was purified by flash column chromatography (7% EtOAc/hexanes) to afford **1.5** as a colourless liquid (1.0 g, 75%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.20 (m, 2H), 6.81 (m, 2H), 4.42 (s, 2H), 4.27 (m, 2H), 3.75 (s, 3H), 3.70 (m, 1H), 3.64 (m, 2H), 3.54 (dd, J = 10.2, 4.2 Hz), 3.46 (dd, J = 10.2, 5.4 Hz), 0.83 (s, 9H), 0.00 (s, 6H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.3, 130.5, 129.4, 113.9, 80.5, 78.5, 74.1, 73.2, 69.8, 63.1, 57.8, 55.4, 26.0, 18.4, -5.2, −5.3.

*(S)-tert-butyl(2-methoxy-3-((4-methoxybenzyl)oxy)propoxy)dimethylsilane (**4.5**)*
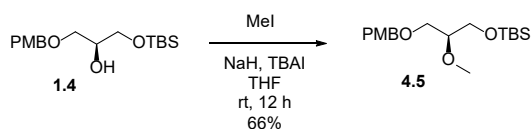
A procedure similar to **1.5** was followed with methyl iodide to afford **4.5** as a colourless liquid (541 mg, 66%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.26 (m, 2H), 6.87 (m, 2H), 4.48 (m, 2H), 3.80 (s, 3H), 3.67 (d, J = 5.2 Hz, 2H), 3.57 (dd, J = 10.0, 4.3 Hz, 1H), 3.49–3.45 (m, 4H), 3.40 (m, 1H), 0.88 (s, 9H), 0.05 (d, J = 1.8 Hz, 6H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.3, 130.6, 129.4, 113.9, 81.2, 73.2, 69.5, 62.6, 58.2, 55.4, 26.0, 18.4, −5.2, −5.2.

*(R)-3-[(4-methoxybenzyl)oxy]-2-(prop-2-yn-1-yloxy)propan-1-ol (**1.6**)*
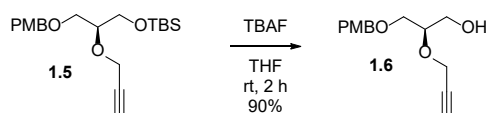
To **1.5** (1.0 g, 2.7mmol, 1 equiv) was added 1M TBAF in THF (7.8 mL, 7.8 mmol, 2.9 equiv) and the mixture was stirred at room temperature for 2 hours. The solvent was then rotary-evaporated and the crude mixture purified by flash column chromatography (40% EtOAc/hexanes) to afford **1.6** as a yellow liquid (618 mg, 90%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.27 (m, 2H), 6.90 (m, 2H), 4.50 (m, 2H), 4.34 (m, 2H), 3.83 (m, 4H), 3.78 (m, 1H), 3.68 (m, 1H), 3.66 (m, 2H), 2.47 (t, J= 2.4 Hz), 2.06 (br, 1H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.4, 130.0, 129.5, 114.0, 80.1, 78.0, 74.7, 73.3, 69.9, 62.9, 57.6, 55.4; LC-MS (ESI): [M+Na]$^+$, *m/z* calculated for C$_{14}$H$_{18}$O$_4$ 273.12, found 273.1.

*(R)-2-methoxy-3-((4-methoxybenzyl)oxy)propan-1-ol (**4.6**)*
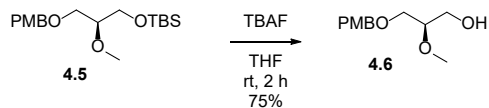
A procedure similar to **1.6** was followed with **4.5** to afford **4.6** as a yellow liquid (555 mg, 75%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.25 (m, 2H), 6.88 (m, 2H), 4.48 (m, 2H), 3.80 (s, 3H), 3.75 (m, 1H), 3.64 (m, 1H), 3.56 (ddd, J = 12.8, 10.0, 5.1 Hz), 3.45 (m, 4H); ^{13}C NMR (100 MHz, CDCl$_3$) δ (159.4, 130.1, 129.5, 114.0, 80.1, 73.4, 69.5, 62.7, 57.9, 55.4.

*tert-butyl (R)-3-{3-[(4-methoxybenzyl)oxy]-2-(prop-2-yn-1-yloxy)propoxy)propanoate (**1.7**)*
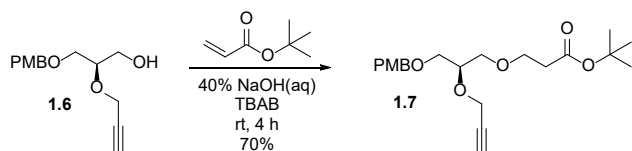
A mixture **1.6** (527 mg, 2.1 mmol, 1 equiv) and 40% NaOH (1.7 mL, 16.8 mmol, 8 equiv) in a round bottom flask was purged with nitrogen and stirred at room temperature for 1 h. A solution of TBAB (340 mg, 1.05 mmol, 0.5 equiv) in water (4.2 mL) was then added and the mixture cooled down to 0 °C. *Tert*-butyl acrylate (0.32 mL, 2.2 mmol, 1.05 equiv) was added dropwise and the mixture allowed to warm to room temperature and stirred for 4 h. The mixture was then diluted with water and the product extracted with Et$_2$O. The combined organic layers were washed with saturated NH$_4$Cl and brine, dried over Mg$_2$SO$_4$ and rotary-evaporated to afford **1.7** as a yellow liquid without further purification (554 mg, 70%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.27 (m, 2H), 6.89 (m, 2H), 4.49 (s, 2H), 4.34 (d, J = 2.4 Hz, 2H), 3.89 (m, 1H), 3.82 (s, 3H), 3.70 (t, J = 6.5 Hz, 2H), 3.58 (m, 4H), 2.49 (t, J = 6.5 Hz, 2H), 2.42 (t, J = 2.4 Hz, 1H), 1.46 (s, 9H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 171.0, 159.3, 130.4, 129.4, 113.9, 80.6, 80.4, 76.7, 74.2, 73.1, 71.3, 70.0, 67.3, 57.8, 55.4, 36.4, 28.2.

*tert-butyl (R)-3-(2-methoxy-3-((4-methoxybenzyl)oxy)propoxy)propanoate (**4.7**)*
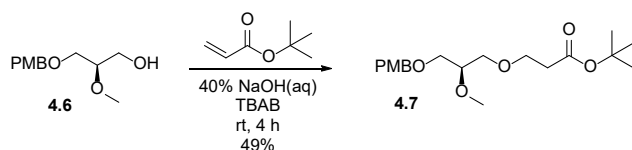
A procedure similar to **1.7** was followed with **4.6** to afford **4.7** as a yellow liquid (413 mg, 49%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.25 (m, 2H), 6.87 (m, 2H), 4.47 (s, 2H), 3.80 (s, 3H), 3.68 (t, J = 6.5 Hz, 2H), 3.52 (m, 5H), 3.44 (s, 3H), 2.47 (t, J = 6.5 Hz, 2H), 1.44 (s, 9H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 171.0, 159.3, 130.5, 129.4, 113.9, 80.6, 79.4, 73.2, 70.8, 69.6, 67.3, 58.1, 55.4, 36.4, 28.2.

*tert-butyl (S)-3-[3-hydroxy-2-(prop-2-yn-1-yloxy)propoxy]propanoate (**L1A**)*
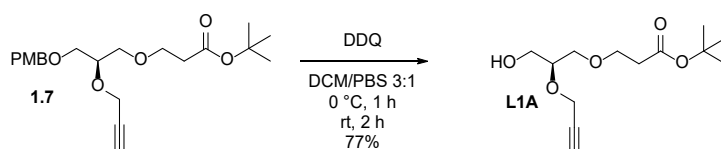
1.7 (534 mg, 1.41 mmol, 1 equiv) was dissolved in 3:1 DCM/PBS (35 mL) and cooled to 0 °C. DDQ (640 mg, 2.82 mmol, 2 equiv) was then added and the mixture, stirred at 0 °C for 1 h and then at room temperature for 2 h. The reaction mixture was then diluted with DCM and washed with saturated NaHCO$_3$. The aqueous layer was extracted with DCM and the organic layers combined, washed with brine, dried over Mg$_2$SO$_4$, filtered and rotary-evaporated. The isolated crude mixture was purified by flash column chromatography (20–40% EtOAc/hexanes) to afford **L1A** as a yellow liquid (280 mg, 77%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 4.29 (m, 2H), 3.78–3.58 (m, 7H), 2.48 (td, J = 6.2, 1.5 Hz, 2H), 2.44 (t, J = 2.4 Hz, 1H), 1.44 (s, 9H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 171.1, 80.9, 80.1, 77.7, 74.7, 71.0, 67.3, 62.5, 57.5, 36.3, 28.2.

*tert-butyl (S)-3-(3-hydroxy-2-methoxypropoxy)propanoate (**L4A**)*
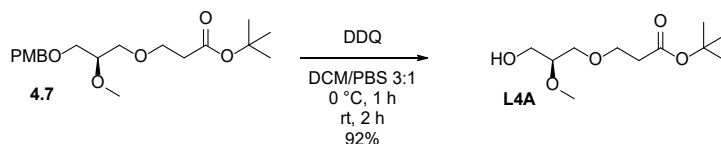
A procedure similar to **L1A** was followed with **4.7** to afford **L4A** as a yellow liquid (173 mg, 92%). ^{1}H NMR (400 MHz, CDCl$_3$) δ3.70 (m, 3H), 3.63 (dd, J = 11.7, 5.3 Hz, 1H), 3.58 (dd, J = 4.9, 3.6 Hz, 2H), 3.44 (s, 3H), 3.41 (m, 1H), 2.49 (td, J = 6.2, 2.0 Hz, 2H), 1.45 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ171.2, 80.9, 79.9, 70.5, 67.3, 62.3, 57.9, 36.3, 28.2.

*tert-butyl (R)-3-(3-((methylsulfonyl)oxy)-2-(prop-2-yn-1-yloxy)propoxy)propanoate (**L1B**)*
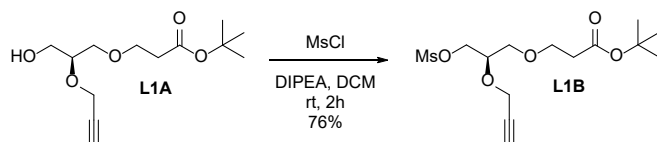
To a solution of **L1A** (172 mg, 0.67 mmol, 1 equiv) in DCM (3.3 mL) at 0 °C was added DIPEA (0.15 mL, 0.87 mmol, 1.3 equiv) and MsCl (62 µL, 0.8 mmol, 1.2 equiv). After stirring at room temperature for 2 h, the reaction was quenched adding cold water. The aqueous layer was extracted with DCM and the combined organic layers washed with brine. The organic layer was then dried over Na$_2$SO$_4$, filtered and concentrated under reduced pressure to afford **L1B** as a brown liquid without further purification (170 mg, 76% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 4.38 (dd, J = 11.0, 3.7 Hz, 1H), 4.31 – 4.27 (m, 3H), 3.96 (m, 1H), 3.7 (m, 2H), 3.59 (t, J = 5.4 Hz, 2H), 3.06 (s, 3H), 2.49 – 2.45 (m, 3H), 1.45 (s, 9H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 170.8, 80.9, 79.4, 75.2, 75.1, 69.5, 69.4, 67.4, 57.9, 37.7, 36.3, 28.2.

*tert-butyl (R)-3-(2-methoxy-3-((methylsulfonyl)oxy)propoxy)propanoate (**L4B**)*
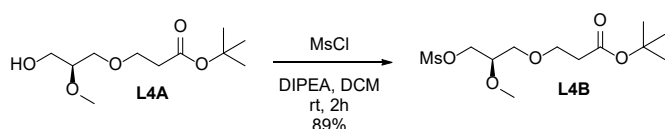
A procedure similar to **L1B** was followed with **L4A** to afford **L4B** as a yellow liquid (193 mg, 89%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 4.37 (dd, J = 11.0, 3.3 Hz, 1H), 4.25 (dd, J = 11.0, 5.3 Hz, 1H), 3.70 (t, J = 6.3 Hz, 2H), 3.59 (m, 1H), 3.55 (m, 2H), 3.46 (s, 3H), 3.04 (s, 3H), 2.48 (t, J = 6.3 Hz, 2H), 1.45 (s, 9H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 170.9, 80.9, 78.0.

*4-[(3-chloro-4-fluorophenyl)amino]-6-nitroquinazolin-7-ol[32] (**Q2**)*
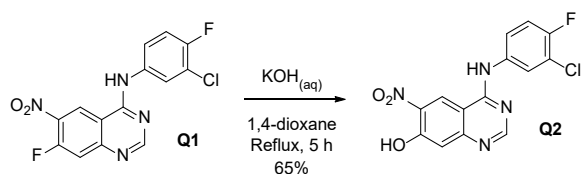
To a solution of *N*-(3-chloro-4-fluorophenyl)-7-fluoro-6-nitroquinazolin-4-amine (**Q1**, 2.0 g, 5.94 mmol, 1 equiv) in 1,4-dioxane (30 mL) was added 10M KOH (7.4 mL) and the mixture was heated to reflux for 5 h. The mixture was then cooled to 0 °C in an ice bath and

neutralized with 1M HCl and then extracted with EtOAc. The combined organic layers were washed with brine, dried over Na$_2$SO$_4$, filtered and rotary-evaporated. The obtained crude mixture was purified by flash column chromatography (30–60% EtOAc/Hexanes) to afford **Q2** as an orange solid as previously reported (1.3 g, 65%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.92 (br), 10.15 (br), 9.21 (s, 1H), 8.58 (s, 1H), 8.14 (dd, J = 6.9, 2.6, Hz, 1H), 7.78 (m, 1H), 7.45 (t, J = 9.0 Hz, 1H), 7.23 (s, 1H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 157.9, 157.0, 154.8, 152.4, 138.2, 136.0, 123.9, 122.7, 122.2, 118.9, 116.6, 112.7.

tert-butyl (R)-3-(3-((4-((3-chloro-4-fluorophenyl)amino)-6-nitroquinazolin-7-yl)oxy)-2-(prop-2-yn-1-yloxy)propoxy)propanoate (1.8)
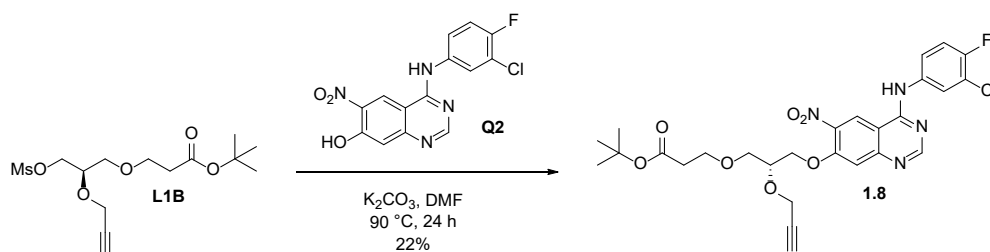
To a mixture of **Q2** (129 mg, 0.385 mmol, 1 equiv) and **L1B** (155 mg, 0.462 mmol, 1.2 equiv) in DMF (1.28 mL) was added K$_2$CO$_3$ (58 mg, 0.423 mmol, 1.1 equiv). The mixture was heated at 90 °C for 24 h, and then the solvent was removed at 90 °C under reduced pressure. The crude mixture was diluted with EtOAc and insoluble salts removed by filtration. The filtrate was rotary-evaporated and purified by flash column chromatography (35% EtOAc/Hx) to afford **1.8** as an orange solid (49 mg, 22%). ^{1}H NMR (400 MHz, acetone-d_6) δ 9.50 (br, 1H), 8.94 (s, 1H), 8.68 (S, 1H), 8.25 (dd, J = 6.7, 2.6 Hz, 1H), 7.83 (m, 1H), 7.48 (s, 1H), 7.34 (t, J = 9.0 Hz, 1H), 4.51 (dd, J = 10.3, 3.8 Hz, 1H), 4.41 (m, 3H), 4.16 (m, 1H), 3.77 – 3.72 (m, 4H), 2.94 (t, J = 2.4 Hz), 2.48 (t, J = 6.2 Hz, 2H), 1.41 (s, 9H); ^{13}C NMR (100 MHz, acetone-d_6) δ 171.1, 159.0, 158.3, 155.2 (d, J = 244 Hz), 155.1, 154.7, 140.4, 137.0, 124.8, 123.1 (d, J = 7.8 Hz), 121.9, 120.6 (d, J = 19,4), 117.3 (d, J = 21.31), 111.6, 109.2, 81.0, 80.5, 76.3, 75.9, 70.8, 70.7, 67.9, 58.3, 36.9, 28.3; HRMS (ESI): [M+H]$^+$, *m/z* calculated for C$_{27}$H$_{28}$ClFN$_4$O$_7$ 575.17034, found 575.16967.

tert-butyl (R)-3-(3-((4-((3-chloro-4-fluorophenyl)amino)-6-nitroquinazolin-7-yl)oxy)-2-methoxypropoxy)propanoate (4.8)
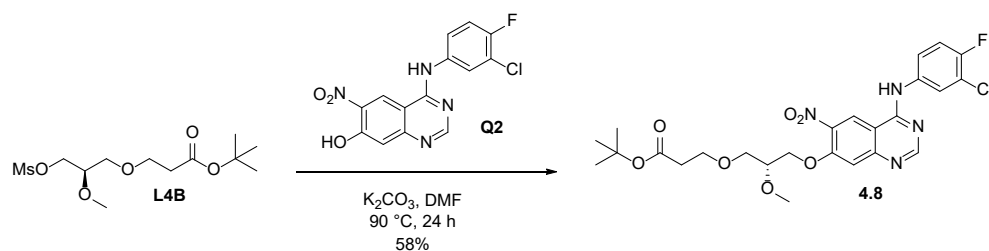
A procedure similar to **1.8** was followed with **L4B** to afford **4.8** as an orange solid (190 mg, 58%). ^{1}H NMR (400 MHz, acetone-d_6) δ 9.51 (br, 1H), 8.93 (s, 1H), 8.68 (s, 1H), 8.25 (dd, J = 6.7, 2.5 Hz, 1H), 7.84 (m, 1H), 7.48 (s, 1H), 7.34 (t, J = 9.0 Hz, 1H), 4.48 (dd, J = 10.2, 3.8 Hz, 1H), 4.36 (dd, J = 10.2, 5.7 Hz, 1H), 3.80 (m, 1H), 3.73 (t, J = 6.2 Hz, 2H), 3.68 (m, 2H), 3.49 (s, 3H), 2.47 (t, J = 6.2 Hz), 1.40 (s, 9H); ^{13}C NMR (100 MHz, acetone-d_6) δ 171.2, 159.0, 158.3, 155.3, 155.2 (d, J = 243.9 Hz), 154.7, 154.0, 140.6, 137.0, 124.8, 123.1, 121.8, 120.6 (d, J = 19.1 Hz), 117.3 (d, J = 22.3 Hz), 111.6, 109.2, 80.5, 79.0, 70.7, 70.4, 67.9, 58.4, 36.9, 28.2; HRMS (ESI): [M+H]$^+$, m/z calculated for $C_{25}H_{28}ClFN_4O_7$ 551.17034, found 551.16977.

(R)-3-(3-((4-((3-chloro-4-fluorophenyl)amino)-6-nitroquinazolin-7-yl)oxy)-2-(prop-2-yn-1-yloxy)propoxy)propanoic acid (1.9)
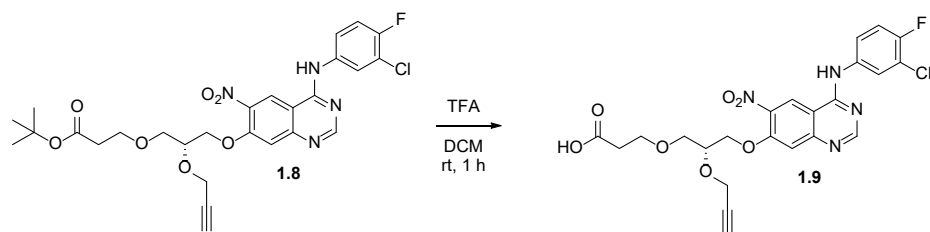
1.8 (18.6 mg, 0.032 mmol, 1 equiv) was dissolved in 10% TFA in DCM (0.65 mL). The solution was stirred at room temperature for 1 h. After this time, the crude mixture was diluted with DCM and rotary-evaporated. The dilution and rotary-evaporation was repeated two more times to afford **1.9** as an orange solid without further purification (16.7 mg, quantitative). ^{1}H NMR (400 MHz, acetone-d_6) δ 9.13 (s, 1H), 8.90 (s, 1H), 8.15 (dd, J = 6.7,

2.6 Hz, 1H), 7.89 – 7.82 (m, 2H), 7.41 (t, J = 9.0 Hz, 1H), 4.58 (dd, J = 10.5, 3.3 Hz, 1H), 4.45 (dd, J = 10.5, 6.1 Hz, 1H), 4.17 (m, 1H), 3.79 (t, J = 6.1 Hz, 2 H), 3.74 (m, 2H), 2.92 (t, J = 2.4 Hz, 1H), 2.56 (t, J = 6.1 Hz, 2H; ^{13}C NMR (100 MHz, acetone-d_6) δ 173.0, 160.2, 156.9, 156.5 (d, J = 245.5), 155.6, 147.8, 141.2, 135.2, 126.4, 124.8 (d, J = 7.2 Hz), 123.1, 121.0 (d, J = 19.5 Hz), 117.7 (d, J = 21.6 Hz), 107.8, 80.9, 76.2, 71.7, 70.3, 67.9, 58.3, 35.3; HRMS (ESI): [M+H]$^+$, m/z calculated for $C_{23}H_{20}ClFN_4O_7$ 519.10774, found 519.10719.

*(R)-3-(3-((4-((3-chloro-4-fluorophenyl)amino)-6-nitroquinazolin-7-yl)oxy)-2-methoxypropoxy)propanoic acid (**4.9**)*
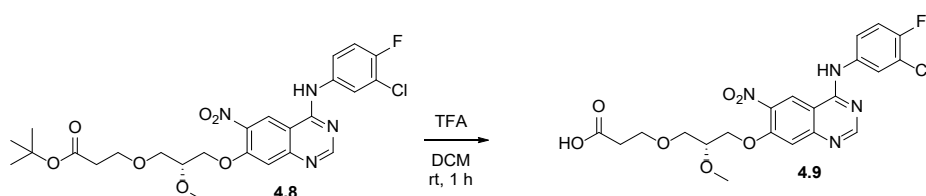
A procedure similar to **1.9** was followed with **4.8** to afford **4.9** as an orange solid (34 mg, quantitative). ^{1}H NMR (400 MHz, acetone-d_6) δ 9.18 (s, 1H), 8.98 (s, 1H), 8.12 (dd, J = 6.27, 2.6 Hz, 1H), 7.95 (s, 1H), 7.83 (m, 1H), 7.43 (t, J = 9.0 Hz, 1H), 4.55 (dd, J = 10.4, 3.3 Hz, 1H), 4.41 (dd, J = 10.4, 6.0 Hz, 1H), 3.82 (m, 1H), 3.77 (t, J = 6.0 Hz, 2H), 3.68 (m, 2H), 3.47 (s, 3H), 2.55 (t, J = 6.0 Hz); ^{13}C NMR (100 MHz, acetone-d_6) δ 173.1, 160.6, 158.1, 157.8 (d, J = 246.8 Hz), 157.5, 154.8, 145.6, 141.4, 134.7 (d, J = 3.6 Hz), 127.0, 125.3 (d, J = 6.9 Hz), 123.3, 121.1 (d, J = 19.4 Hz), 117.8 (d, J = 21.7 Hz), 107.7, 106.1, 78.8, 71.8, 69.9, 67.8, 58.4, 35.3; HRMS (ESI): [M+H]$^+$, m/z calculated for $C_{21}H_{20}ClFN_4O_7$ 495.10774, found 495.10685.

*Succinimidyl (R)-3-(3-((4-((3-chloro-4-fluorophenyl)amino)-6-nitroquinazolin-7-yl)oxy)-2-(prop-2-yn-1-yloxy)propoxy)propanoate (**P1**)*

EDC, NHS

DCM
rt, 8 h
37%

1.9 → **P1**
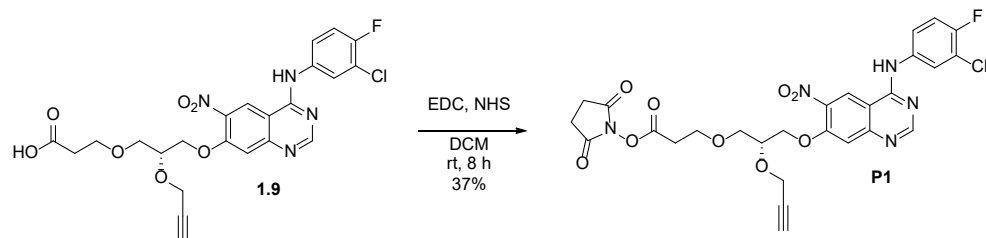
To a mixture of **1.9** (19.4 mg, 0.037 mmol, 1 equiv) and NHS (5.2 mg, 0.045 mmol, 1.2 equiv) in DCM (0.75 mL) at 0 °C was added EDC (10.8 mg, 0.056 mmol, 1.5 equiv). The mixture was then allowed to warm up to room temperature and stirred for 12 h. The solvent was rotary-evaporated and the crude mixture purified by flash column chromatography (40% EtOAc/Hx) to afford **P1** as a yellow solid (8.4 mg, 37%). ^{1}H NMR (400 MHz, acetone-d_6) δ 9.52 (br, 1H), 8.94 (s, 1H), 8.70 (s, 1H), 8.26 (dd, J = 6.8, 2.7 Hz, 1H), 7.85 (m, 1H), 7.49 (s, 1H), 7.35 (t, J = 9.0 Hz, 1H), 4.54 (dd, J = 10.4, 3.5 Hz, 1H), 4.43 (m, 3H), 4.17 (m, 1H), 3.90 (t, J = 5.9 Hz, 2H), 3.78 (m, 2H), 2.96 – 2.92 (m, 3H), 2.89 (m, 4H); ^{13}C NMR (100 MHz, acetone-d_6) δ 170.5, 168.1, 159.0, 158.3, 155.2, 154.7, 140.5, 137.0, 124.8, 123.1 (d, J = 6.9 Hz), 121.9, 120.6 (d, J = 18.4 Hz), 117.3 (d, J = 20.7 Hz), 111.8, 109.2, 81.0, 76.4, 75.9, 71.0, 70.6, 67.0, 58,4, 32.8, 26.3; HRMS (ESI): [M+H]$^+$, m/z calculated for $C_{27}H_{23}ClFN_5O_9$ 616.12412, found 616.12371

*Succinimidyl (R)-3-(3-((4-((3-chloro-4-fluorophenyl)amino)-6-nitroquinazolin-7-yl)oxy)-2-methoxypropoxy)propanoate (**I1**)*

EDC, NHS

DCM
rt, 8 h
45%

4.9 → **I1**
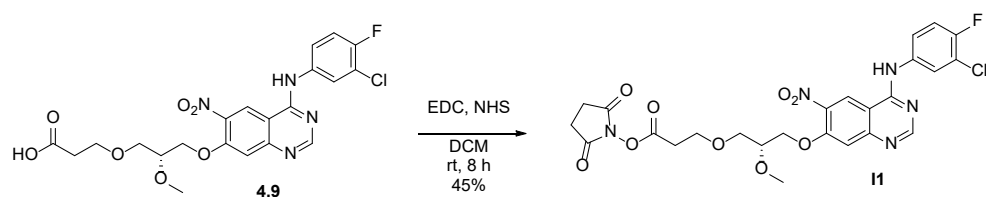
A procedure similar to **P1** was followed with **4.9** to afford **I1** as a yellow solid (18 mg, 40%). ^{1}H NMR (400 MHz, acetone-d_6) δ 9.58 (br, 1H), 8.95 (s, 1H), 8.74 (br, 1H), 8.26 (m, 1H), 7.84 (m, 1H), 7.56 (s, 1H), 7.36 (t, J = 9.0 Hz), 4.51 (dd, J = 10.3, 3.4 Hz), 4.38 (dd, J = 10.3, 3.6 Hz), 3.89 (t, J = 5.9 Hz), 3.83 (m, 1H), 3.74 (d, J = 5.3 Hz, 2H), 3.49 (s, 3H), 2.91 (t, J = 5.9 Hz, 2H), 2.88 (br, 4H); ^{13}C NMR (100 MHz, acetone-d_6) δ 170.5, 168.1, 155.3, 124.8,

123.2, 121.8, 117.3 (d, J = 22.3 Hz), 111.8, 79.0, 71.0, 70.3, 67.0, 58.4, 32.8, 26.3; HRMS (ESI): $[M+H]^+$, m/z calculated for $C_{25}H_{23}ClFN_5O_9$ 592.12412, found 592.12332.

2.3.2.2 Synthesis of linker L2A and corresponding quinazoline derivatives

3-[(4-methoxybenzyl)oxy]propan-1-ol[30] (2.2)
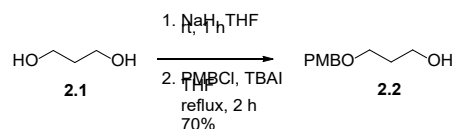
To a solution of 1,3-propanediol (**2.1**, 3.35 g, 44 mmol, 2 equiv) in THF (44 mL) at 0 °C was added NaH (60% w/w, 1.76 g, 44 mmol, 2 equiv). After stirring at room temperature for 1 h, TBAI (1.62 g, 4.4 mmol, 0.2 equiv) and PMBCl (3.44 g, 22 mmol, 1 equiv) were added and the mixture was stirred under reflux for 2 h. The reaction was then quenched at 0 °C by dropwise addition of saturated NH$_4$Cl. THF was rotary-evaporated and the obtained crude mixture extracted with EtOAc, dried over Na$_2$SO$_4$, filtered and rotary-evaporated. The obtained oil was purified by flash column chromatography (20–60% EtOAc/hexanes) to afford **2.2** as a yellow liquid as previously reported (2.96 g, 70%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.25 (m, 2H), 6.88 (m, 2H), 4.45 (s, 2H), 3.80 (s, 3H), 3.77 (q, J = 5.4 Hz, 2H), 3.64 (t, J = 5.8 Hz, 2H), 2.31 (t, J = 5.2 Hz, 1H), 1.85 (quin, J = 5.7 Hz, 2H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.4, 130.3, 129.4, 114.0, 73.1, 69.3, 62.2, 55.4, 32.2.

3-[(4-methoxybenzyl)oxy]propyl methanesulfonate[30] (2.3)
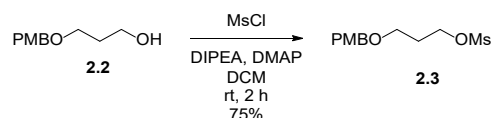
To a solution of **2.2** (2.96 g, 15 mmol, 1 equiv) in DCM (75 mL) at 0 °C was added DIPEA (3.4 mL, 19.5 mmol, 1.3 equiv), MsCl (1.4 mL, 18 mmol, 1.2 equiv) and DMAP (0.37 g, 3 mmol, 0.2 equiv). After stirring at room temperature for 2 h, the reaction was quenched by adding cold water. The aqueous layer was extracted with DCM and the combined organic

layers washed with 0.1 M HCl, saturated NaHCO$_3$ and brine. The organic layer was then dried over Na$_2$SO$_4$, filtered and rotary-evaporated to afford **2.3** as a yellow liquid without further purification as previously reported (3.12 g, 75%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.25 (m, 2H), 6.88 (m, 2H), 4.36 (s, 2H), 4.34 (t, J = 6.3 Hz, 2H), 3.80 (s, 3H), 3.56 (t, J = 5.8 Hz, 2H), 2.96 (s, 3H), 2.04 (quin, J = 6.1 Hz, 2H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.4, 130.2, 129.5, 114.0, 73.0, 67.5, 65.3, 55.4, 37.3, 29.7.

(S)-4-[[3-((4-methoxybenzyl)oxy)propoxy]methyl]-2,2-dimethyl-1,3-dioxolane (2.4)
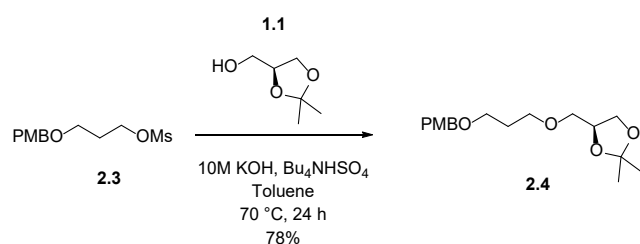
To a stirred solution of **1.1** (1.0 g, 7.6 mmol, 1 equiv) in toluene (75 ml) was added tetrabutylammonium hydrogensulfate (2.6 g, 7.6 mmol, 1 equiv), 10 M KOH (57 mL) and **2.3** (3.1 g, 11.3 mmol, 1.5 equiv). The mixture was heated at 70 °C for 24 h. The mixture was then diluted with water; the organic and aqueous layers were separated, and the aqueous layer was extracted with EtOAc. The combined organic layers were washed with saturated NH$_4$Cl until pH = 7, dried over Na$_2$SO$_4$, filtered and rotary-evaporated. The crude mixture was purified by flash column chromatography (25% EtOAc/hexanes) to afford **2.4** as a yellow liquid (1.8 g, 78%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.25 (m, 2H), 6.81 (m, 2H), 4.42. (s, 2H), 4.24 (quin, J = 6 Hz, 1H), 4.03 (dd, J = 8.3, 6.4 Hz, 1H), 3.80 (s, 3H), 3.70 (dd, J = 8.3, 6.4 Hz, 1H), 3.60–3.49 (m, 5H), 3.41 (dd, J = 9.8, 5.5 Hz, 1H), 1.87 (quin, J = 6.4 Hz, 2H), 1.41 (s, 3H), 1.36 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.3, 130.7, 129.4, 113.9, 109.5, 74.8, 72.8, 72.1, 68.8, 67.0, 55.4, 30.1, 26.9, 25.6.

*(R)-3-[3-[(4-methoxybenzyl)oxy]propoxy]-propane-1,2-diol (**2.5**)*
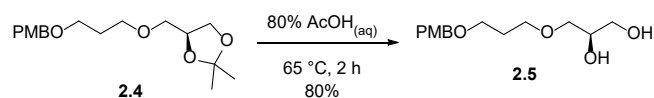
A solution of **2.4** (2.0 g, 6.6 mmol, 1 equiv) in 80% AcOH$_{(aq)}$ (13 mL) was heated at 65 °C for 2 hours. The solvent was then rotary-evaporated at 50 °C. The crude mixture was diluted with toluene and rotary-evaporated three times to remove any residual AcOH. The isolated crude mixture was purified by flash column chromatography (60–100% EtOAc/hexanes) to afford **2.5** as a colourless liquid (1.4 g, 80%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.25 (m, 2H), 6.88 (m, 2H), 4.43 (s, 2H), 3.82 (m, 1H), 3.79 (s, 3H), 3.68 (dd, J = 11.4, 3.9 Hz, 1H), 3.62–3.48 (m, 7H), 1.86 (quin, J = 6.2 Hz, 2H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.3, 130.5, 129.5, 113.9, 72.8, 72.6, 70.6, 68.9, 67.0, 64.3, 55.4, 30.0.

*(S)-1-(4-methoxyphenyl)-11,11,12,12-tetramethyl-2,6,10-trioxa-11-silatridecan-8-ol (**2.6**)*
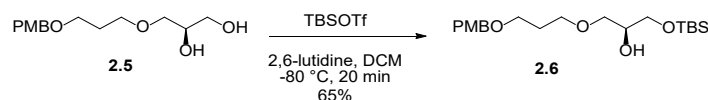
To a solution of **2.5** (1.4 g, 5.2 mmol, 1 equiv) in DCM (52 mL) at −78°C (dry ice/*i*PrOH bath) was added 2,6-lutidine (1.2 mL, 10.4 mmol, 2 equiv) and TBSOTf (1.1 mL, 5.0 mmol, 0.95 equiv) dropwise. The mixture was stirred at −78 °C for 20 min, and then allowed to warm up to room temperature. The solvent was rotary-evaporated and the crude mixture purified by flash column chromatography (25% EtOAc/Hexanes) to afford **2.6** as a colourless liquid (1.3 g, 65%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.25 (m, 2H), 6.87 (m ,2H), 4.43 (s, 2H), 3.81–3.78 (m, 4H), 3.64–3.44 (m, 8H), 2.49 (d, J = 4.9 Hz, 1H), 1.87 (quin, J = 6.3 Hz, 2H), 0.90 (s, 9H), -0.07 (s, 6H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.3, 130.7, 129.4, 113.9, 72.8, 71.7, 70.8, 68.7, 67.1, 64.2, 55.4, 30.2, 26.0, 18.4, −5.3.

(S)-1-(4-methoxyphenyl)-11,11,12,12-tetramethyl-8-(prop-2-yn-1-yloxy)-2,6,10-trioxa-11-silatridecane (2.7)
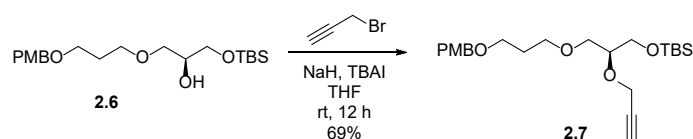
To a suspension of NaH (60% w/w, 0.65 g, 16.4 mmol, 5 equiv) in THF (16 mL) was at 0 °C was added **2.6** (1.3 g, 3.3 mmol, 1 equiv) dropwise. After stirring at room temperature for 1 h, TBAI (0.12 g, 0.33 mmol, 0.1 equiv) was added followed by propargyl bromide (0.97 g, 6.6 mmol, 2 equiv) and the mixture was stirred at room temperature overnight. The reaction was quenched by dropwise addition of saturated NH$_4$Cl. THF was rotary-evaporated and the obtained crude mixture diluted with DCM and washed with brine. The organic layer was dried over Na$_2$SO$_4$, filtered and rotary-evaporated. The isolated crude mixture was purified by flash column chromatography (10% EtOAc/hexanes) to **2.7** as a yellow liquid (0.96 g, 69%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.25 (m, 2H), 6.87 (m, 2H), 4.43 (s, 2H), 4.31 (dd, J = 2.3, 0.9 Hz, 2H), 3.80 (s, 3H), 3.73–3.67 (m, 3H), 3.57–3.47 (m, 6H), 2.39 (t, J = 2.4 Hz, 1H), 1.87 (quin, J = 6.3 Hz, 2H), 0.89 (s, 9H), 0.07 (s, 6H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.3, 130.8, 129.4, 113.9, 80.5, 78.5, 74.1, 72.8, 70.9, 68.7, 67.2, 63.2, 57.9, 55.4, 30.3, 26.0, 18.4, −4.6.

(R)-3-[3-[(4-methyoxybenzyl)oxy]propoxy]-2-(prop-2-yn-1-yloxy)propan-1-ol (L2A)
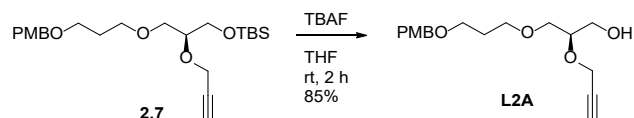
To **2.7** (416 mg, 0.98 mmol, 1 equiv) was added 1M TBAF in THF (2.8 mL, 2.8 mmol, 2.8 equiv) and the mixture was stirred at room temperature for 2 hours. The solvent was then rotary-evaporated and the crude mixture purified by flash column chromatography (50% EtOAc/hexanes) to afford of **L2A** as colourless liquid (258 mg, 85%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.25 (m, 2H), 6.88 (m, 2H), 4.43 (s, 2H), 4.30 (m, 2H), 3.80 (s, 3H), 3.77 (m, 1H),

3.57–3.50 (m, 6H), 2.44 (t, $J = 2.4$ Hz, 1H), 1.86 (quin, $J = 6.2$ Hz, 2H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.3, 130.7, 129.4, 113.9, 80.1, 77.8, 74.7, 72.8, 71.1, 68.9, 67.0, 62.9, 57.6, 55.4, 30.1.

(S)-3-(3-((4-methoxybenzyl)oxy)propoxy)-2-(prop-2-yn-1-yloxy)propyl methanesulfonate (L2B)
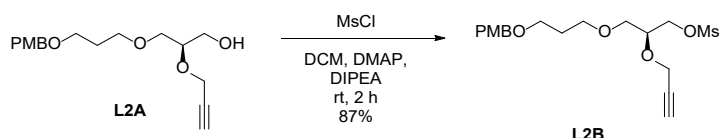
To a solution of **L2A** (1.46 g, 4.74 mmol, 1 equiv) in DCM (24 mL) at 0 °C was added DIPEA (1.1 mL, 6.16 mmol, 1.3 equiv), MsCl (0.44 mL, 5.68 mmol, 1.2 equiv) and DMAP (116 mg, 0.95 mmol, 0.2 equiv). After stirring at room temperature for 2 h, the reaction was quenched adding cold water. The aqueous layer was extracted with DCM and the combined organic layers washed with 0.1 M HCl, saturated NaHCO$_3$ and brine. The organic layer was then dried over Na$_2$SO$_4$, filtered and rotary-evaporated to afford **L2B** as a brown liquid without further purification. ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.24 (m, 2H), 6.88 (m, 2H), 4.42 (s, 2H), 4.36 (dd, $J = 11.0, 3.5$ Hz, 1H), 4.27 (m, 3H), 3.94 (m, 1H), 3.80 (s, 3H), 3.52 (m, 6H), 3.04 (s, 3H), 2.46 (t, $J = 2.5$ Hz, 1 H), 1.85 (quin, $J = 6.3$ Hz, 2H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 159.3, 130.7, 129.4, 113.9, 79.4, 75.3, 75.2, 72.8, 69.4, 69.3, 68.3, 66.8, 57.8, 55.4, 37.7, 30.1.

(S)-N-(3-chloro-4-fluorophenyl)-7-(3-(3-((4-methoxybenzyl)oxy)propoxy)-2-(prop-2-yn-1-yloxy)propoxy)-6-nitroquinazolin-4-amine (2.8)
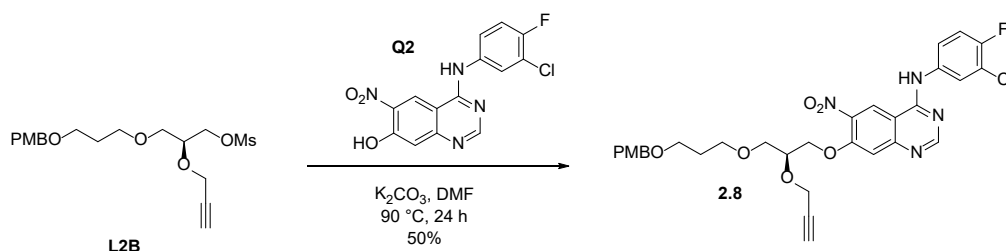
To a mixture of **L2B** (447 mg, 1.34 mmol, 1 equiv) and **Q2** (593 mg, 1.54 mmol, 1.15 equiv) in DMF (4.4 mL) was added K_2CO_3 (203 mg, 1.47 mmol, 1.1 equiv). The mixture was heated at 90 °C for 24 h, and then the solvent was removed at 90 °C under reduced pressure. The crude mixture was diluted with EtOAc and insoluble salts removed by filtration. The filtrate was rotary-evaporated and purified by flash column chromatography (40 – 50% EtOAc/Hx) to afford **2.8** as an orange solid (418 mg, 50%). ^{1}H NMR (400 MHz, DMSO-$d6$) δ 10.15 (s, 1H), 9.22 (s, 1H), 8.67 (s, 1H), 8.15 (dd, J = 6.9, 2.6 Hz, 1H), 7.79 (m, 1H), 7.46 (m, 2H), 7.16 (m, 2H), 6.83 (m, 2H), 4.42 (dd, J = 10.6, 3.8 Hz, 1H), 4.34 (m, 3H), 4.30 (s, 2H), 4.01 (m, 1H), 3.70 (s, 3H), 3.59 (d, J = 5.5 Hz, 2H), 3.50 (t, J = 6.4 Hz, 2H), 3.41 (m, 3H), 1.75 (quin, J = 6.4 Hz, 2H); ^{13}C NMR (100 MHz, DMSO-$d6$) δ 158.5, 157.8, 157.4, 153.7, 153.6 (d, J = 245.0 Hz), 153.3, 138.7, 135.9, 130.4, 128.9, 123.8, 122.6, (d, J = 7.1 Hz), 122.0, 118.9 (d, J = 18.4 Hz), 116.7 (d, J = 21.3 Hz), 113.5, 110.3, 108.1, 80.3, 77.1, 74.9, 71.5, 69.2, 67.7, 66.2, 57.0, 54.9, 29.5.

(S)-3-(3-((4-((3-chloro-4-fluorophenyl)amino)-6-nitroquinazolin-7-yl)oxy)-2-(prop-2-yn-1-yloxy)propoxy)propan-1-ol (2.9)
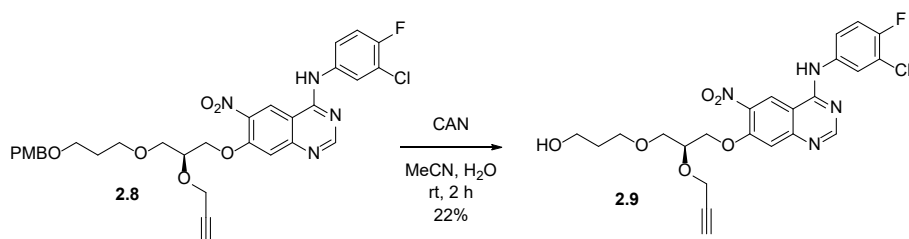
To a solution of **2.8** (133 mg, 0.21 mmol, 1 equiv) in 80% acetonitrile in water (8 mL) at 0 °C was added CAN (700 mg, 1.28 mmol, 6 equiv). After stirring at room temperature for 2 h, the crude mixture was diluted with EtOAc and water. The organic layer was extracted with EtOAc. The combined organic layers were then washed with saturated $NaHCO_3$, dried over $MgSO_4$, filtered and rotary-evaporated. The crude mixture product was purified by flash column chromatography (40 – 100% EtOAc/Hx) to afford **2.9** as an orange solid. ^{1}H NMR (400 MHz, DMSO-$d6$) δ 10.16 (s, 1H), 9.23 (s, 1H), 8.67 (s, 1H), 8.16 (dd, J = 6.8, 2.7 Hz,

1H), 7.79 (m, 1H), 7.46 (m, 2H), 4.44 (dd, J = 3.7, 0.7 Hz, 1H), 4.38 (t, J = 5.1 Hz, 1H), 4.34 (m, 3H), 4.02 (m, 1H), 3.59 (d, J = 5.3 Hz, 2H), 3.50 (t, J = 6.4 Hz, 2H), 3.44 (m, 2H), 3.41 (t, J = 2.3 Hz, 1H), 1.66 (quin, J = 6.4 Hz, 2H); ^{13}C NMR (100 MHz, DMSO-$d6$) δ 157.9, 157.4, 153.7, 153.6 (d, J = 244.4 Hz), 153.3, 138.8, 135.9, 123.8, 122.6 (d, J = 6.8 Hz), 122.0, 118.9 (d, J = 18.1 Hz), 116.6 (d, J = 22.4 Hz), 110.3, 108.1, 80.4, 77.1, 75.0, 69.4, 69.2, 67.9, 57.7, 57.0, 32.6.

2.3.2.3 Synthesis of linker L3A and corresponding quinazoline derivatives

3-(benzyloxy)propan-1-ol[33] (3.1)
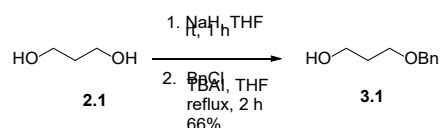
To a solution of 1,3-propanediol (**2.1**, 6.0 g, 79 mmol, 2 equiv) in THF (79 mL) at 0 °C was added NaH (60% w/w, 3.15 g, 79 mmol, 2 equiv). After stirring at room temperature for 1 h, TBAI (2.91 g, 7.9 mmol, 0.1 equiv) and BnCl (5.0 g, 39.5 mmol, 1 equiv) were added and the mixture was stirred under reflux for 2 h. The mixture was then quenched at 0 °C by dropwise addition of saturated NH$_4$Cl. THF was rotary-evaporated and the crude mixture extracted with EtOAc, dried over Na$_2$SO$_4$, filtered and concentrated under reduced pressure. The obtained oil was purified by flash column chromatography (10–40% EtOAc/Hexanes) to afford **3.1** as a yellow liquid (4.36 g, 66%) as previously reported. ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.38–7.27 (m, 5H), 4.53 (s, 2H), 3.79 (t, J = 5.7 Hz, 2H), 3.67 (t, J = 5.7 Hz, 2H), 1.87 (quin, J = 5.7 Hz, 2H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 138.2, 128.6, 127.9, 127.8, 73.4, 69.5, 62.1, 32.2.

3-(benzyloxy)propyl methanesulfonate[33] (3.2)
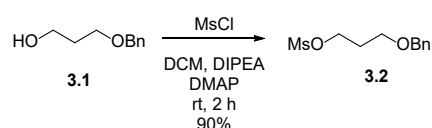
To a solution of **3.1** (4.34 g, 26 mmol, 1 equiv) in DCM (130 mL) at 0 °C was added DIPEA (5.9 mL, 34 mmol, 1.3 equiv), MsCl (2.4 mL, 31 mmol, 1.2 equiv) and DMAP (0.64 g, 5 mmol, 0.2 equiv). After stirring at room temperature for 2 h, the reaction was quenched by adding cold water. The aqueous layer was extracted with DCM and the combined organic layers washed with 0.1 M HCl, saturated $NaHCO_3$ and brine. The organic layer was then dried over Na_2SO_4, filtered and concentrated under reduced pressure to afford **3.2** as a brown liquid without further purification as previously reported (5.7 g, 90%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.38–7.27 (m, 2H), 4.51 (s, 2H), 4.36 (t, J = 6.2 Hz, 2H), 3.59 (t, J = 5.9 Hz, 2H), 2.96 (s, 3H), 2.04 (quin, J = 6.0 Hz, 2H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 138.2, 128.6. 127.9, 127.8, 73.3, 67.5, 65.6, 37.3, 29.7.

(S)-4-[3-(benzyloxy)propoxy]methyl-2,2-dimethyl-1,3-dioxolane[34] (3.3)
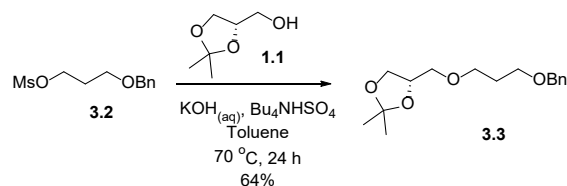
To a stirred solution of **1.1** (2.07 g, 15.6 mmol, 1 equiv) in toluene (156 ml) was added tetrabutylammonium hydrogensulfate (5.3 g, 15.6 mmol, 1 equiv), 10 M KOH (117 mL) and **3.2** (5.7 g, 23.4 mmol, 1.5 equiv). The mixture was heated at 70°C for 24 h. The mixture was then diluted with water; the organic and aqueous layers were separated, and the aqueous layer was extracted with EtOAc. The combined organic layers were washed with saturated NH_4Cl until pH = 7, dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude mixture was purified by flash column chromatography (15% EtOAc/Hx) to afford **3.3** as a colourless liquid (2.8 g, 64%) as previously reported. ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.37– 7.27 (m, 5H), 4.50 (s, 2H), 4.24 (quin, J = 6.0 Hz, 1H), 4.03 (dd, J = 8.2, 6.4 Hz, 1H), 3.70 (dd, J = 8.2, 6.4 Hz, 1H), 3.63–3.54 (m, 4H), 3.51 (dd, J = 9.9, 5.7 Hz, 1H), 3.42 (dd, J = 9.9,

5.5 Hz, 1H), 1.89 (quin, J = 6.3 Hz, 2H), 1.42 (s, 3H), 1.36 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 138.6, 128.5, 127.8, 127.7, 109.5, 74.8, 73.1, 72.1, 68.8, 67.3, 67.0, 30.1, 26.9, 25.6.

(S)-3-[(2,2-dimethyl-1,3-dioxolan-4-yl)methoxy]propan-1-ol[34] (**L3A**)
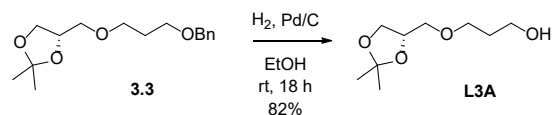
To a solution of **3.3** (960 mg, 3.42 mmol, 1 equiv) in EtOH (8.6 mL) was added 10% Pd/C (364 mg, 0.34 mmol, 0.1 equiv) in a round bottom flask. The system was evacuated and then back-filled with H$_2$ three times, and then stirred under 1 atm of H$_2$ (balloon) while heating at 50 °C for 18 h. The mixture was then allowed to cooled downed to room temperature and filtered through celite. The filtered was washed with EtOH five times. The combined organic extracts were rotary-evaporated and the crude mixture purified by flash column chromatography (60% EtOAc/Hexanes) to afford **L3A** as a colourless liquid as previously reported. ^{1}H NMR (400 MHz, DMSO-d_6) δ 4.37 (t, J = 5.1 Hz, 1H), 4.15 (quin, J = 6.0 Hz, 1H), 3.97 (dd, J = 8.1, 6.5 Hz, 1H), 3.59 (dd, J = 8.2, 6.5 Hz, 1H), 3.48–3.37 (m, 6H), 1.63 (q, J = 6.4 Hz, 2H), 1.30 (s, 3H), 1.26 (s, 3H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 108.4, 74.3, 71.3, 67.8, 66.0, 57.6, 32.6, 26.6, 25.4.

*(S)-3-[(2,2-dimethyl-1,3-dioxolan-4-yl)methoxy]propyl methanesulfonate (**L3B**)*
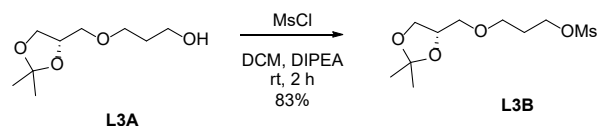
To a solution of **L3A** (320 mg, 1.68 mmol, 1 equiv) in DCM (8.4 mL) at 0 °C was added DIPEA (0.38 mL, 2.19 mmol, 1.3 equiv) and MsCl (0.16 mL, 2.02 mmol, 1.2 equiv). After stirring at room temperature for 4 h, the reaction was quenched by adding cold water. The aqueous layer was extracted with DCM and the combined organic layers washed with brine. The organic layer was then dried over Na$_2$SO$_4$, filtered and concentrated under reduced

pressure to afford **L3B** as a yellow liquid without further purification (374 mg, 83%). ^{1}H NMR (400 MHz, CDCl₃) δ 4.33 (t, J = 6.2 Hz, 2H), 4.26 (m, 1H), 4.05 (dd, J = 8.3, 6.4 Hz, 1H), 3..70 (dd, J = 8.3, 6.4 Hz, 1H), 3.70 (dd, J = 8.3, 6.4 Hz, 1H), 3.59 (m, 2H), 3.51 (dd, J = 10, 5.9 Hz, 1H), 3.46 (dd, J = 10, 5.6 Hz, 1H), 3.01 (s, 3H), 2.01 (m, 2H), 1.41 (s, 3H), 1.35 (s, 3H); ^{13}C NMR (100 MHz, CDCl₃) δ 109.6, 74.8, 72.3, 67.3, 67.0, 66.7, 37.4, 29.5, 26.9, 25.5.

*(S)-N-(3-chloro-4-fluorophenyl)-7-(3-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)propoxy)-6-nitroquinazolin-4-amine (**1.10**)*
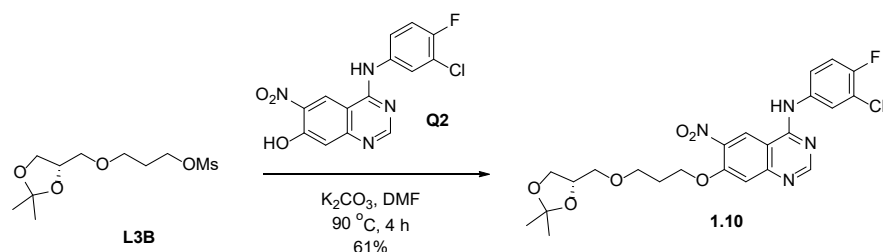
To a mixture of **Q2** (728 mg, 2.17 mmol, 1 equiv) and **L3B** (700 mg, 2.6 mmol, 1.2 equiv) in DMF (7.25 mL) was added K₂CO₃ (331 mg, 2.39 mmol, 1.1 equiv). The mixture was heated at 90 ℃ for 4 h, and then the solvent was removed at 90 ℃ under reduced pressure. The crude mixture was diluted with EtOAc and insoluble salts removed by filtration. The filtrate was rotary-evaporated and purified by flash column chromatography (45 – 100% EtOAc/Hx) to afford **1.10** as an orange solid (678 mg, 61%). ^{1}H NMR (400 MHz, acetone-d_6) δ 10.14 (br, 1H), 9.21 (s, 1H), 8.66 (s, 1H), 8.15 (dd, J = 6.8, 2.5 Hz, 1H), 7.79 (m, 1H), 7.46 (m, 2H), 4.35 (t, J = 6.1 Hz, 2H), 4.17 (quin, J = 6.0 Hz, 1H), 3.96 (dd, J = 8.1, 6.5 Hz, 1H), 3.60 (m, 3H), 3.44 (m, 2H), 2.02 (quin, J = 6.1 Hz), 1.29 (s, 3H), 1.24 (s, 3H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 157.8, 157.4, 153.8, 153.6 (d, J = 244 Hz), 153.3, 138.9, 136.0, 123.8, 122.6 (d, J = 6.2 Hz), 121.8, 118.9, (d, J = 18.6), 116.6 (d, J = 21.7 Hz), 110.2, 108.4, 107.9, 74.2, 71.3, 66.9, 66.7, 65.9, 28.5, 26.6, 25.3; LC-MS (ESI): [M+H]⁺, *m/z* calculated for C₂₃H₂₄ClFN₄O₆ 507.14, found 507.1.

*(R)-3-(3-((4-((3-chloro-4-fluorophenyl)amino)-6-nitroquinazolin-7-yl)oxy)propoxy)propane-1,2-diol (**1.11**)*

A solution of **1.10** (615 mg, 1.21 mmol, 1 equiv) in 80% AcOH$_{(aq)}$ (6.1 mL) was heated at 80 °C for 24 h. The reaction mixture was then cooled on an ice bath and neutralized to pH = 7 with 10 M NaOH. The neutralized crude mixture was then extracted with EtOAc, dried over MgSO$_4$ and rotary-evaporated. The obtained crude mixture was purified by flash column chromatography (60 – 100 % EtOAc/Hx to 0 – 10% EtOH/EtOAc) to afford **1.11** as an orange solid (299 mg, 53%). ^{1}H NMR (600 MHz, DMSO-d_6) δ 10.14 (br, 1H), 9.20 (s, 1H), 8.66 (s, 1H), 8.15 (dd, J = 6.8, 2.5 Hz, 1H), 7.79 (m, 1H), 7.46 (m, 1H), 4.63 (d, J = 5.0 Hz, 1H), 4.47 (t, J = 5.6 Hz, 1H), 4.36 (t, J = 6.1 Hz, 2H), 3.58 (m, 3H), 3.42 (dd, J = 9.9, 4.6 Hz, 1H), 3.33 (m, 3H), 2.02 (quin, J = 6.2 Hz, 2H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 157.8, 157.3, 153.8, 153.6 (d, J = 243.1 Hz), 153.2, 138.9, 135.9, 123.8, 122.6 (d, J = 7.0 Hz), 121.8, 118.9 (d, J = 18.4 Hz), 116.6 (d, J = 21.9 Hz); LC-MS (ESI): [M+H]$^+$, *m/z* calculated for C$_{20}$H$_{20}$ClFN$_4$O$_6$ 467.11, found 467.1.

*(S)-1-((tert-butyldimethylsilyl)oxy)-3-(3-((4-((3-chloro-4-fluorophenyl)amino)-6-nitroquinazolin-7-yl)oxy)propoxy)propan-2-ol (**1.12**)*

To a mixture of **1.11** (100 mg, 0.21 mmol. 1 equiv) and imidazole (29 mg, 0.43 mmol, 2 equiv) in DMF (2.1 mL) was added TBSCl (48 mg, 0.32 mmol, 1.5 equiv). After stirring at
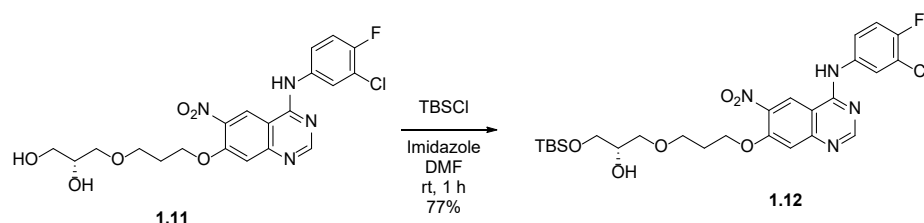
room temperature for 1 h, the solvent was removed under reduced pressure at 50 °C. The obtained crude mixture was purified by flash column chromatography (30–50% EtOAc/Hx) to afford **1.12** as an orange solid (95 mg, 77%). ^{1}H NMR (600 MHz, DMSO-d_6) δ 10.13 (br, 1H), 9.20 (s, 1H), 8.65 (s, 1H), 8.15 (dd, J = 6.8, 2.5 Hz, 1H), 7.79 (m, 1H), 7.45 (m, 2H), 4.69 (d, J = 5.3 Hz, 1H), 4.35 (t, J = 6.2 Hz, 2 H), 3.57 (m, 3H), 3.49 (m, 2H), 3.41 (dd, J = 9.8, 4.6 Hz, 1H), 3.33 (m, 1H), 2.01 (m, 2H), 0.81 (s, 9H), -0.01 (s, 6H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 157.8, 157.3, 154.8, 153.8, 153.6 (d, J = 245.0 Hz), 153.2, 152.4, 138.9, 136.0, 123.8, 122.6, (d, J = 6.6 Hz), 121.7, 118.9 (d, J = 18.8 Hz), 116.6 (d, J = 21.7 Hz), 110.2, 107.9, 71.9, 70.0, 66.7, 66.6, 64.4, 28.6, 25.7, 17.9, −5.4.

2.3.2.4 Synthesis of quinazoline derivatives 1.14 – 1.16

*(S)-N-(3-chloro-4-fluorophenyl)-7-((2,2-dimethyl-1,3-dioxolan-4-yl)methoxy)-6-nitroquinazolin-4-amine (**1.14**)*
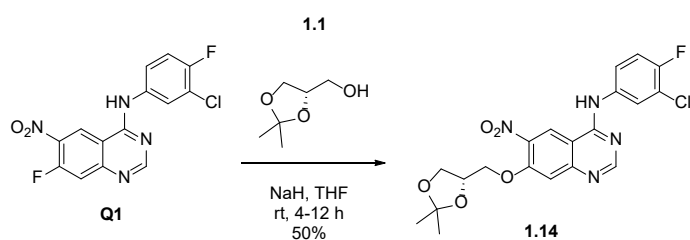
To a suspension of sodium hydride (60% dispersion in mineral oil, 143 mg, 3.6 mmol, 4 equiv) in THF (18 mL) at 0 °C was added **1.1** (0.22 mL, 1.8 mmol, 2 equiv) dropwise. After stirring at room temperature for 1 h, **Q1** was added in small portions. The mixture was stirred at room temperature for 3 hours. THF was rotary-evaporated and the residue was diluted with water yielding a precipitate. The precipitate was collected by filtration, washed with water, dried and purified by flash column chromatography (30–40% EtOAc/hexanes) to afford **1.14** as a yellow solid (200 mg, 50%). ^{1}H NMR (600 MHz, DMSO-d_6) δ 10.13 (br, 1H), 9.20 (s, 1H), 8.66 (s, 1H), 8.15 (dd, J = 6.8, 2.6 Hz, 1H), 7.79 (m, 1H), 7.52 (s, 1H), 7.45 (t, J = 8.9 Hz, 1H), 4.46 (m, 1H), 4.40 (dd, J = 10.6, 3.9 Hz, 1H), 4.34 (dd, J = 10.6, 5.4 Hz, 1H), 4.11 (dd, J = 8.4, 6.6 Hz, 1H), 3.85 (dd, J = 8.5, 6.5 Hz, 1H), 1.35 (s, 3H), 1.32 (s, 3H); ^{13}C NMR

(150 MHz, DMSO-d_6) δ 157.8, 157.4, 153.7 (d, J = 244.6 Hz), 153.5, 153.2, 138.8. 135.9, 123.8, 122.5 (d, J = 8.9 Hz), 121.7, 118.9 (d, J = 20.7 Hz), 116.6 (d, J = 20.7 Hz), 110.6, 109.0, 108.1, 73.3, 69.9, 65.2, 26.3, 25.5; LCMS (ESI): [M + H]$^+$, m/z calculated for C$_{20}$H$_{18}$ClFN$_4$O$_5$ 449.09, found 449.1.

(R)-3-((4-((3-chloro-4-fluorophenyl)amino)-6-nitroquinazolin-7-yl)oxy)propane-1,2-diol
(1.15)
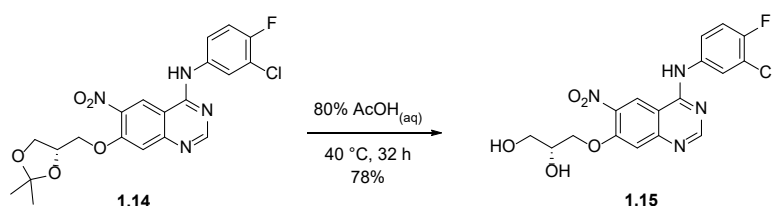
A solution of **1.14** (225 mg, 0.5 mmol, 1 equiv) in 80% AcOH$_{(aq)}$ (2.5 mL) was heated at 40 °C for 32 h. The reaction mixture was then neutralized to pH = 7 with 10 M NaOH. The neutralized crude was then extracted with EtOAc, dried over MgSO$_4$ and rotary-evaporated to afford **1.15** as a yellow solid without further purification (159 mg, 78%). ^{1}H NMR (600 MHz, DMSO-d_6) δ 10.14 (br, 1H), 9.20 (s, 1H), 8.66 (s, 1H), 8.15 (dd, J = 6.5, 2.3 Hz, 1H), 7.79 (m, 1H), 7.47 (s, 1H), 7.45 (t, J = 9.0 Hz, 1H), 5.06 (br, 1H), 4.73 (br, 1H), 4.31 (dd, J = 10.1, 4.1 Hz, 1H), 4.22 (dd, J = 10.2, 5.6 Hz, 1H), 3.86 (m, 1H), 3.49 (m, 2H); ^{13}C NMR (150 MHz, DMSO-d_6) δ 157.8, 157.3, 153.6 (d, J = 243.3 Hz), 154.1, 153.2, 139.0. 135.9, 123.8, 122.6 (d, J = 6.1 Hz), 121.7, 118.9 (d, J = 19.6 Hz), 116.6 (d, J = 22 Hz), 110.3, 107.9, 71.3, 69.5, 62.3; LCMS (ESI): [M + H]$^+$, m/z calculated for C$_{17}$H$_{14}$ClFN$_4$O$_5$ 409.06, found 409.0.

(S)-1-((tert-butyldimethylsilyl)oxy)-3-((4-((3-chloro-4-fluorophenyl)amino)-6-nitroquinazolin-7-yl)oxy)propan-2-ol (1.16)
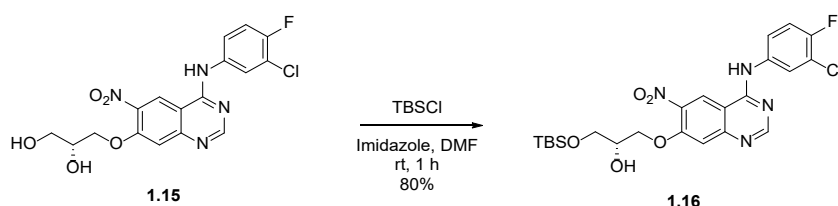
To a mixture of 1.15 (44 mg, 0.11 mmol, 1 equiv) and imidazole (22 mg, 0.22 mmol, 2 equiv) in DMF (1.1 mL) was added TBSCl (49 mg, 0.33 mmol, 3 equiv). After stirring at room temperature for 1 h, the solvent was lyophilized. The residue was extracted with EtOAc and the organic layer washed with water, dried over $MgSO_4$ and rotary-evaporated. The obtained crude mixture was purified by flash column chromatography (0–40% EtOAc/Hx) to afford 1.16 as a yellow solid (45 mg), 80%. ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.16 (br, 1H), 9.22 (s, 1H), 8.67 (s, 1H), 8.16 (dd, J = 6.7, 2.3 Hz), 7.80 (m, 1H), 7.47 (m, 2H), 5.16 (d, J = 5.2 Hz, 1H), 4.30 (dd, J = 10.1, 4.3 Hz, 1H), 4.25 (dd, J = 10.1, 4.9 Hz, 1H), 3.89 (m, 1H), 3.70 (dd, J = 10.1, 6.3 Hz, 1H), 3.66 (dd, J = 10.2, 5.3 Hz, 1H), 0.84, (s, 9H), -0.04 (s, 6H); LCMS (ESI): [M + H]$^+$, m/z calculated for $C_{23}H_{28}ClFN_4O_5Si$ 523.15, found 523.1.

2.4 REFERENCES

1. Fry, D. W.; Kraker, A. J.; McMichael, A.; Ambroso, L. A.; Nelson, J. M.; Leopold, W. R., . . . Bridges, A. J., A specific inhibitor of the epidermal growth factor receptor tyrosine kinase. *Science* **1994,** *265* (5175), 1093-1095.
2. Ward, W. H.; Cook, P. N.; Slater, A. M.; Davies, D. H.; Holdgate, G. A.; Green, L. R., Epidermal growth factor receptor tyrosine kinase. Investigation of catalytic mechanism, structure-based searching and discovery of a potent inhibitor. *Biochem. Pharmacol.* **1994,** *48* (4), 659-666.
3. Barker, A. J.; Gibson, K. H.; Grundy, W.; Godfrey, A. A. B., Jeffrey J. ; Healy, M. P.; Woodburn, J. R., . . . Richards, L., Studies leading to the identification of ZD1839 (Iressa): An orally active, selective epidermal growth factor receptor tyrosine kinase inhiitor targeted to the treatment of cancer. *Bioorg. Med. Chem. Lett.* **2001,** *11*, 1911-1914.
4. Stamos, J.; Sliwkowski, M. X.; Eigenbrot, C., Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.* **2002,** *277* (48), 46265-46272.
5. Yun, C. H.; Boggon, T. J.; Li, Y.; Woo, M. S.; Greulich, H.; Meyerson, M.; Eck, M. J., Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* **2007,** *11* (3), 217-27.
6. Solca, F.; Dahl, G.; Zoephel, A.; Bader, G.; Sanderson, M.; Klein, C., . . . Adolf, G. R., Target binding properties and cellular activity of afatinib (BIBW 2992), an irreversible ErbB family blocker. *J. Pharmcol. Exper. Ther.* **2012,** *343* (2), 342-350.
7. Gajiwala, K. S.; Feng, J.; Ferre, R.; Ryan, K.; Brodsky, O.; Weinrich, S., . . . Stewart, A., Insights into the Aberrant Activity of Mutant EGFR Kinase Domain and Drug Recognition. *Structure* **2013,** *21*, 209-219.
8. Brehmer, D.; Greff, Z.; Godl, K.; Blencke, S.; Kurtenbach, A.; Weber, M., . . . Daub, H., Cellular Targets of Gefitinib. *Cancer Res.* **2005,** *65* (2), 379-382.
9. Salazar Estrada, I. J. Molecular tools for elucidating signaling networks of the protein kinase EGFR. Macquarie University, Sydney, Australia, 2015.
10. Blair, J. A.; Rauh, D.; Kung, C.; Yun, C. H.; Fan, Q. W.; Rode, H., . . . Shokat, K. M., Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. *Nat. Chem. Biol.* **2007,** *3* (4), 229-38.
11. Sun, H.; Ren, Y.; Hou, W.; Li, L.; Zeng, F.; Li, S., . . . Zhang, Z., Focusing on probe-modified peptides: a quick and effective method for target identification. *Chem. Commun.* **2016,** *52*, 10225-10228.
12. Conradt, L.; Godl, K.; Schaab, C.; Tebbe, A.; Eser, S.; Diersch, S., . . . Schneider, G., Disclosure of Erlotinib as a Multikinase Inhibitor in Pancreatic Ductal Adenocarcinoma. *Neoplasia* **2011,** *13* (11), 1026-1034.

13. Lanning, B. R.; Whitby, L. R.; Dix, M. M.; Douhan, J.; Gilbert, A. M.; Hett, E. C., . . . Cravatt, B. F., A road map to evaluate the proteome-wide selectivity of covalent kinase inhibitors. *Nat. Chem. Biol.* **2014,** *10* (9), 760-7.

14. Niessen, S.; Dix, M. M.; Barbas, S.; Potter, Z. E.; Lu, S.; Brodsky, O., . . . Cravatt, B. F., Proteome-wide Map of Targets of T790M-EGFR-Directed Covalent Inhibitors. *Cell Chem. Biol.* **2017,** *24*, 1388-1400.

15. Shindo, N.; Fuchida, H.; Sato, M.; Watari, K.; Shibata, T.; Kuwata, K., . . . Ojida, A., Selective and reversible modification of kinase cysteines with chlorofluoroacetamides. *Nat. Chem. Biol.* **2019,** *15*, 250-258.

16. Choi, S.; Connelly, S.; Reixach, N.; Wilson, I. A.; Kelly, J. W., Chemoselective small molecules that covalently modify one lysine in a non-enzyme protein in plasma. *Nat. Chem. Biol.* **2010,** *6* (2), 133-139.

17. Dalton, S. E.; Dittus, L.; Thomas, D. A.; Convery, M. A.; Nunes, J.; Bush, J. T., . . . Campos, S., Selectively Targeting the Kinome-Conserved Lysine of PI3Kδ as a General Approach to Covalent Kinase Inhibition. *J. Am. Chem. Soc.* **2018,** *140*, 932-939.

18. Pettinger, J.; Jones, K.; Cheeseman, M. D., Lysine-Targeting Covalent Inhibitors. *Angew. Chem. Int. Ed.* **2017,** *56*, 15200-15209.

19. Cuesta, A.; Taunton, J., Lysine-Targeted Inhibitors and Chemoproteomic Probes. *Annu. Rev. Biochem.* **2019,** *88*, 10.7-10.17.

20. Hacker, S. M.; Backus, K. M.; Lazear, M. R.; Forli, S.; Correia, B. E.; Cravatt, B. F., Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **2017,** *9*, 1181-1190.

21. Ward, C. C.; Kleinman, J. I.; Nomura, D. K., NHS-Esters As Versatile Reactivity-Based Probes for Mapping Proteome-Wide Ligandable Hotspots. *ACS Chem. Biol.* **2017,** *12*, 1478-1483.

22. Chavez, J. D.; Weisbrod, C. R.; Zheng, C.; Eng, J. K.; Bruce, J. E., Protein Interactions, Post-translational Modifications and Topologies in Human Cells. *Mol. Cell. Proteomics* **2013,** *12* (5), 1451–1467.

23. Hayashi, T.; Yasueda, Y.; Tamura, T.; Takaoka, Y.; Hamachi, I., Analysis of Cell-Surface Receptor Dynamics through Covalent Labeling by Catalyst-Tethered Antibody. *J. Am. Chem. Soc.* **2015,** *137* (16), 5372-5380.

24. Yamaura, K.; Kuwata, K.; Tamura, T.; Kioi, Y.; Takaoka, Y.; Kiyonaka, S.; Hamachi, I., Live cell off-target identification of lapatinib using ligand-directed tosyl chemistry. *Chem. Commun.* **2014,** *50* (91), 14097-14100.

25. Tamura, T.; Tsukiji, S.; Hamachi, I., Native FKBP12 engineering by ligand-directed tosyl chemistry: labeling properties and application to photo-cross-linking of protein complexes in vitro and in living cells. *J. Am. Chem. Soc.* **2012,** *134* (4), 2216-2226.

26. Tamura, T.; Ueda, T.; Goto, T.; Tsukidate, T.; Shapira, Y.; Nishikawa, Y., . . . Hamachi, I., Rapid labelling and covalent inhibition of intracellular native proteins using ligand-directed N-acyl-N-alkyl sulfonamide. *Nat. Commun.* **2019,** *9*, 1870.

27. Tan, D.; Li, Q.; Zhang, M.-J.; Liu, C.; Ma, C.; Zhang, P., . . . Lei, X., Trifunctional cross-linker for mapping protein-protein interaction networks and comparing protein conformational states. *Elife* **2016,** *5*, e12509.

28. Bauer, F.; Rueb, K.-P.; Lieflander, M., Zur stereoselektiven Synthese von 2-0-Benzyl-1-0-hexadecyl-sn-glycerol. *Liebigs Ann. Chem.* **1992,** *1992*, 47–50.

29. Dumbre, S.; Derouaux, A.; Lescrinier, E.; Piette, A.; Joris, B.; Terrak, M.; Herdewijn, P., Synthesis of modified peptidoglycan precursor analogues for the inhibition of glycosyltransferase. *J. Am. Chem. Soc.* **2012,** *134* (22), 9343-9351.

30. Urbanek, R. A.; Sabes, S. F.; Forsyth, C. J., Efficient Synthesis of Okadaic Acid. 1. Convergent Assembly of the C15-C38 Domain. *J. Am. Chem. Soc.* **1998,** *120*, 2523–2533.

31. Raghupathy, R.; Anilkumar, A. A.; Polley, A.; Singh, P. P.; Yadav, M.; Johnson, C., . . . Mayor, S., Transbilayer lipid interactions mediate nanoclustering of lipid-anchored proteins. *Cell* **2015,** *161* (3), 581-594.

32. Wu, J.; Chen, W.; Xia, G.; Zhang, J.; Shao, J.; Tan, B., . . . Yu, Y., *ACS Med. Chem. Lett.* **2013,** *4* (10), 974–978.

33. Mian, F.; Bottaro, G.; Seraglia, R.; Cavazzini, M.; Quici, S.; Armelao, L., The Role of Ligand Topology in the Decomplexation of Luminescent Lanthanide Complexes by Dipicolinic Acid. *ChemPhysChem* **2016,** *17*, 3229–3236.

34. Meimetis, L. G.; Williams, D. E.; Mawji, N., R.; Banuelos, C. A.; Lal, A. A.; Park, J. J., . . . Andersen, R. J., *J. Med. Chem.* **2012,** *55*, 503–514.

CHAPTER 3

Reverse Chemical Proteomics with a

Non-Covalent EGFR-Directed Probe

3.1 INTRODUCTION

3.1.1 Reverse chemical proteomics with T7 phage display

Reverse chemical proteomics borrows its name from chemical genetics studies, referring to the link going in the direction from gene to phenotype.[1-3] While in forward chemical proteomics a cell's proteome functions as a library from which binding partners for a particular small-molecule are isolated, reverse chemical proteomics makes use of the cell's genome (transcriptome) as the starting point. In both cases, selection is based on small-molecule/proteins interactions. However, the main characteristic of reverse chemical proteomics is the physical link between the protein of interest and its encoding gene, which makes possible the isolation, amplification and identification of target proteins. This physical link is made possible by display technologies, which comprise two major groups: virus or cell based methods (phage, bacterial, yeast, human) and *In vitro* technologies (mRNA, DNA, plasmid, ribosome).[4]

Among these display technologies, phage display is the most commonly used and probably also the best understood.[5] Phage display in turn can be based in two types of phage particles: filamentous (M13) phage, which replicate and assemble without killing their host and are therefore secreted through the outer membrane, and lytic phage (T7), which are released from the host by cell lysis. Since its invention in 1996 by Rosenberg et al and commercialization by Novagen (Merck-Millipore),[6] the T7 phage display has become the most popular method for reverse chemical proteomics due to its advantages over filamentous systems. T7 phage particles are easy to grow, replicate rapidly, do not require transport through the periplasmic space and are extremely robust. Each particle can display peptides up to about 50 residues in high copy number (415 per phage) and up to 1200 residues in low copy number (0.1-1 per phage) or mid copy number (5-15 per phage).[7] T7 cDNA libraries from different tissues are commercially available.[8] To generate these libraries, mRNA is reversed transcribed to its

corresponding cDNA, and inserts between 300 and 3000 bp are cloned directionally into *EcoR I*/*Hind III* from the T7 select vector. This allows the inserts to be displayed as surface fusion proteins to the C–terminus of the T7 gene 10 major capsid protein.

Although the T7 phage display is a powerful tool that can allow the rapid identification of potential drug targets, it also presents some limitations. PTMs of displayed proteins are not possible in a bacterial host, and the limit in the size of display proteins makes the display and correct folding of multi-domain and membrane proteins challenging. However, fragments or single protein domains can still retain their ability to bind to the target small-molecule under certain conditions.[9]

3.1.2 Overview of the reverse chemical proteomics workflow.

The workflow for reverse chemical proteomics using the T7 phage display is well established (**Scheme 3.1**) and has been previously used for the identification of targets of small molecules.[10-12] The drug or small molecule of interest is usually modified intro a biotinylated probe to allow its use as a bait. Although the probe immobilization can be done on beads, the most popular method makes use of streptavidin-coated microplates.

Affinity purification starts with the incubation of the initial T7 cDNA phage library with the immobilized probe (**Scheme 3.1**, step 1). In order to reduce non-specific binders, a control probe that usually consist of the biotinylated linker without the small-molecule of interest can be used for pre-incubation with the phage library to remove some non-specific binders. Once the protein/small-molecule binding has occurred, unbound phage particles are washed away (**Scheme 3.1**, step 2) and the remaining particles are eluted (**Scheme 3.1**, step 3). Eluted phage particles are then amplified in the *E. coli* host system (**Scheme 3.1**, step 5) to generate the next phage library enriched with phage particles displaying proteins that can bind to the immobilized probe for another round of selection (**Scheme 3.1**, step 6). A single round of

selection is usually referred to as biopanning. This iterative process allows the enrichment of more specific binders.
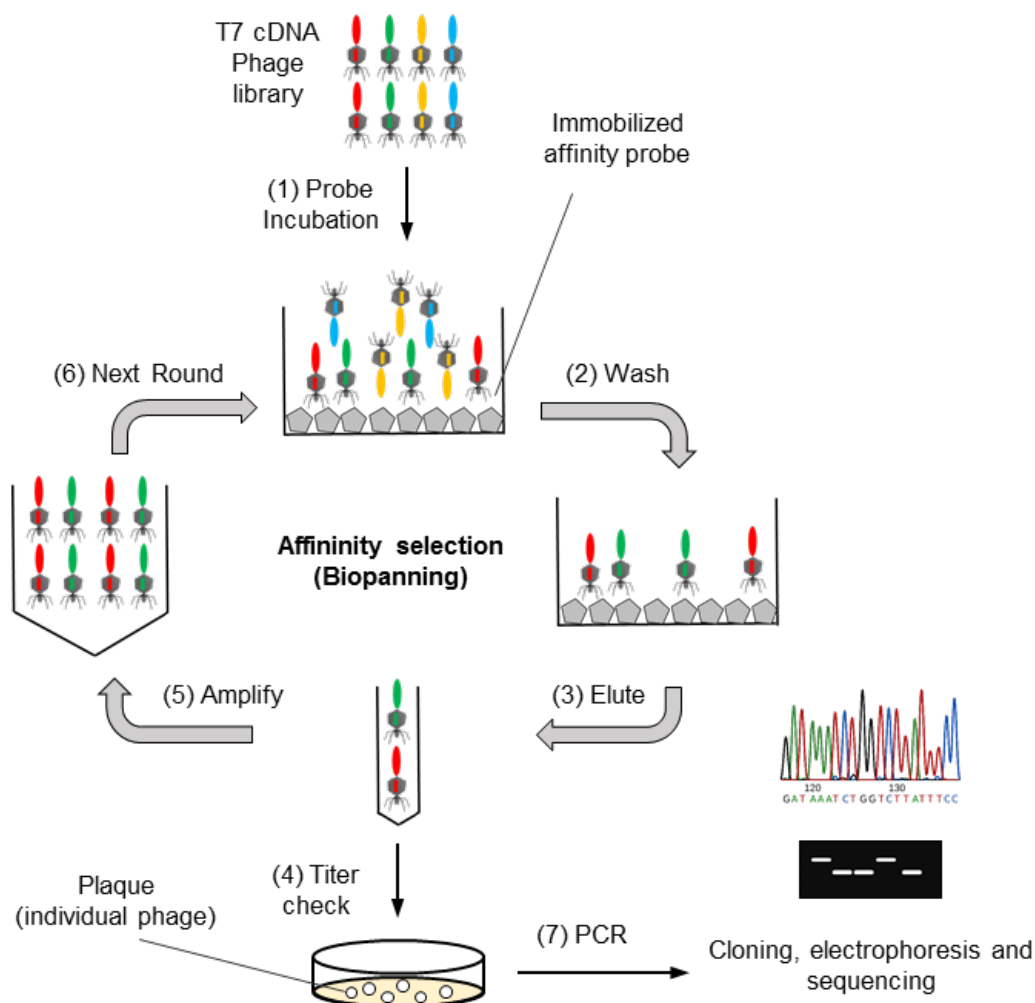
Scheme 3.1 Workflow of reverse chemical proteomics using the T7 phage display system

After several rounds, the degree of enrichment can be determined by titering the number of phage particles obtained after each round (**Scheme 3.1**, step 4). To this end, a small volume of the eluted fraction is serially diluted and a small aliquot of each dilution is dropped on a thin layer of *E. coli*-infected agarose plate. Individual phage particles are observed as clear plaques against the opaque bacterial lawn, and the dilutions containing a countable number of plaques are used to determine the phage particle titer of each round. Generally five to six biopanning rounds are necessary for convergence to a manageable number of clones. Individual plaques are then isolated, and their DNA products are extracted and amplified by PCR for further

analysis (**Scheme 3.1**, step 7). Digestion of PCR products with a restriction endonuclease and separation by agarose gel electrophoresis allows the quick identification of clones with similar DNA inserts, which are then finally sequenced and searched against the BLAST protein database to determine the identity of the target protein.

3.1.3 Identification of targets of small-molecules

The first example of the potential of phage display for the identification of binding proteins of small-molecules was reported by Austin et al in 1999.[13] Using a biotinylated version of the immunosuppressant drug FK506, they were able to isolate and identify its known target FK506-binding protein (FKBP) from a human brain cDNA library. After this proof of concepts, this methodology has been successfully applied in the identification of targets of small molecules, including bioactive natural products, clinically approved drugs and carcinogenic agents (**Table 3.1**).[10-12]

Table 3.1 Selected reported examples of the identification of small-molecule binding targets by reverse chemical proteomics with T7 or M13 phage display

Entry	Small-molecule	Target identified	Reference
1	FK506	FKBP	13
2	Doxorubicin	hNopp140	14
3	HBC	Ca^{2+}/Calmodulin	15
4	Paclitaxel	Bcl-2, NSC-1, NFX	16-17
5	Demethylasterriquinone (DAQ) B1	GADPH	18
6	Kahalalide F	RPS25	19
7	Bisphenol A	TACC3	20
8	Cyclosporine A	CypA/B	21
9	Terpestacin	UQCRB	22
10	Camptothecin	hnRNP A1	23
11	Sulfoquinovosyl diacylglyceride	FAK	24
12	Artesunate	BAD	25

Due to its intrinsic limitations in expressing membrane proteins, the use of T7 phage display system for profiling binding targets of kinase inhibitors has not been extensively explored. However, other expression cloning techniques, such as the yeast three-hybrid system, have been used to this end.[26] In 2011, Chidley et al used this technique with an Erlotinib derivative probe to identify potential off-targets (**Fig 3.1**).[27] Unlike most EGFR probes, they derivatized Erlotinib through the aniline rather than the quinazoline ring, which may have interfered with its binding to EGFR and other kinases. Accordingly, they were unable to isolate any kinase, but identified oxysterol-binding protein-related protein (ORP7) as the first non-kinase target of Erlotinib, demonstrating the utility of this approach in expanding the known drug interactome.
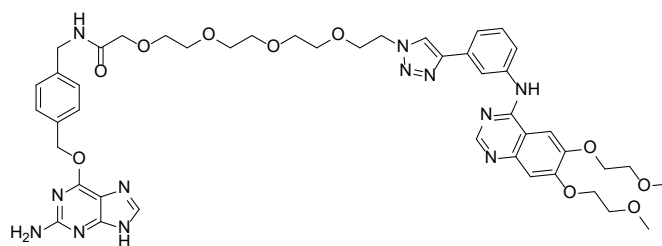
Fig. 3.1 Structure of Erlotinib-derivative probe
used in yeast three-hybrid system.

Since kinase inhibitors off-targets are not limited to kinases (the target landscape of clinical kinase drugs),[28] the T7 phage display represents an interesting approach complementary to traditional forward chemical proteomics analysis for the identification of binding proteins. In this chapter, we describe the use of a non-covalent EGFR-directed chemical probe for 1) testing if it is possible to find the intended target EGFR and 2) identifying potential off-target binders in cDNA T7 phage tumour libraries in a completely agnostic manner.

3.2 RESULTS AND DISCUSSION

3.2.1 Biopanning

To generate the affinity supports for display cloning, a biotinylated derivative of Gefitinib, probe **P3,** and the corresponding control probe **P4** were immobilized on neutravidin-coated polystyrene microtiter plates (**Fig. 3.2**). These probes had been previously synthesized in our group and demonstrated to retain binding affinity for EGFR in the low nanomolar range as indicated by their K_d values.[29] Similar probes based on the structure of Gefitinib and Erlotinib but covalently attached to agarose or sepharose beads have been used previously for the identification of drugs off-targets in forward chemical proteomics analysis. Thus, probe **P3** was selected in a first attempt to validate the use of reverse chemical proteomics based on T7 phage display as a complementary approach to traditional forward proteomics methods for target and off-target identification.
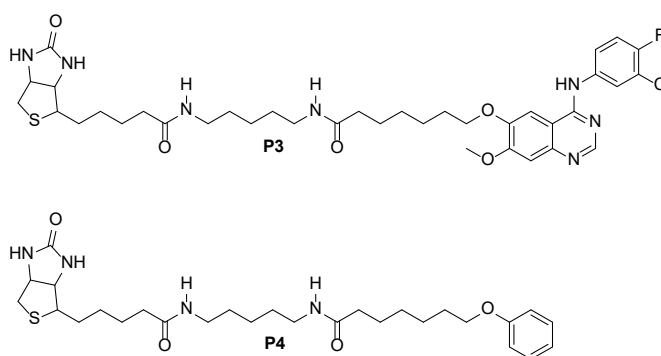
Fig. 3.2 Structure of EGFR probe **P3** and control probe
P4 used for biopanning experiments

Two human cDNA libraries (lung tumour and breast tumour) were subjected to eight rounds of selection. The library phage lysate was first incubated with the control **P4** for one hour to remove non-specific displayed proteins capable of binding to the plastic plate, neutravidin, biotin or the probe linker. This pre-cleared lysate was then transferred to the immobilized probe **P3** and incubated for two hours. Following incubation, the phage solutions were

aspirated and weakly bound proteins were washed away with phage wash buffer (PWB). A mild washing condition (1 wash, < 3 s) was used for the first round of selection to prevent the loss of non-abundant moderate affinity binders. After each successive round of selection, the stringency of the washing step was increased gradually by using a maximum of 20 washes (< 3 s each) after round 6. Finally, retained phage were eluted by the disruption of the avidin-biotin interaction with 1% SDS. The eluted fraction was then used as the sub-library for the next round of selection. The procedure was repeated until eight rounds of selection had been completed.

3.2.2 Enrichment determination by phage titer and agarose gel electrophoresis

A quick way to determine the degree of specific enrichment is to measure the overall number of phage particles eluted after each round. To determine this phage titer, an aliquot of the eluted fraction from each round of selection was serially diluted from 10^{-1} to 10^{-6}. An aliquot of each dilution was dropped onto an agar plate containing a thin layer of *E. coli*-infected LB agarose. The plates were incubated at 37 °C for 2-3 hours, after which clear plaques became visible against the opaque bacterial lawn. The dilutions containing a countable (5-50) number of plaques were used to calculate the phage titer for each round of selection (**Fig. 3.3**).

The lung tumour library reached a plateau with a 10-fold increase in phage titer after four rounds of selection (**Fig. 3.3a**). This suggests that selective enrichment of one of more clones by probe **P3** had occurred by round four. A second 10-fold increase in phage titer with respect to round four (and 100-fold from round two) was observed by round eight, suggesting the enrichment of a second population of clones. A similar behaviour was observed for the breast tumour library (**Fig 3.2b**), with 10- and 100-fold increase in phage titer after rounds four and seven, respectively. The drop of phage titer for round five in breast tumour library and round eight for lung tumour library may be due to a dilution error.
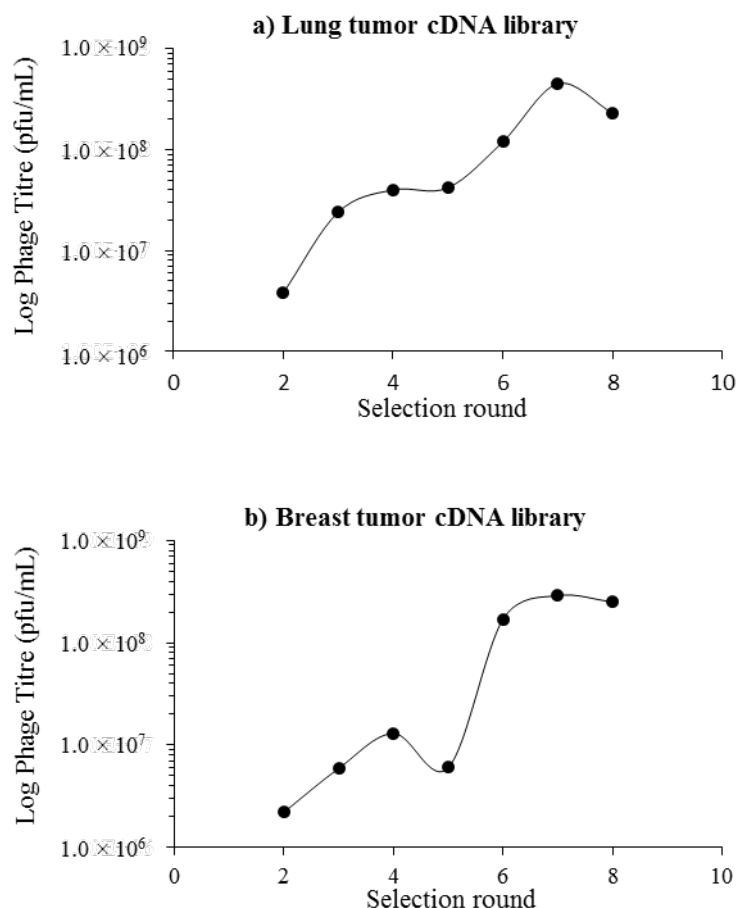
Fig. 3.3 Phage titer of each round of selection for a) lung tumour and b) breast tumour cDNA library using immobilized probe **P3**

The DNA inserts of phage enriched after each round of selection were amplified by PCR with generic T7 primers and then subjected to agarose gel electrophoresis (**Fig. 3.4**). For both libraries, the DNA amplified from the first 2–3 rounds appeared as a smear on the gel from 400–1000 bp, characteristic of the presence of a wide range of different sized DNA inserts in equal proportions. A faint DNA band around 800 bp was observed from round three in the lung tumour library (**Fig. 3.4a**). The intensity of the band increased with each round of selection, and the enrichment of this clone was observed clearly by round six. Similarly, the enrichment of DNA inserts around 500 bp from round four was apparent in the breast tumour library (**Fig. 3.3b**).

a) Lung tumour library

b) Breast tumour library
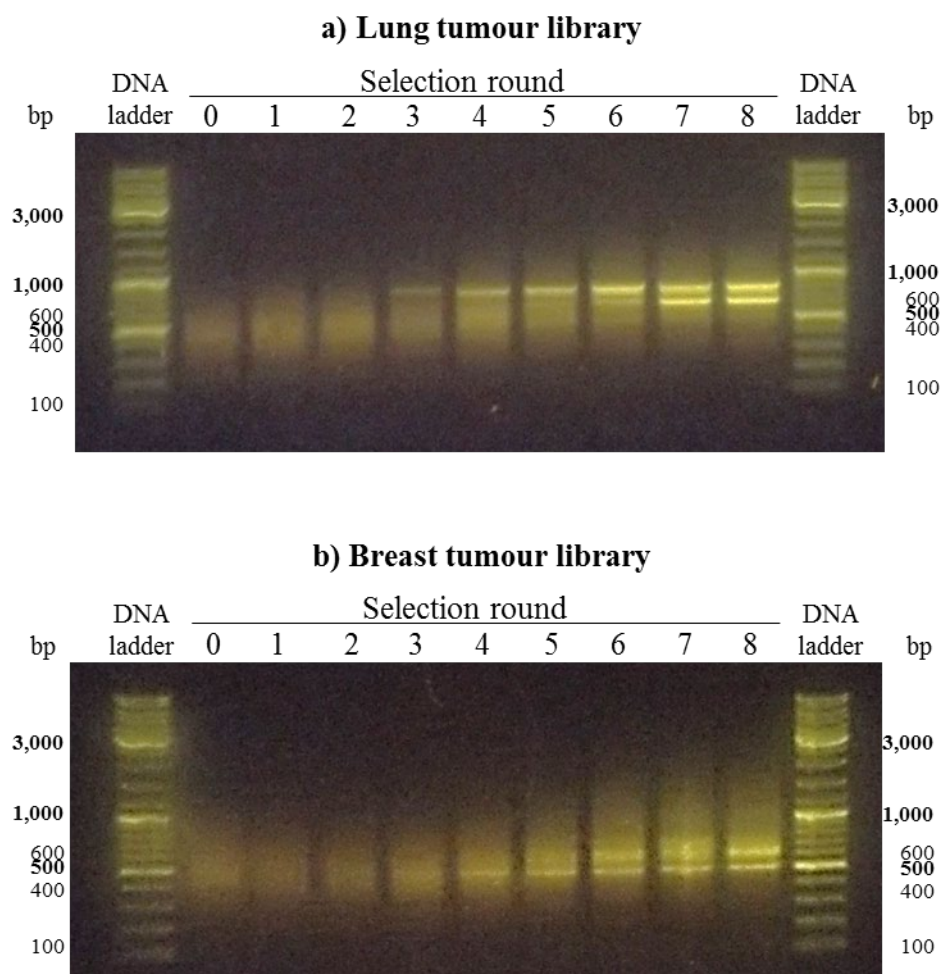
Fig. 3.4 Agarose gel electrophoresis of phage DNA inserts amplified by PCR with T7 primers from a) lung tumour and b) breast tumour cDNA libraries after each of eight rounds of selections with immobilised probe **P3**.

Despite the difference in bp of the DNA inserts enriched in the lung and breast tumour libraries from the first four rounds, it was interesting to observe that both libraries converged to a common population of 600 bp after seven rounds of selection. The enrichment of two different groups of DNA inserts after round four and seven observed by gel electrophoresis was consistent with the increments and plateaus determined by the quantification of phage titer. The displayed proteins from the enriched phage clones may include both common and specific targets for different biological libraries, as seen from the agarose gels of breast and lung tumour libraries.

3.2.3 Identification of target proteins

To determine the identity of the proteins displayed on enriched phage particles, individual phage particles were isolated and their DNA inserts amplified by PCR. Phage lysate from round eight was serially diluted and an aliquot mixed with *E. coli*-infected molten LB agarose. The molten infected solutions were poured onto LB agar plates, left to set for 15 minutes at room temperature, and then incubated at 37 °C for 2–3 hours until clear plaques (around 50 plaques per plate) were visible against the opaque lawn of bacteria. Twelve random plaques were isolated for each library and their DNA inserts amplified using generic T7 primers. A fraction of the amplified DNA from each individual plaque was digested with *HinfI* to produce a unique DNA fingerprinting. Both digested and non-digested samples were subjected to agarose gel electrophoresis (**Fig. 3.5**). The PCR products of clones containing unique DNA inserts (as determined by DNA fingerprinting) were purified and sequenced (**Table 3.2**). Clones containing similar DNA fingerprints were then assigned according to the identity of the only unique sequenced clone.

Table 3.2 DNA sequencing of PCR products obtained after the eighth round of selection with probe **P3**

Library	Plaque	Gene identified from DNA Sequence	Notes	Frame
LuT	E3	Homo sapiens sorting nexin 1 (SNX1), transcript variant 1, mRNA	First 591 of 1569 bp CDS	1
LuT	E4	Homo sapiens DNA topoisomerase I (TOP1), mRNA	319 of 2298 bp CDS	1
LuT	E5	Homo sapiens nucleolar and coiled-body phosphoprotein 1 (NOLC1), transcript variant 1, mRNA	660 of 2139 bp CDS	1
LuT	E7	Homo sapiens UBX domain protein 4 (UBXN4), mRNA.	389 of 1527 bp CDS	2
BrT	A1	Homo sapiens DNA topoisomerase I (TOP1), mRNA	417 of 2298 bp CDS	1
BrT	A2	No significant similarity	-	-
BrT	A5	Homo sapiens BCL2 associated agonist of cell death (BAD), transcript variant 2, mRNA	Full CDS, 507 bp	1
BrT	A9	Homo sapiens dystrobrevin beta (DTNB), transcript variant 15, mRNA	Last 65 Of 1353	3
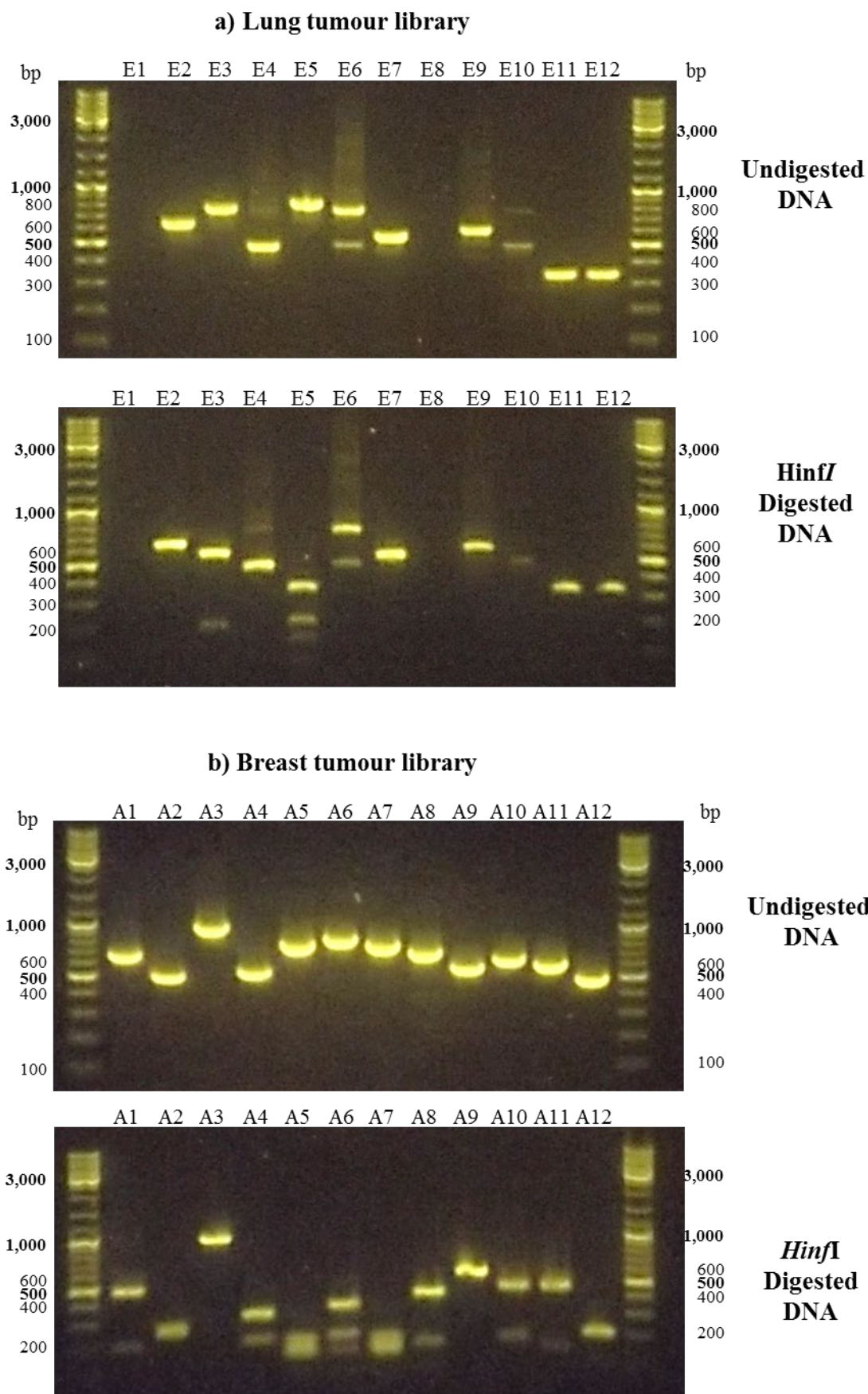
Fig. 3.5 Agarose gel electrophoresis of PCR products obtained from a) lung tumour and b) breast tumour individual plaques after the eighth round of selection with probe **P3**

The majority of the identified DNA inserts (5 out of 8) are in frame with the viral coat and within the coding region (CDS, **Table 3.2**), which suggest that the phage particles displayed sizable protein domains as potential binders of probe **P3**. However, neither EGFR nor other known kinase off-targets were identified. As previously mentioned, in these commercial libraries cDNA cloning is done with the restriction enzymes *EcoRI/HindIII*. Analysis of the mRNA nucleotide sequencing showed two *EcoR I* and one *HindIII* restriction sites at positions 258/3168 and 3890 respectively. The first possible cDNA insert comprises most of the *EGFR* coding sequence, contains more than 3,000 bp, and therefore is not inserted into the T7 vector. The second possible cDNA insert comprises 723 bp, which translates into a peptide of 241 residues corresponding to the C-terminal tail of EGFR. Although the restriction digestion was only partially in the library preparation to potentially allow some of the EGFR kinase domain to be displayed, the lack of EGFR capture indicates that it is more likely that the EGFR kinase domain is not displayed in this library using these restriction sites. In the future, the construction of libraries with different restriction enzymes and with relevant levels of *EGFR* mRNA could afford phage particles capable of displaying the kinase domain to overcome this difficulty.

Among the identified targets, the most promising was DNA topoisomerase I (TOPI) as it was the only one common to both library selections. In addition, TOPI was the dominant clone of the breast tumour library, with 3 out of 12 single plaques potentially corresponding to this gene as determined by their DNA fingerprint (lanes A1, A8 and A10 in **Fig. 3.5 b**; only A1 DNA was sequenced). The displayed peptide from this TOPI insert corresponds to 106 residues out of 765 of the target protein close to the N-terminus (aa 89–194).

TOPI plays an important role in maintaining the DNA topology during DNA replication and transcription. It is overexpressed in many types of cancers, and is a validated drug target of chemotherapeutic drugs such as camptothecins.[30] Cross-talk between EGFR and TOPI

through various mechanisms in cancer has been previously reported in the literature, and combined inhibition of both proteins has been shown to have a synergistic effect capable to overcome resistance to chemotherapy.[31] In 2014, Peleg et al identified TOP1 as a target of Erlotinib through a series of biochemical assays.[32] Erlotinib reduced TOP1 DNA relaxation activity in a dose dependant manner, which was not accompanied by reduction of TOP1 levels. Since a direct interaction of Erlotinib and DNA was excluded, it was hypothesized that Erlotinib might bind directly to TOP1, triggering a conformational change that modifies its DNA binding activity. Our identification of TOP1 as a target of probe **P3** supports this hypothesis of a direct interaction.

Despite current literature evidence of TOP1 as a target of Erlotinib,[32] further experiments are required for the validation of TOP1 as a target of probe **P3** in reverse chemical proteomics. For example, TOP1 could be cloned and overexpressed in mammalian cells to determine its interaction with probe **P3**. Competed experimented by pre-treatment with **P3** analogue free inhibitor could be useful for the determination of a K_d value. Another approach may include the use of on-phage binding assay, in which the affinity of the TOP1-displaying phage for **P3**-coated and a control plate (using a phage particle without cDNA insert as a negative control) is compared. Other methods for the identification of direct targets of small molecules such as DARTS or the use of proteome chips could also be helpful for this end.[33]

Recently, a dual inhibitor of EGFR and Topoisomerase II (which requires ATP binding) was reported based on a 2,4-bis(arylamino) substituted quinolone,[34] and similarly, it has been proposed that the information on the direct physical interaction of EGFR and TOP1 would bring opportunities for the design of dual inhibitors less prone to drug resistance.[31] This proof of concept demonstrates the utility of reverse chemical proteomics as a complementary approach to identify targets that might be concealed in conventional forward methodologies, which can then lead to further studies to expand the utility of known drugs.

3.2.4 Conclusions

In summary, T7 phage display is a powerful technique and complementary approach for the identification of small-molecule/protein interactions, with its own advantages and limitation. The nature of this method can yield unique information, as prior knowledge of potential targets is not required. Unlike forward chemical proteomics, protein abundance is irrelevant for target identification since any binder can be amplified. The fast growth rate of the phage also reduces the analysis time in comparison with forward approaches. In this study, these advantages were reflected in the identification of TOP1, a non-kinase protein, as a potential target of probe **P3**. Although further studies are required for the validation of this target, recent literature evidence from biochemical assays supports the possibility of TOP1 being a direct target of EGFR inhibitors.[32] Unlike forward chemical proteomics, the starting point of reverse chemical proteomics is the genome (transcriptome) rather than the proteome. This represents a major drawback for the technique, as not all proteins are equally inserted and expressed in the display system. This was reflected in the absence of EGFR identification as a target of probe **P3**, due to the specific restrictions enzymes utilized in the construction of the library. Moreover, the use of a bacterial display system implies that the eukaryotic posttranslational machinery, which is of great importance for cell signaling, is absent. The construction of new libraries using various restriction enzymes, as well as the use of eukaryotic display systems such as yeast display might be able to overcome these shortcomings and potentially allow the identification of EGFR and other EGFR inhibitors off-targets by reverse chemical proteomics.

3.3 Experimental

3.3.1 Phage Display Overview

Biosafety approval was obtained from Macquarie University Biosafety Committee (Approval number NLRD 5201600329). The display cloning protocols presented below were adapted from those described in the Novagen T7 Select System Manual.[7]

3.3.2 Materials

Potassium dihydrogen phosphate, sodium dodecyl sulfate, DMSO, LB broth with agar, GenElute™ PCR clean-up kit were obtained from Sigma-Aldrich (USA). Ammonium chloride, disodium hydrogen phosphate were obtained from BDH (Germany). Sodium chloride, glucose, PBS tablets, agarose, IPTG, glycerol, and Tween 20 were obtained from Astral Scientific (Australia). 1M magnesium chloride, high resolution agarose, 25× TAE, nuclease free water were obtained from AMRESCO (USA). Tryptone, yeast extract were obtained from BD (USA). Carbenicilin disodium was obtained from AG Scientific (USA). *Taq* DNA polymerase, 10× CoralLoad PCR Buffer were obtained from QIAGEN (USA). *Hinf*I restriction endonuclease, CutSmart® buffer were obtained from New England Biolabs (USA). SYBR® safe DNA gel stain, DNA molecular weight markers, NeutrAvidin coated high capacity 8-well strip plates, nucleotides (dNTPs) and oligonucleotides (primers) were obtained from ThermoFisher Scientific (USA). Human disease T7 select libraries, *E. coli* strain BLT5615 were obtained from Novagen (USA). Reagents and media were prepared as described in **Table 3.3**.

Table 3.3 Preparation of reagents and media used in phage display.

Reagent	Ingredients	Instructions
20×M9 salts	20 g NH$_4$Cl 60 g KH$_2$PO4 120 g Na$_2$HPO$_4$.12H$_2$O Water to make 1 L of solution	Adjust pH to 7.4 Autoclave 20 min / 121 °C
20% Glucose	50 g glucose Water to make 250 mL of solution	Sterile filter (0.22 μm) into sterile bottle
10% Carbenicilin	1 g carbenicillin disodium Water to make 20 mL of solution	Sterile filter (0.22 μm) Store at -20 °C in 500 μL aliquots
M9TB	9.28 g tryptone 4.64 g NaCl 50 mL 20×M9 salts 1 mL MgCl2 (1 M)	Autoclave 20 min / 121 °C Cool to room temperature Add 20 mL sterile 20% glucose Add 500 μL sterile 10% carbenicilin

	Water to make 1 L of solution	
2×YT	4.25 g tryptone 2.50 g yeast extract 1.25 g NaCl Water to make 250 mL of solution	Adjust pH to 7.4 Autoclave 20 min / 121 °C
1M IPTG	2.4 g isopropyl β-D-1-thiogalactopyranoside (IPTG) Water to make 10 mL of solution	Sterile filter (0.22 μm) Store at -20 °C in 500 μL aliquots
80 % glycerol	80 mL glycerol Water to make 100 mL of solution	Autoclave 20 min / 121 °C
1% SDS	0.5 g sodium dodecyl sulfate Water to make 50 mL of solution	Sterile filter (0.22 μm)
1% Tween 20	0.5 mL Tween 20 Water to make50 mL of solution	Sterile filter (0.22 μm
Phosphate Buffer Saline (PBS)	5 PBS tablets Water to make 500 mL of solution	Adjust pH to 7.4 Autoclave 20 min / 121 °C
Phage Wash Buffer (PWB)	25 μL Tween 20 PBS to make 50 mL of solution	Adjust pH to 7.4 Autoclave 20 min / 121 °C
LB Agar	40 g LB broth with agar Water to make 1 L solution	Adjust pH to 7.4 Autoclave 20 min / 121 °C Cool to 50 °C in water bath Add 500 μL sterile 10% carbenicillin Pour plates (15 mL) and store at 4 °C
LB broth	10 g tryptone 5 g yeast extract 10 g NaCl Water to make 1 L solution	Adjust pH to 7.4 Autoclave 20 min / 121 °C
LB Agarose	5 g tryptone 2.6 g yeast extract 2.6 g NaCl 3 g agarose Water to make 500 mL of solution	Adjust pH to 7.4 Autoclave 20 min / 121 °C
1×TAE	40 mL 25× TAE Water to make 1L of solution	Use immediately
PCR Master Mix	796 μL nuclease free water 100 μL CoralLoad PCR buffer (10×) 20 μL dNTPs (10 mM each dNTP) 20 μL T7seqF primer (10 μM) 20 μL T7seqR primer (10 μM)	Store frozen at -20 °C Defrost and add 5 μL *Taq* DNA polymerase before use Add 1 μL phage lysate to 24 μL PCR master mix
DNA Fingerprint Mix	240 μL nuclease free water 50 μL CutSmart buffer 10 μL Hinf*I* enzyme (1000 U/mL)	Store at 4 °C for up to 1 week

3.3.3 Equipment

Bacterial cultures were incubated in a heated orbital shaker (Ratek, Australia). Optical densities were recorded in 1 cm polystyrene cuvettes using a SmartSpec Plus UV spectrophotometer containing a 600 nm filter (BioRad, USA). Solutions were centrifuged with a Jouan CR312 refrigerated centrifuge. DNA was amplified with an Eppendorf Nexus GSX1 Thermal Cycler (ThermoFisher Scientific, USA). DNA sequencing was performed by Macrogen (Korea). Agarose gel electrophoresis was performed using a Mini-Sub Cell GT system (BioRad, USA) and gels were visualized via SYBR® safe DNA gel stain with a Safe

Imager 2.0 blue light transilluminator (ThermoFisher Scientific, USA). Water was purified using a Mili-Q Ultrapure Water Purification System (Millipore, USA).

3.3.4 Preparation of bacterial cultures

A stock of *E. coli* strain BLT5615 was stored at -80 °C in 10% glycerol stock. After thawing, an initial culture was prepared by spreading 25 µL of this stock onto an LB agar plate and incubating the plate at 37 °C for 16 h. The plate was stored at 4 °C and used to inoculate subsequent cultures for up to three weeks.

A saturated overnight culture of *E. coli* BLT5615 was prepared by inoculating 20 mL of M9TB with a single bacterial colony from an LB agar plate and then incubating at 37 °C for 12 h with gentle swirling (120 rpm).

A fresh culture of BLT5615 cells ready for T7 phage infection was prepared by inoculating 100 mL of M9TB with 2.5 mL of saturated overnight culture and incubating at 37 °C with vigorous shaking (260 rpm) until an OD_{600} of 0.4 was reached (2–3 h). 100 µL of 1M IPTG were added and incubation continued for further 30 min. The culture was then stored on slushy ice for up to 24 h until required.

3.3.5 Growth of T7 library lysates

100 mL of IPTG-treated BLT5615 cells were infected with 1 µL of a T7Select cDNA library and incubated at 37 °C with vigorous shaking (260 rpm) until lysis had occurred (1–2 h), as indicated by a marked decrease in OD_{600}. Immediately following lysis, 20 mL of the lysate were centrifuged at 5000 rpm for 10 min at 4 °C to precipitate cellular debris. The supernatant was decanted into a clean tube containing 20 µL of 1% Tween 20. The clarified lysate containing 0.01% Tween 20 was stored on slushy ice until required.

3.3.6 Functionalization of microtiter strip plates with biotinylated probes

1 mM stock solutions of the biotinylated control probe and affinity probe were prepared in DMSO and stored at −80 °C. The stock solutions were then diluted 1:100 with PBS to obtain 10 μM solutions (1% DMSO).

All microtiter plate wells were pre-incubated with 250 μL of PBS for 1 h at room temperature before use. PBS was discarded and the wells incubated with 200 μL of 10 μM solution of control or affinity probe for 1 h at room temperature. The solution was removed and the wells washed five times with 250 μL of PBS and stored with PBS at 4 °C until required.

3.3.7 Biopanning

250 μL of clarified T7 phage lysate were added to one well of a NeutrAvidin-coated plate functionalized with the biotinylated control probe. After 1 h of incubation at room temperature, the lysate was transferred to a well of the second plate functionalized with the biotinylated affinity probe. After 2 h of incubation at room temperature, the lysate was discarded, and the well washed once with 250 μL of PWB (< 5 s). Bound phage particles were eluted incubating with 100 μL of 1% SDS for 30 min at room temperature. Finally, the elute was diluted with 900 μL of 2×YT and stored at 4 °C overnight; during this time a portion of the SDS precipitated to the bottom of the tube.

The following day, 20 μL of the elute (1:10 in 2×YT) were removed, taking care not to disturb any precipitated SDS. This elute aliquot was used to infect 20 mL of fresh IPTG-treated BLT5615 cells for the next round of selection. This procedure was repeated until 8 rounds of selection had been completed. The stringency of the washing step was increased with each successive round of selection from 1 to 20 aliquots of 250 μL PWB.

3.3.8 Titering

LB agar plates were pre-warmed at 37 °C for 30 min. LB agarose was completely molten in a microwave oven and allowed to cool to 50 °C. 3 mL of the cooled agarose solution was combined with 250 µL of IPTG-treated BLT5615 cells, and the mixture was poured onto one pre-warmed LB agar plate. The uncovered plate was left to stand at room temperature for 30 minutes to allow the agarose to set completely. The phage elute retained from each round of selection was serially diluted with 2×YT by 100 to a million fold in a flexible 96-wel assay plate. A 2 µL aliquot of each dilution from each round of selection was dropped onto the surface of the solidified agarose using a multi-channel micropipette (8 × 6 array). The uncovered plate was left to stand at room temperature until the drops had absorbed completely into the agarose. The plate was incubated for 2–3 hours at 37 °C until plaques were clearly visible against the lawn of bacteria. The particular phage dilution from each selection round that contained a countable number of plaques (5–50) was used to calculate the phage titer.

3.3.9 Picking plaques

From the previous titering experiment, the volume required to give 50 plaques for the specific round of interest was calculated. LB agarose was melted in a microwave oven and allowed to cool to 50 °C. A 3 mL solution of the cooled agarose was combined with 250 µL of IPTG-treated BLT5615 cells and the calculated amount of the diluted phage lysate. The mixture was poured onto one pre-warmed LB agar plate, and the plate was allowed to set for 30 min. The plate was then incubated for 2–3 hours at 37 °C until plaques were clearly visible against the lawn of bacteria. Individual plaques were collected by stabbing the centre of each plaque with a 10 µL micropipette tube and transferring the tip to 100 µL of IPTG-treated BLT5615 cells in a 96-well microtiter plate. After the required number of plaques had been picked, the microtiter plate was incubated at 37 °C until complete lysis of the bacterial cells in each well

was observed (2–3 h). 45 µL of each lysate were transferred into a clean 96-well microtiter plate containing 5 µL of 80% glycerol and stored at -80 °C until required.

3.3.10 Amplification, sequencing, and fingerprinting of cDNA inserts

24 µL of PCR master mix were added to 1 µL of phage lysate and the resulting solution was subjected to 30 rounds of thermocycling using the protocol shown in **Table 3.4.** A 3 µL aliquot of amplified DNA solution was then incubated with 3 µL of the DNA fingerprint mix at 37 °C for 1 h.

Table 3.4 Standard thermocycler program for PCR of cDNA inserts

Cycles	Temperature	Time
1	94 °C	3 min
30	94 °C	30 s
	57 °C	30 s
	72 °C	45 s
1	72 °C	10 min
1	4 °C	10 min

3.3.11 Gel electrophoresis

1.5 g of agarose was suspended in 100 mL of 1× TAE buffer, and the suspension was boiled in a microwave oven until the agarose had dissolved completely. The 1.5% agarose solution was cooled to 50 °C and 10 µL of DNA gel stain were added. The solution was mixed thoroughly, poured into a casting tray (15 × 10 cm) containing two 26 well combs and allowed to set for 20 min at room temperature. The solidified gel was then transferred to a gel tank, flooded with 1× TAE, and the combs were removed. For gel electrophoresis of PCR products from single plaques picked from the round of interest, the agarose concentration was increased to 2%. 3 µL of undigested and 5 µL of digested amplified DNA were loaded onto

the gel with a micropipette. After all samples had been loaded, the gel was run at 96 V for 35 min. The well was then removed from the tank and visualized using a blue light transilluminator.

3.3.12 DNA sequencing

A 15 µL aliquot of PCR-amplified DNA was purified using a GenElute™ PCR clean-up kit following the manufacturer's instructions. The purified DNA was then centrifuged at 13,000 rpm for 5 min at 25 °C to precipitate any resin that had carried through from the GenElute™ column. Finally, a 10 µL aliquot of the purified DNA was submitted for DNA sequencing (Macrogen, Korea) along with T7SeqF primers.

3.4 REFERENCES

1. Spring, D. R., Chemical genetics to chemical genomics: small molecules offer big insights. *Chem. Soc. Rev.* **2004,** *34*, 472-482.
2. Zanders, E. D., Overview of chemical genomics and proteomics. *Methods Mol. Biol.* **2012,** *800*, 3-10.
3. Palcy, S.; Chevet, E., Integrating forward and reverse proteomics to unravel protein function. *Proteomics* **2006,** *6*, 5467-5480.
4. Sergeeva, A.; Kolonin, M. G.; Molldrem, J. J.; Pasqualini, R.; Arap, W., Display technologies: Application for the discovery of drug and gene delivery agents. *Adv. Drug Deliv. Rev.* **2006,** *58*, 1622-1654.
5. Paschke, M., Phage display systems and their applications. *Appl. Microbiol. Biotechnol.* **2006** *70*, 2-11.
6. Rosenberg, A.; Griffin, G.; Studier, F. W.; McCormick, M.; Berg, J.; Mierendorf, R., T7Select Phage Display System: A Powerful New Protein Display System Based on Bacteriophage T7. *inNovations* **1996,** *6*, 1-6.
7. Novagen, *T7Select System Manual User Protocol TB178*. Rev. D 0311JN, USA, 2011.
8. Novagen, *Pre-Made T7Select Libraries User Protocol* TB270.
9. Savinov, S. N.; Austin, D. J., The Cloning of Human Genes Using cDNA Phage Display and Small-Molecule Chemical Probes. *Comb. Chem. High Throughput Screening* **2001,** *4*, 593-597.
10. Piggott, A. M.; Karuso, P., Identifying the cellular targets of natural products using T7 phage display. *Nat. Prod. Rep.* **2016,** *33*, 626.
11. Van Dorst, B.; Mehta, J.; Rouah-Martin, E.; Blust, R.; Robbens, J., Phage display as a method for discovering cellular targets of small molecules. *Methods* **2012,** *58*, 56-61.
12. Takakusagi, Y.; Takakusagi, K.; Sugawara, F.; Sakaguchi, K., Use of phage display technology for the determination of the targets for small-molecule therapeutics. *Exp. Opin. Drug Discov.* **2010,** *5* (4), 361-389.
13. Sche, P. P.; McKenzie, K. M.; White, J. D.; Austin, D. J., Display cloning: functional identification of natural product receptors using cDNA-phage display. *Chem. Biol.* **1999,** *6* (10), 707-716.
14. Jin, Y.; Yu, J.; Yu, Y. G., Identification of hNopp140 as a Binding Partner for Doxorubicin with a Phage Display Cloning Method. *Chem. Biol.* **2002,** *9*, 157-162.
15. Shim, J. S.; Lee, J.; Park, H.-J.; Park, S.-J.; Kwon, H. J., A New Curcumin Derivative, HBC, Interferes with the Cell Cycle Progression of Colon Cancer Cells via Antagonization of the Ca2+/Calmodulin Function. *Chem. Biol.* **2004,** *2004*, 1455-1463.
16. Boehmerle, W.; Splittgerber, U.; Lazarus, M. B.; McKenzie, K. M.; Johnston, D. G.; Austin, D. J.; Ehrlich, B. E., Paclitaxel induces calcium oscillations via an inositol 1,4,5-trisphosphate receptor and neuronal calcium sensor 1-dependent mechanism. *Proc. Nat.l Acad. Sci. USA* **2006,** *103* (48), 18356-18361.

17. Aoki, S.; Morohashi, K.; Sunoki, T.; Kuramochi, K.; Kobayashi, S.; Sugawara, F., Screening of Paclitaxel-Binding Molecules from a Library of Random Peptides Displayed on T7 Phage Particles Using Paclitaxel-Photoimmobilized Resin. *Bioconjug. Chem.* **2007,** *18*, 1981-1986.

18. Kim, H.; Deng, L.; Xiong, X.; Hunter, W. D.; Long, M. C.; Pirrung, M. C., Glyceraldehyde 3-Phosphate Dehydrogenase Is a Cellular Target of the Insulin Mimic Demethylasterriquinone B1. *J. Med. Chem.* **2007,** *50*, 3423-3426.

19. Piggott, A. M.; Karuso, P., Rapid Identification of a Protein Binding Partner for the Marine Natural Product Kahalalide F by Using Reverse Chemical Proteomics. *Chembiochem* **2008,** *9*, 524-530.

20. Van Dorst, B.; Mehta, J.; Rouah-Martin, E.; Somers, V.; De Coen, W.; Blust, R.; Robbens, J., cDNA phage display as a novel tool to screen for cellular targets of chemical compounds. *Toxicol. In Vitro* **2010,** *24*, 1435-1440.

21. He, Q.-L.; Jiang, H.; Zhang, F.; Chen, H.-B.; Tang, G.-L., Simultaneous identification of multiple receptors of natural product using an optimized cDNA phage display cloning. *Bioorg .Med. Chem. Lett.* **2008,** *18*, 3995-3998.

22. Jung, H. J.; Shin, J. S.; Lee, J.; Song, Y. M.; Park, K. C.; Choi, S. H., . . . Kwon, H. J., Terpestacin Inhibits Tumor Angiogenesis by Targeting UQCRB of Mitochondrial Complex III and Suppressing Hypoxia-induced Reactive Oxygen Species Production and Cellular Oxygen Sensing. *J. Biol. Chem.* **2010,** *285*, 11584-11595.

23. Manita, D.; Toba, Y.; Takakusagi, Y.; Matsumoto, Y.; Kusayanagi, T.; Takakusagi, K., . . . Sugawara, F., Camptothecin (CPT) directly binds to human heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) and inhibits the hnRNP A1/topoisomerase I interaction. *Bioorg. Med. Chem.* **2011,** *19*, 7690-7697.

24. Izaguirre-Carbonell, J.; Kawakubo, H.; Murata, H.; Tanabe, A.; Takeuchi, T.; Kusayanagi, T., . . . Sugawara, F., Novel anticancer agent, SQAP, binds to focal adhesion kinase and modulates its activity. *Sci. Rep.* **2015,** *5*, 15136.

25. Gotsbacher, M. P.; Cho, S. M.; Kim, N. H.; Liu, F.; Kwon, H. J.; Karuso, P., Reverse Chemical Proteomics Identifies an Unanticipated Human Target of the Antimalarial Artesunate. *ACS Chem. Biol.* **2019,** *14*, 636-643.

26. Becker, F.; Murthi, K.; Smith, C.; Come, J.; Costa-Roldan, N.; Kaufmann, C., . . . Kley, N., A Three-Hybrid Approach to Scanning the Proteome for Targets of Small Molecule Kinase Inhibitors. *Chem. Biol.* **2004,** *11*, 211-233.

27. Chidley, C.; Haruki, H.; Gronlund Pedersen, M.; Muller, E.; Johnsson, K., a yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nat. Chem. Biol.* **2011,** *7*, 375-383.

28. Klaeger, S.; Heinzlmeir, S.; Wilhelm, M.; Polzer, H.; Vick, B.; Koenig, P.-A., . . . Kuster, B., The target landscape of clinical kinase drugs. *Science* **2017,** *358*, 1148.

29. Salazar Estrada, I. J. Molecular tools for elucidating signaling networks of the protein kinase EGFR. Macquarie University, Sydney, Australia, 2015.

30. Pommier, Y., DNA Topoisomerase I Inhibitors: Chemistry, Biology and Interfacial Inhibition. *Chem. Rev.* **2009,** *109* (7), 2894-2902.

31. Chauhan, M.; Sharma, G.; Joshi, G.; Kumar, R., Epidermal Growth Factor Receptor (EGFR) and its Cross-talks with Topoisomerases: Challenges and Opportunities for Multi-Target Anticancer Drugs. *Curr. Pharm. Des.* **2016,** *22*, 3226-3236.

32. Peleg, R.; Bobilev, D.; Priel, E., Topoisomerase I as a target of erlotinib and gefitinib: Efficacy of combined treatments with camptothecin. *Int. J. Oncol.* **2014,** *44*, 934-942.

33. Lomenick, B.; Olsen, R.; Huang, J., Identification of Direct Protein Targets of Small Molecules. *ACS Chem. Biol.* **2010,** *6* (1), 34-46.

34. Chauhan, M.; Joshi, G.; Kler, H.; Kashyap, A.; Amrutkar, S. M.; Sharma, P., . . . Kumar, R., Dual inhibitors of epidermal growth factor receptor and topoisomerase IIa derived from a quinoline scaffold. *RSC Adv.* **2016,** *6*, 77717-77734.

CHAPTER 4

Forward Chemical Proteomics Studies

with Covalent EGFR Probes

4.1 INTRODUCTION

4.1.1 Forward chemical proteomics with covalent click probes

The use of covalent click probes has expanded the potential of non-directed global proteomics studies and traditional probe-immobilized-based chemical proteomics.[1-3] The combination of a click handle and a reactive group for covalent labelling provides the probe unique properties that allow the interrogation of a protein of interest in its endogenous context. Thus, information on the highly-regulated and compartmentalized small-molecule/protein interactions *in situ* can be obtained.[4-5]
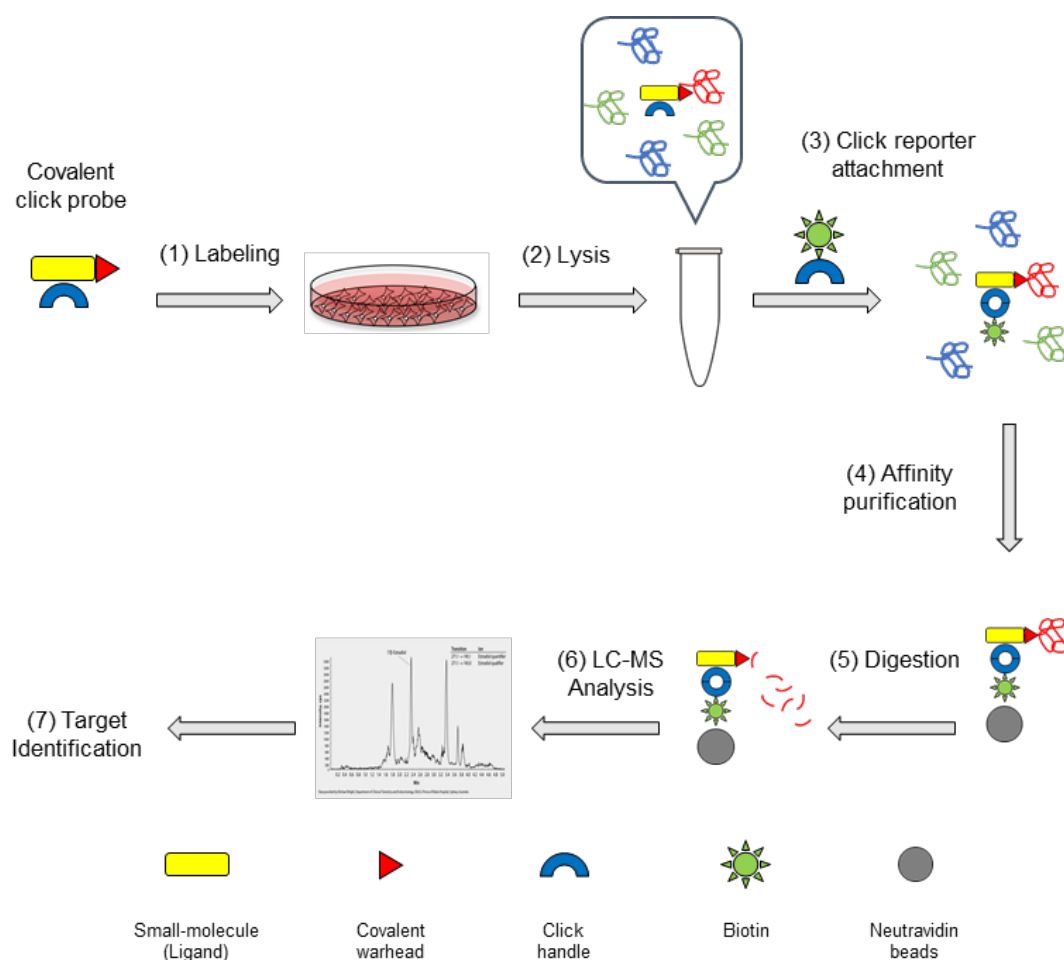
Scheme 4.1 Workflow of forward chemical proteomics with covalent click probes

Chemical proteomics studies with covalent click probes are based on the well-established methodology of ABPP (**Scheme 4.1**).[6-7] Experiments start with the labelling of live cells *in situ* with the probe of interest (**Scheme 4.1**, step 1). The inclusion of control experiments

such as competition with free drug or treatment with linker at this stage facilitates the identification of true targets from non-specific background. Quantitative isotope-labelled methods (e.g. SILAC, ICAT, ITRAQ) are also commonly used at this or later stage, but quantitative label-free methods can also be used. Once the labelling has occurred, cells are lysed (**Scheme 4.1** step 2) and the reporter of interest is attached to the click handle under click chemistry conditions, usually with CuAAC (Cooper(I)-catalyzed azide-alkyne cycloaddition, **Scheme 4.1**, step 3). For proteomics analysis, a biotin tag is typically attached; fluorophores can also be attached to allow visualization of labeled proteins in gel after SDS-PAGE separation. Biotinylated proteins are then captured on neutravidin beads for the unbound and non-specific proteins to be washed away (**Scheme 4.1**, step 4). Enriched proteins can then be digested directly on the beads (**Scheme 4.1**, step 5) or by other methods after elution from the beads. Finally LC-MS/MS analysis allows the identification of labeled proteins (**Scheme 4.1**, steps 6 and 7).

4.1.2 EGFR-targeted chemical proteomics with covalent click probes

In 2014, Cravatt et al reported a proteome-wide analysis of the selectivity of the irreversible kinase inhibitors Afatinib and Ibrutinib.[8] To this end, they modified Afatinib into the covalent click probe **P5** (**Fig. 4.1a**), which was then used in quantitative isotope-labelled (SILAC) chemical proteomics analysis of A431 cells. A431 is a model epidermoid carcinoma cell commonly used in EGFR studies due to EGFR overexpression, with approximately 3.5×10^6 EGFR copies per cell.[9] Fluorescent SDS-PAGE analysis of the membrane fraction labeled proteome allowed the visualization of labeled proteins (**Fig. 4.1b**). EGFR labelling was observed at low nM concentrations; the labelling was confirmed to be specific as pre-treatment with the corresponding inhibitor **I5** (**Fig 4.1a**) outcompeted probe labelling. In addition to EGFR, the concentration-dependant labelling of other proteins was also observed, and many of these were not competed by the inhibitor **I5**. This suggested that the labelling of

these proteins was mediated by the electrophilic warhead independently of ligand binding. Chemical proteomics analysis allowed the identification of 18 targets enriched by the probe **P5** in comparison with a DMSO blank control as determined by their SILAC ratio (**Fig. 4.1c**). Half of these targets, including the previously known targets EGFR and ErbB2 were outcompeted by inhibitor **I5** or labeled selectivity by Afatinib but not Ibrutinib probe, and therefore classified as specific.
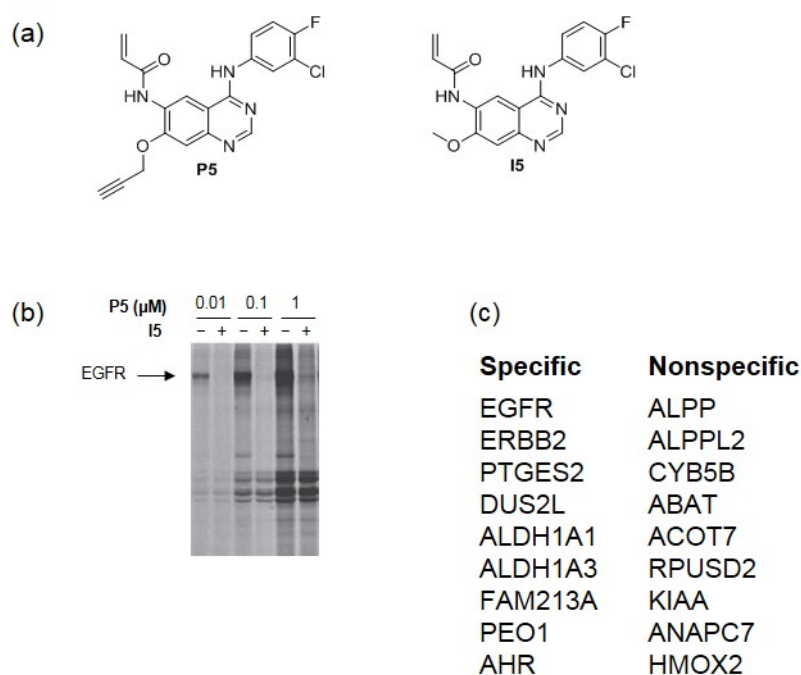
Fig. 4.1 Summary of results for EGFR chemical proteomics studies reported by Cravatt et al.[8]
(Adapted from Lanning, B. R. et al *Nat. Chem. Biol.* **2014**, *10*, 760-767).

Interestingly, **P3** specific targets were not limited to kinases but included other enzymes (e.g. DUS2L, PTGES2, ALDH1A1), receptors (e.g. AHR) and uncharacterized proteins (e.g. FAM213A). Although these proteins did not share any structural or functional similarity with kinases, they all present conserved catalytic or functional cysteines, suggesting the importance of the electrophilic warhead in their capture. The time and concentration dependent off-reactivity of probe **P5** was then associated to the failure of Afatinib to effectively inhibit T790M EGFR: at the concentrations required for T790M EGFR inhibition, cell death occurs independently of EGFR activity and is mediated by this off-target reactivity.

The same methodology was later used for profiling the off-target reactivity of third-generation EGFR inhibitors, including the FDA-approved Osimertinib.[10] Osimertinib was found to target cathepsins *in vivo* and *in situ* but not *in vitro*. This was attributed to lysosomal accumulation of the drug in endogenous conditions, highlighting the advantages of covalent click probes over traditional kinobeads methods in maintaining endogenous interactions. Together, these cases support the importance of characterizing the quality of chemical probes and the experimental conditions used in a specific context.[11-13]

4.1.3 Chemical proteomics beyond cysteine targeting

In 2014, Hamachi et al reported the identification of Lapatinib off-targets in live cells using a different warhead, a ligand-directed tosyl probe (**Fig. 4.2**).[14] In this approach, the drug of interest (ligand) is connected to a reporter group (in this case, click handle) through a tosylate group. Ligand binding to the protein of interest allows the labelling of nucleophilic residues in close proximity to the tosylates group by an S_N2 reaction, which results in the endogenous covalent modification of the target protein, followed by cleavage of the ligand (**Fig. 4.2a**). Six Lapatinib derivative click probes (**P6, Fig. 4.2b**) with varied patterns of substitution at the tosyl unit and different linker length were used for profiling the drug interactome in live NCI-N87 cells, a human gastric cancer cell line endogenously overexpressing HER2. After probe labelling, the lysate was reacted with biotin-azide and labeled proteins visualized by WB against biotin (**Fig. 4.2c**). The labelling efficiency was shown to be highly dependent on the linker length. The highest level of HER2 labelling was observed with probe **P6-1p**, in which the reactive group is directly attached to the quinazoline core (**Lane 4, Fig. 4.2c**). This suggested that labelling had occurred in the kinase domain. Interestingly, the probe **P6-2m** with a longer linker length shifted the selectivity of the probe from HER2 to second off-target with a MW of 55kDa (**Lane 6, Fig. 4.2c**). Chemical proteomics analysis allowed the identification of this off-target as protein disulfide isomerase. Previous studies by the same

group using a KFPB-directed tosyl probe had also shown the importance of the linker structure in controlling the proximity effect and consequently the labelling properties of the probe.[15] Compared with Cravatt's probe **P5**, indiscriminate warhead-mediated labelling was not significant, which can be attributed to the lower reactivity of the tosylate group. However, this is also a disadvantage, as significant labelling of HER2 requires long incubation times (up to eight hours for this case). These cases support the idea that it is possible to achieve good selectivity of covalent probes targeting non-cysteine residues with a carefully structure-guided designed probe.
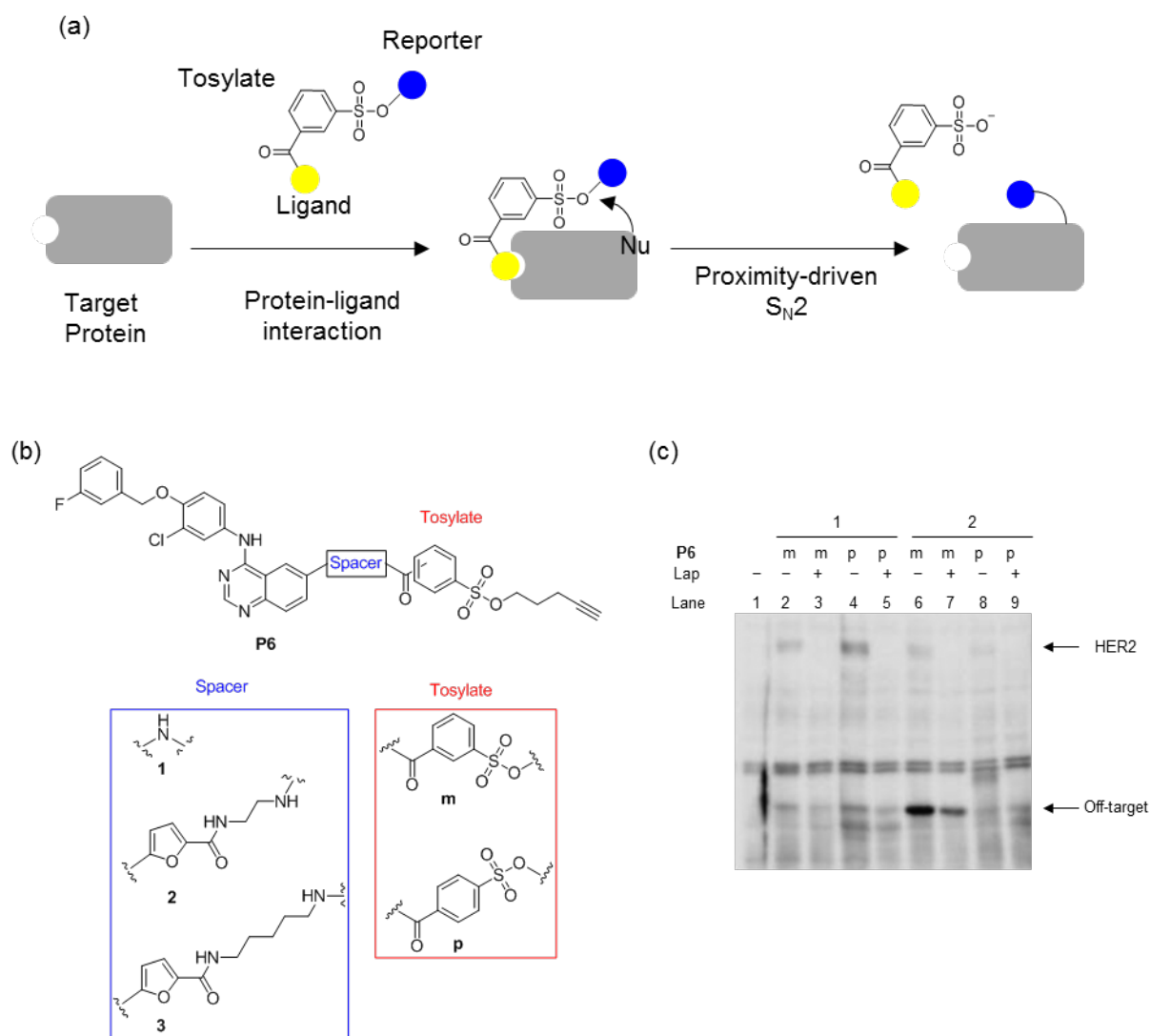
Fig. 4.2 Overview of identification of Lapatinib off-targets in situ reported by Hamachi et al.[14] (adapted from *Chem. Commun.* **2014**, *50*, 14097-14100).

4.1.4 MDA-MB-468 cells as a model for EGFR studies in Triple Negative Breast Cancer (TNBC)

TNBC comprises a series of highly heterogeneous cancers, each with unique molecular characteristics.[16] Several cell lines have been established as models of TNBC; among them, MDA-MB-468 has been extensively used for EGFR studies.[17-18] This cell line was first isolated in 1977 by Cailleau et al from a pleural effusion of a 51-year-old Black female patient with metastatic adenocarcinoma of the breast.[19] MDA-MB-468 cells belong to the basal A subtype of breast cancer,[20-21] which is characterized by a high expression of proliferation and DNA damage response genes with expression of basal markers (e.g. cytokeratins, integrins); this last feature is characteristic of a more differentiated state compared with the basal B type, characterized by a more mesenchymal-like appearance. EGFR mutations are not observed in MDA-MB-468 cells, but *EGFR* gene amplification originates EGFR protein overexpression (ca. 1×10^6 copies per cell, similar to A431 cells).[22]

Numerous studies with MDA-MB-468 have shed light on the molecular mechanisms of EGFR activity in specific subcellular localization contexts. For example, despite EGFR overexpression, EGF treatment of MDA-MB-468 cells promotes apoptosis and inhibition of apoptosis.[23] This counterintuitive observation resembles a trend in which EGFR activity shifts from oncogenic in primary BC tumours to tumour suppressor in metastatic states.[24] EGFR apoptosis-promoting activity has been shown to be initiated by endosomal EGFR but not plasma membrane EGFR (the latter promoting cell growth).[25] In MDA-MB-468, EGFR accumulation in early endosomes is possible due to slow endocytic trafficking and low receptor degradation.[26] Recently, EGFR pro-apoptotic activity was also associated with the ability of the receptor to traffic to the nucleus and phosphorylate STAT1; selective inhibition of STAT1 in the nucleus restored EGFR-mediated cell proliferation in response to EGF treatment.[27] In 2014, Wheeler et al showed that MDA-MB-468 cells express high levels of

nuclear EGFR, and nuclear trafficking was triggered by SFK-mediated phosphorylation of Y1101 in EGFR.[28] Treatment with the SFK inhibitor Dasatinib impeded EGFR nuclear translocation, shifting the equilibrium towards plasma membrane localization and sensitizing the cells to Cetuximab treatment. In light of the previous reports, this might due in part to the proliferation signalling characteristic of cell surface EGFR.

In addition to these localization-specific information, recent studies have shed light on the metabolic characteristics of TNBC.[29] In MDA-MB-468 cells, EGFR accelerates glucose uptake and lactate production.[30] Proteomics studies allowed the identification of pyruvate kinase M2 (PKM2) as a direct interactor and kinase substrate of EGFR in response to EGF treatment.[31] PKM2 regulates the last step of glycolysis (dephosphorylation of phosphoenolpyruvate to pyruvate); PKM2 phosphorylation at Y148 by EGFR reduced PKM2 activity promoting cell proliferation, glycolysis and lactate production. In addition to this direct interaction, EGF treatment also reduced the levels of mRNA and protein expression of hexokinase 2 (HK2, catalyzes the first step in glycolysis, glucose phosphorylation) and its enzymatic activity. Upregulation of HK2 and downregulation of PKM2 resulted in the accumulation of glycolytic metabolites; among these, fructose 1,6-biphosphate (F1,6BP) was shown to enhance EGFR phosphorylation, probably through direct EGFR binding. Combinatorial targeting of EGF signalling and glycolysis showed a synergistic effect in blocking cell proliferation.

Together, these cases demonstrate the importance of the localization and time specific cellular context in determining EGFR functions and phenotype, and the potential identification of combinatorial therapies capable to overcome drug resistance based on endogenous EGFR interactions. They also validate the use of MDA-MB-468 as a model cell line for TNBC.

4.1.5 Overview of chemical proteomics studies with newly synthesized probes

Given the high expression of EGFR and the extensive information on EGFR interactors, as well as the need to validate new molecular targets in TNBC breast cancer, MDA-MB-468 cells were chosen as a model system for this forward chemical proteomics study. An isotope-label free forward chemical proteomics method was used with the previously reported EGFR-directed cysteine-targeting probe **P5**[8] and the newly designed EGFR-directed lysine-targeting probe **P1** (**Fig. 4.3**). The corresponding inhibitors **I5** and **I1**, as well as the non-directed NHS probes **P7** and **P8** were used as controls to facilitate the identification of specific (drug-mediated) and non-specific (warhead-mediated) targets.
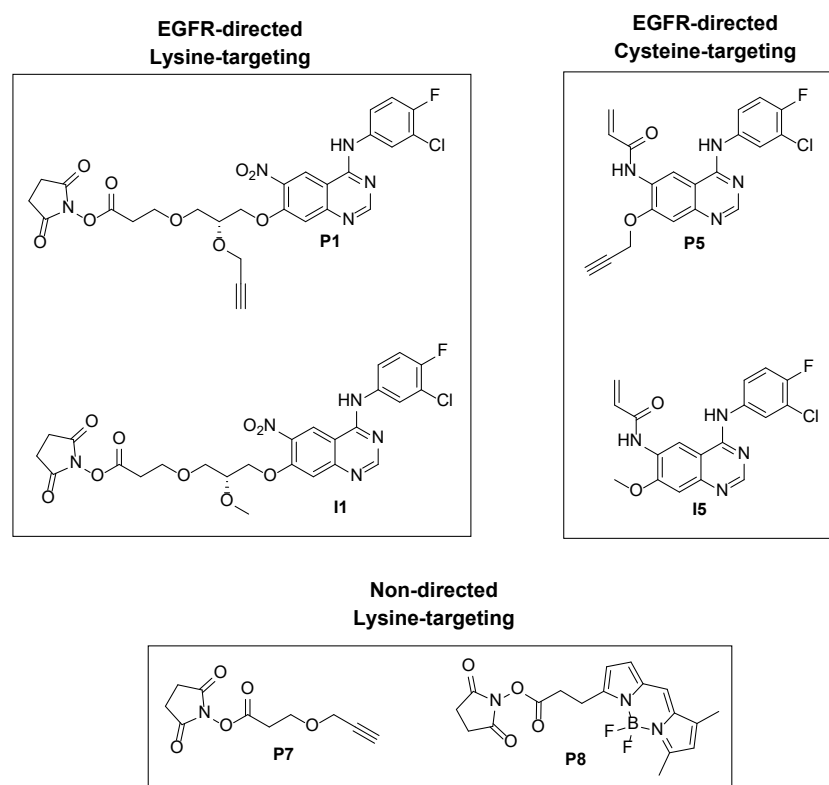
Fig. 4.3 Structure of probes used in forward chemical proteomics studies

A label-free quantification method based on the the normalized spectral abundance factor (NSAF)[32-33] in combination with a negative DMSO and corresponding control inhibitors or probes was used to identify probe-targeted proteins (**Scheme 4.2**). To this end, each experimental point consisted of at least four experiments: 1) a positive experiment consistent

of treatment with the corresponding probe (**P1** or **P5**); 2) a competed experiment in which cells were pre-incubated with inhibitor (**I1** or **I5**) followed by treatment with the corresponding inhibitor (to block labelling of probe-specific targets); 3) a negative DMSO control subject to same treatment as 1 and 2 (bead incubation and on-bead digestion; 4) a second DMSO negative control directly subjected to in-solution digestion. Proteins appearing in the DMSO-bead control (3) or whose NSAF and spectral count ratio between probe/DMSO was less than two, as well as proteins containing only one unique peptide were assigned as background proteins. The remaining proteins were assigned as probe targets. Probe targets whose NSAF and spectral count probe/inhibitor ratio was greater than two were classified as specific (inhibitor competed, drug-mediated), and those below this ratio were considered non-specific (warhead-mediated).
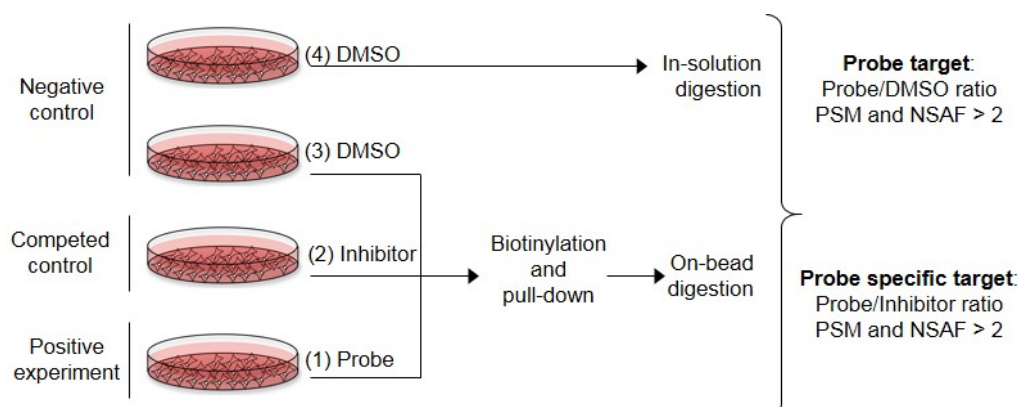
Scheme 4.2 Overview of experimental design for the identification of probes targets

Besides MDA-MB-468 cells, a colorectal cancer HCT 116 cell line was used to gauge the labelling proficiency of probe **P5** at physiological levels of EGFR, since this cell line does not overexpress EGFR (1.5×10^5 copies per cell).[34] Probe labelling was also combined with nuclear fractionation to assess the efficiency of the probes in capturing endogenous targets in specific subcellular compartments.

4.2 RESULTS AND DISCUSSION

Chemical proteomics analysis were performed following the method previously reported by Cravatt et al with slight modifications.[8] The main difference consisted on the use of a quantitative isotope-labelled free approach instead of the commonly used SILAC method. Each of these methods have specific advantages and disadvantages. The isotope-label free approach was chosen as it can provide faster, cleaner and simpler quantification results.[35] Lysis and fractionation of labelled cells was also modified. Instead of membrane fractionation by ultracentrifugation, the whole cell lysate was obtained in RIPA buffer (in both cases lysis was done by sonication). A nuclear fractionation was also used in a second set of experiments (section 4.2.4). All other steps were performed without change.

4.2.1 Labelling of HCT 116 cells with probe P5

HCT 116 cells, which express low levels of EGFR, were first used as a control experiment to test the performance and limitations of the previously reported probe **P5**. The cells were incubated with **P5** at a final concentration of 5 μM at 37 °C for one hour followed by sonication in RIPA buffer. The labelled lysate was then "clicked" with biotin-azide; biotinylated proteins were pulled down on neutravidin beads, subjected to on-bead digestion and analysed by LC-MS/MS.

Despite the difference of an order of magnitude in EGFR expression in comparison with A431 cells, EGFR was selectively labelled and enriched in HCT 116 cells by probe **P5**. In addition to EGFR, other probe **P5** specific targets previously observed in A431 cells such as ErbB2 and prostaglandin synthase 2 were also identified in HCT 116 cells (**Table 4.1**).[8] The specificity of these targets was confirmed by competition with the corresponding inhibitor **I5**. Among the specific target proteins, EGFR presented the highest number peptides: 18 unique peptides representing all the different protein domains and covering 24% of the total protein

sequence were observed. Due to the lower amount of EGFR, the peptide coverage in HCT 116 cells was lower compared to the one previously observed for A431 cells (49 peptides, 64% sequence). However, it is remarkable that this lower amount of EGFR can be labelled and enriched and by probe **P5** and detected by MS.

Table 4.1 Selected **P5** specific targets in HCT 116 cells.

Entry	UniProt ID	Gene	Name
1	P00533	*EGFR*	Epidermal growth factor receptor
2	P04626	*ERBB2*	Receptor tyrosine protein kinase ErbB2
3	Q9H727	*PGES2*	Prostaglandin E synthase 2
4	Q9NX74	*DUS2L*	tRNA-dihydrouridine(20) synthase [NAD(P)+]-like

4.2.2 Labelling of MDA-MB-468 cells with probe P5

After testing the performance of probe **P5** in HCT 116 cells, MDA-MB-468 cells were labelled with the same probe under similar conditions. Before processing the samples for pull-down and LC-MS/MS analysis, the labelled lysate was subjected to click chemistry with rhodamine-azide, followed by SDS-PAGE separation and visualization by fluorescent scanning (**Fig. 4.4a**). EGFR labelling was observed as a fluorescent band at around 180 kDa. Similarly to the previous report for A431 cells,[8] EGFR labelling was observed at submicromolar concentration and was competed by pre-incubation with the corresponding inhibitor **I5**. Non-specific, concentration-dependent labelling of several other proteins without being outcompeted by inhibitor **I5** was also observed. Under our experimental conditions, EGFR labelling was observed only when cells were labelled *in situ*, but not if labelling was done directly on the cell lysate (**Fig. 4.4b**), in which only non-specific labelling was observed. This suggests that under *in vitro* conditions labelling occurs mostly non-specifically due to intrinsic acrylamide reactivity rather than mediated by drug-target interactions. This might be

attributed to the denaturing conditions of the RIPA buffer reducing probe binding affinity for EGFR. The results observed by fluorescent gels were also reflected in chemical proteomics analysis, in which EGFR was detected only in samples treated *in situ*.
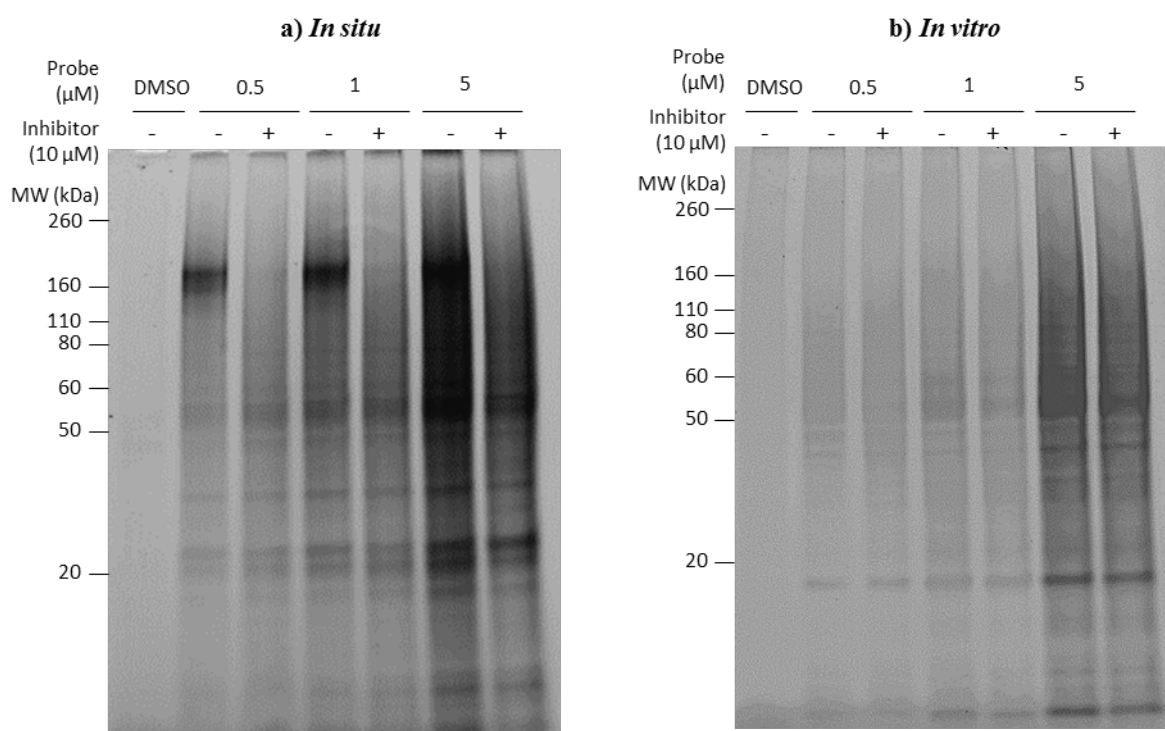
Fig. 4.4 Fluorescent gel of MDA-MB-468 cells labelled with probe **P5** a) *in situ* b) *in vitro*

A total of 703 proteins in the **P5** treated sample were identified by chemical proteomics. Among these, only 10 proteins were enriched by **P5** in comparison with a DMSO-treated whole cell lysate, as indicated by their differences in spectral counts (**Table 4.2**). The 10 enriched proteins were assigned as probe targets, while the rest of the proteins were assigned as background. Background proteins include various cytoskeletal proteins, histones, hnRNP proteins, heath shock proteins, ribosomal proteins and translation and initiation factors, all of which have been previously reported as common proteins binding non-specifically to sepharose beads.[36]

Table 4.2 Selected **P5** specific targets in MDA-MB-468 cells.

Entry	UniProt ID	Gene	Name
1	P00533	*EGFR*	Epidermal growth factor receptor
2	Q9H7Z7	PTGES2	Prostaglandin E synthase 2
3	P04150	NR3C1	Glucocorticoid receptor
4	Q8N2G8	GHDC	GH3 domain-containing protein
5	Q02318	CYP27A1	Sterol 26-hydroxylase, mitochondrial

From the 10 target proteins, five were competed by pre-incubation with inhibitor **I5**, and thus identified as specific probe targets (**Table 4.2**). EGFR was the target with greater abundance, comprising about 4% of the total proteins identified as indicated by the normalized spectral abundance factor (NSAF). Accordingly, EGFR also presented the highest coverage of protein sequence (46%) and number of unique peptides (43), which were slightly lower than those previously reported for A431 cells, but higher than the results for HCT 116 cells. After EGFR, prostaglandin E synthase was the second most abundant specific target, representing about 0.7% of identified proteins by its NSAF value. Prostaglandin E synthase was also enriched by probe **P5** in A431 and HCT 116 cells, suggesting that this off-target protein is commonly present among different cell lines. After EGFR, ErbB2 presents the second highest affinity for afatinib. Although ErbB2 had been observed in A431 and HCT 116 cells, it was not detected in MDA-MB-468, in accordance to the TNBC characteristics of this cell line. In addition to these common targets, other proteins not previously observed in A431 or HCT 116 cells were also identified, and these were not restricted to kinases. Despite the relatively high concentration of probe (5 μM) and the high number of background proteins identified, the number of specific targets was relatively low. This confirmed the utility of **P5** as a highly selective EGFR probe.

4.2.3 Labelling of MDA-MB-468 cells with probe P1

After the identification of probe **P5** targets in MDA-MB-468 cells, the newly designed lysine-targeting probe **P1** was finally tested in the same cell line. All previous *in situ* labelling with probe **P5** were done in serum free DMEM cell culture media. Under the same conditions, no labelling at all (not even non-specific) was observed by fluorescent SDS-PAGE gel or LC-MS/MS. However, labelling did take place *in vitro* (**Fig. 4.5b**). This suggested that the probe had been quenched when diluted in cell culture media. Therefore, PBS was used for *in situ* labelling with **P1** and all subsequent NHS-containing probes. A fluorescent gel of the cell lysate labelled *in situ* confirmed the stability of the probe in PBS allowing labelling to occur (**Fig. 4.5a**). Unlike probe **P5**, EGFR labelling with probe **P1** was interestingly not observed by fluorescent gel *in situ*, and none of the targeted proteins seemed to be competed by pre-incubation with inhibitor **I1**. As expected from the results with probe **P5**, no specific labelling was observed *in vitro* either.
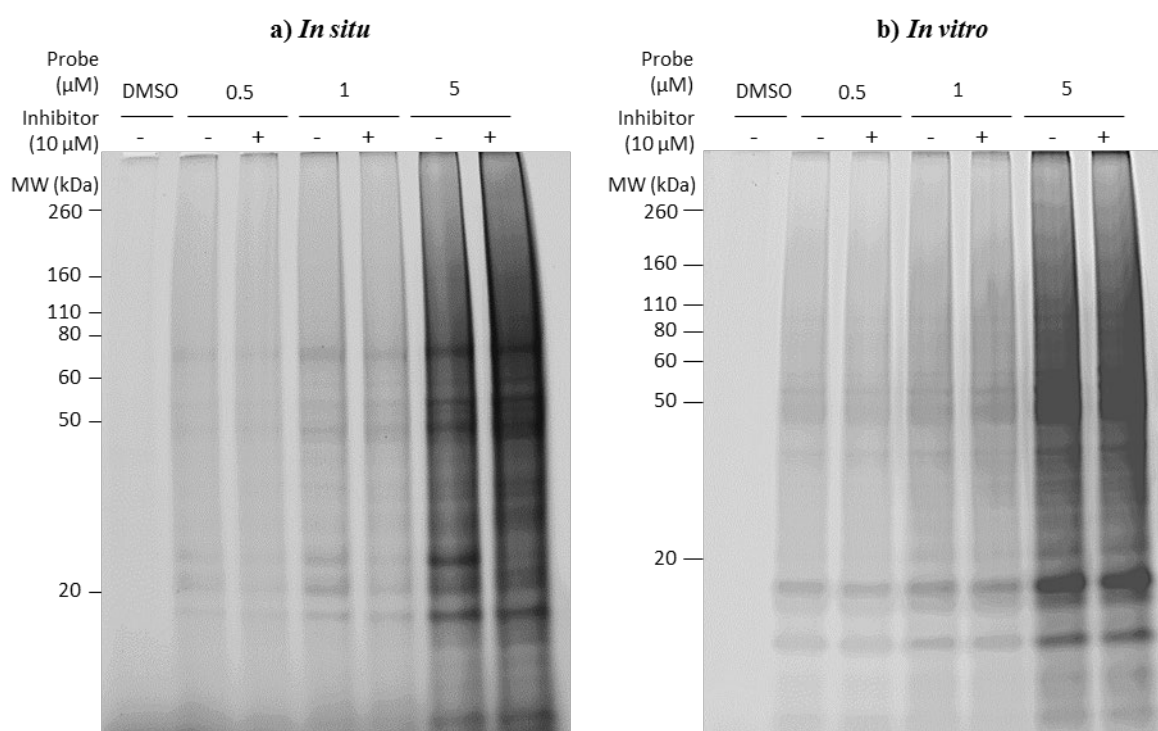
Fig. 4.5 Fluorescent gel of MDA-MB-468 cells labelled with probe **P1** a) *In situ* b) *In vitro*

Reflecting the results observed by fluorescent SDS-PAGE gels, EGFR was not identified as a target of probe **P1** in chemical proteomics analysis. However, seven other proteins were identified as specific targets of probe **P1** (**Table 4.3**), as they were competed by pre-treatment with inhibitor **I1**. None of these target proteins were common to other Afatinib off-targets previously identified for probe **P5** (e.g. PTGES2). Targeted proteins included several kinases (PFKL, PFKP, PIP4K2C), metabolic enzymes (FDFT1, SGPL1) and nucleotide binding proteins (GTPBP1). Among the identified specific target proteins, ATP-dependent 6-phosphofructokinase was the most abundant, with each type (liver and platelet) representing 1% of the total proteins by its NSAF value. The higher reactivity of the probe **P1** compared with **P5** was reflected in a higher number of nonspecific targets, with a 10-fold increase in the number of proteins for which labelling was not outcompeted by adding inhibitors.

Table 4.3 Selected **P1** specific targets in MDA-MB-468 cells.

Entry	UniProt ID	Gene	Name
1	P17858	PFKL	ATP-dependent 6-phosphofructokinase, liver type
2	Q01813	PFKP	ATP-dependent 6-phosphofructokinase, platelet type
3	P37268	FDFT1	Squalene synthase
4	Q8TBX8	PIP4K2C	Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma
5	Q9NRV9	HEBP1	Heme-binding protein 1
6	O00178	GTPBP1	GTP-binding protein 1
7	O95470	SGPL1	Sphingosine-1-phosphate lyase 1

Given the good selectivity of the 4-anilinoquinazoline scaffold for EGFR and the results observed for probe **P5** in HCT 116 and MDA-MB-468 cells, the absence of EGFR labelling with probe **P1** was initially surprising. Three different hypotheses were formulated below to explain the absence of EGFR labelling by probe **P1**: 1) probe quenching prevents EGFR binding; 2) structural bias for probe off-targets; 3) capture of EGFR interactors.

4.2.3.1 Hypotheses for the absence of EGFR labelling by probe P1

4.2.3.1.1 Probe quenching prevents EGFR binding

Probe **P1** presented a higher reactivity towards nucleophiles compared with the previously reported probe **P5**. This higher reactivity was reflected in the hydrolysis of **P1** in DMEM cell culture media, which resulted in a quenched probe that prevented any labelling to occur (unlike probe **P5**), as well as a higher number of non-specific probe targets. The high reactivity of the NHS ester can bias probe **P1** labelling from thermodynamic to kinetic control, effectively quenching the probe before binding to EGFR can occur. If this is the case, most of the **P1**-labelled proteins would be expected to be non-specific cell surface proteins that prevent the internalization of probe to the cell and consequent probe-EGFR binding. To test this hypothesis, MDA-MB-468 cells were pre-incubated with the fluorophore-NHS ester probe **P8** (**Fig. 4.3**) to preclear some of the reactive lysine sites followed by treatment with probe **P1**. Under these conditions, EGFR was not identified in the mass spectroscopic analysis. Unlike pre-treatment with inhibitor **I1**, none of the seven **P1** specific targets described in **Table 4.3** were outcompeted by pre-treatment with the non-specific probe **P8**, confirming the requirement of drug binding for labelling and thus their specificity. Target specificity for these proteins was also supported by the fact that they were not enriched by labelling with the non-directed alkyne-NHS probe **P7** (**Fig. 4.3**), with which only warhead-mediated labelling is possible. Since specific labelling of other proteins did occur, it is unlikely that probe **P1** quenching was the reason why EGFR was not observed.

While most unlikely, this hypothesis in the future could be further investigated by testing other 4-anilinoquinazoline derivative probes with more stable electrophilic warheads. Previous works with other lysine-directed irreversible inhibitors have shown the importance of fine-tuning the reactivity of the electrophilic warhead for obtaining a good target selectivity, especially through the kinetic control by electronic effects.[37-38] The kinetics

characteristics of several electrophilic warheads commonly used for the endogenous labelling of proteins with ligand-directed strategies has been recently reviewed.[5] In addition to the intrinsic differences between different reactive groups, the characteristics of the probe linker such as length and rigidity have also been reported to affect the labelling kinetics characteristics.[15] Therefore, new EGFR-directed probes with various warhead and linker characteristics could provide further information on the nature of newly identified targets of probe **P1**.

4.2.3.1.2 Structural bias for probe off-targets

As previously describe in section 4.1.3 for the Lapatinib derived probe **P6**, the modifications of a drug of interest with various linkers can shift the binding preference of a drug from the main protein of interest to an off-target.[14] Therefore, it is possible that the structural modifications in probe **P1** may bias its binding affinity for proteins other than EGFR. This hypothesis is supported by the identification of **P1** specific proteins in table **4.3** whose labelling is mediated by probe binding (confirmed by competition with inhibitor pre-incubation). Interestingly, although **P1** specific targets include a variety of proteins, the major target is in fact a kinase: ATP-dependent 6-phosphofructokinase (PFK, table **4.3**). To the best of our knowledge, non-protein kinases have not been reported as targets of probe **P5** by chemical proteomics studies. Further studies are required to validate PFK as an off-target of probe **P1**.

To validate PFK as a specific target of **P1**, the corresponding labelled sample (gel) could be immunoblotted against PFK. PFK has a molecular weight of 85 kDa, which is not observed in the corresponding fluorescent gel (**Fig. 4.5a**). In previous works where new off-targets of known drugs are identified, the potential new targets have been recombinantly expressed in mammalian cells and screened with various probe and inhibitors concentrations to validate the probe-target interaction, providing rough estimate of K_d or IC$_{50}$ values.[8] Assuming that probe

P1 binding to PFK inhibits its enzymatic activity, in vitro biochemical assays could also be used to validate the putative probe-target interaction. The determination of probe **P1** K_d value for EGFR (and if possible PKF) compared with known 4-anilinoquinazoline drugs and probes could also be helpful to explain a possible bias of this probe for PFK binding.

The determination of the probe-labelled residue in PFK could also be useful in validating PFK as an off-target of probe **P1**. The structural characteristics of this probe would limit the site of lysine labelling to the proximity of the ATP-binding pocket in PFK. These lysine residues could be mapped in the crystal structure of PFK to predict possible sites of modification. These predicted sites could then be compared to the actual labelling site in PFK, which can be determined by MS by eluting the probe-modified peptides from the neutravidin beads (this fraction was not analysed in our experiments). If probe **P1** does bind to PFK at the ATP-binding site, the modified (labelled) residue would be limited to those in close proximity to the ATP pocket.

If PFK is in fact validated as a direct probe **P1** off-target, an interesting question arises: why are none of the previously reported **P5** off-targets identified with **P1**? The differences in both the specific structure (linker) and reactivity of the warhead (NHS ester vs acrylamide) can have an influence in the subset of labelled proteins. Thus, the use of analogous probes in which the electrophilic warhead is replaced by a non-specific photocrosslinker (or various lysine and cysteine-targeted warheads) could provide further information and help dissect the influence of each component of the probe in the specific subset of labelled proteins.

To the best of our knowledge, PFK has not been previously reported in the literature as direct off-target of Afatinib. If PFK is in fact validated as direct **P1** off target, this could open a

novel research opportunity for repurposing not only Afatinib, but many other kinase inhibitors expanding their current utility.[39-40]

4.2.3.1.3 Capture of EGFR interactors

In virtually all covalent inhibitors the electrophilic warhead is directly attached to the pharmacophore to achieve selectivity. However in probe **P1** the activated ester is separated from the quinazoline core by a relatively long linker. In addition, the lysine of interest in EGFR is located in a more solvent exposed region compared with the cysteine targeted by probe **P5**. Therefore, it is possible that probe **P1** does bind to EGFR but the NHS ester then labels a protein in close proximity to EGFR rather than EGFR itself. In fact, several of the identified **P1** target proteins have been reported to interact with EGFR with important implications in drug resistance.

In 2018, Lee et al demonstrated that EGFR-mediated phosphorylation of phosphofructokinase (PFKP) promoted PI3K activation and enhanced glycolysis in glioblastoma cells.[41] In response to EGFR activation, lysine acetyl transferase 5 (KAT5) mediated the acetylation of PFKP at K395, promoting its translocation to the membrane and subsequent phosphorylation at Y64 by EGFR. Phosphorylated PFKP was then recruited by p85α and promoted PI3K activation leading to enhanced AKT-dependent PFK2 and PFK1 activation and glucose transporter type 1 (GLUT1) expression. This mechanisms was then linked to the observed Warburg effect, cell proliferation and tumorigenesis. Similarly, in A431 and MDA-MB-231 cells (TNBC, basal B type), EGFR activation was shown to promote PFKP stability mediated by Akt, resulting in enhanced glycolysis and cell proliferation.[42] In MDA-MB-468 cells, EGFR-PFKP interaction was predicted by Ingenuity Pathway Analysis, but not confirmed by coimmunoprecipitation experiments.[31]

Although evidence of direct EGFR interaction with phosphatidylinositol 5-phosphate 4-kinase (PIP4K2C) has not been reported in the literature, other phosphatidylinositol phosphate (PIP) kinases are known to regulate EGFR activity.[43-46] In 2007, Anderson et al demonstrated that the PIP kinase PIPKIγ661 is a direct substrate of EGFR involved in cell migration.[43] EGFR-mediated phosphorylation of PIPKIγ661at Y634 disrupts a PPI between PIPKIγ661and PLCγ, allowing PIPKIγ66 to associate with talin, thus regulating adhesion formation and facilitating EGF-induced migration. In another report, the same group showed that PIPKIγi5 controls EGFR lysosomal sorting.[45] Through its association with sortin nexin 5 (SN5) and Hrs, PIPKIγi5 prevents Hrs ubiquitination and facilitates the EGFR-Hrs interaction required for lysosomal sorting. Similarly, through its interaction with NEDD4-1, PIPKIγi5 prevents ubiquitination and degradation of the known EGFR tumor suppressor Mig6.[46]. To the best of our knowledge, this is the first experimental to suggest a direct interaction between EGFR and PIP4K2C. Together, these known interactions of EGFR support the hypothesis that probe **P1** might potentially target EGFR interactors such as PFKP and PIP4K2C.

Despite these reported interactions in the literature, more evidence is required to elucidate if the identified targets of probe **P1** are truly EGFR interactors or simply probe off-targets. The proposed dual cysteine-lysine targeting probe **P2** could help differentiate between these two possibilities by providing a physical link between the two labelled peptides. A similar approach was reported by Hamachi et al in 2012.[15] Using an FKBP-targeting ligand-directed tosyl chemistry probe, FKBP was covalently labeled with a diazirine-based photocrosslinker; UV irradiation of labelled FKBP in the presence of its known interactors mTOR or calcineurin allowed the crosslinking of FKBP-interactor complex *in vitro* but not under endogenous conditions, which was attributed to the low expression levels of the interacting proteins in that particular cell line.

Cross-linking mass spectrometry (XL-MS) is regarded as a powerful technique that provides unique topological information on protein complexes.[47] XL-MS make use of non-directed crosslinkers that are not biased to a particular protein of interest. Just as chemical proteomics increases the potential of global chemical proteomics studies by providing unique small-molecule/protein interaction information, XL-MS with probes targeting a specific protein of interest would enhance the potential of this technique. The newly designed probe **P1** might be the first step towards this goal that could eventually afford molecular tools that facilitate the characterization of endogenous EGFR complexes. Since the labelling characteristics of the probe are highly dependent on the nature of the linker and the reactivity of the warhead, next-generation chemical probes are required to test these hypotheses.

In addition to this chemical approach, other biochemical techniques used for the identification of PPI such as immunoprecipitation[48] and proximity ligation assays[49] could be used to confirm or reject the identified **P1** targets as EGFR interactors. It is important to notice that the current proteomic experiments with MDA-MB-468 cells were done in the absence of any externally added ligand (e.g. EGF). In the future, additional studies with EGF treatment can also facilitate the validation of probe **P1** targets as EGFR interactors, as normally PPI would be expected to increase with EGF treatment and be disrupted with an EGFR inhibitor. However, it is important to remember that several EGFR PPI have been reported to be independent of its kinase activity.[50-51] Therefore, further validation studies are required for the potential probe **P1** interactors.

4.2.4 Nuclear fractionation of P1- and P5-labelled MDA-MB-468 cells

In most chemical proteomics studies with covalent click probes, cell labelling is usually followed by the analysis of the whole cell lysate or partition methods to separate the membrane and soluble proteome. Beyond membrane fractionation, the isolation of other subcellular compartments in combination with drug-directed chemical probes has not been

explored extensively. Recently, some organelle-specific ABPP have also been developed for subcellular chemical proteomics analysis in specific cellular compartments.[35] As demonstrated by the vast amount of literature, EGFR function is highly dependent on its cellular localization and context; EGFR-directed subcellular proteomics analysis is therefore of great interest.[52-53] Taking advantage of high levels of endogenous EGFR present in MDA-MB-468 cells, cell labelling with probes **P1** and **P5** was combined with nuclear fractionation to characterize localization specific interactions.

Nuclear fractionation of labelled MDA-MB-468 cells was done following a previously reported method for this same cell line.[54] The effectiveness of the nuclear fractionation was confirmed by WB against lamin A as a nuclear marker (data not shown). Analysis of **P1** and **P5** treated nuclear (N) and cytosolic/membrane fractions (C) by SDS-PAGE fluorescent gel showed that both probes retained their labelling characteristics in both fractions, as they presented a very similar labelling profile (**Fig. 4.6**). Consistent with the results for whole cell lysate, EGFR labelling was observed only with probe P5 but not with probe **P1** (**Table 4.5**). **P5**-labelled EGFR was also observed in the nuclear fraction (**Table 4.4**).

Most of the targets of probes **P1** and **P5** identified in the whole cell lysate by proteomic analysis were also observed in both the nuclear and cytosolic/membrane fractions (**Table 4.4** and **Table 4.5**). Some of these proteins, such as PTGES2 and GHDC were selectively enriched only in the nuclear fraction, consistent with their reported cellular localization. This proof of concept demonstrates that it is possible to use currently available EGFR-directed probes in combination with subcellular fractionation beyond the plasma membrane to characterize localization specific functions.
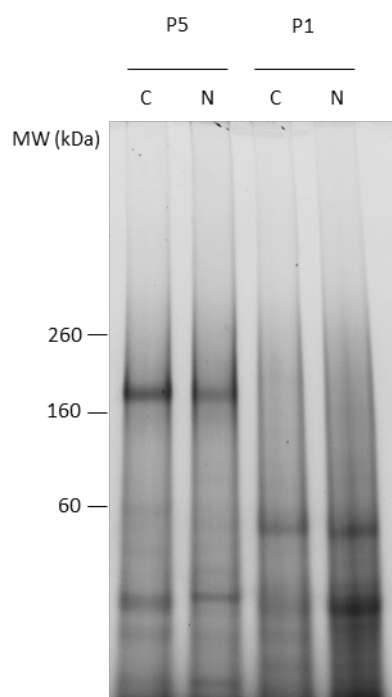
Fig. 4.6 Fluorescent gel for nuclear (N) and cytosolic/membrane (C) fractions of **P1**- and **P5**-labelled MDA-MB-468 cells In Situ

Table 4.4 Selected **P5** specific targets in nuclear and cytosolic/membrane fractions of MDA-MB-468 cells.

Entry	UniProt ID	Gene	Name	Cytosol/ membrane	Nucleus
1	P00533	EGFR	Epidermal growth factor receptor	Yes	Yes
2	Q9H7Z7	PTGES2	Prostaglandin E synthase 2	No	Yes
3	P04150	NR3C1	Glucocorticoid receptor	Yes	Yes
4	Q8N2G8	GHDC	GH3 domain-containing protein	No	Yes

Table 4.5 Selected **P1** specific targets in nuclear and cytosolic/membrane fractions of MDA-MB-468 cells.

Entry	UniProt ID	Gene	Name	Cytosol/ membrane	Nucleus
1	P17858	PFKL	ATP-dependent 6-phosphofructokinase, liver type	Yes	Yes
2	Q01813	PFKP	ATP-dependent 6-phosphofructokinase, platelet type	Yes	Yes
3	P37268	FDFT1	Squalene synthase	Yes	Yes
4	Q8TBX8	PIP4K2C	Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma	Yes	Yes

Under these experimental conditions, it is not possible to ascertain if probe **P5** is transported to the nucleus for labelling to occur, or cell-surface-labelled EGFR is transported into the nucleus, or a both mechanisms occurred simultaneously. However, to the best of our knowledge this is the first time that a small-molecule EGFR-directed probe is shown to label EGFR in the nuclear fraction. Recently, a nuclear-EGFR-directed inhibitor was obtained through the modification of Gefitinib with a nuclear targeting peptoid.[55] The inhibitor was shown to accumulate in the nucleus, allowing the selective inhibition of nuclear EGFR while sparing EGFR activity in the plasma membrane.[27, 55] Given the modular approach of probe synthesis, this selective nuclear EGFR inhibitor could be modified with our synthesized linkers to afford new chemical probes for future characterization of the endogenous nuclear EGFR interactome.

4.2.5 Conclusions

Here chemical proteomics analysis validated the use of the previously reported probe **P5** as a specific EGFR-directed probe. The specificity of the probe was confirmed by the small number of proteins that were competed by pre-incubation with the corresponding inhibitor **I5**, demonstrating the requirement of protein-ligand binding for labelling to occur. Most probe **P5** specific targets were common to MDA-MB468, HCT 116 and A431 cells. **P5** also retained its capacity to label EGFR at much lower expression levels such as in HCT116 cells. While EGFR labelling was not observed with the newly designed lysine-targeting probe **P1**, potential EGFR interactors may have been captured. This hypothesis is supported by the structural characteristics of the probe and the EGFR interactome previously reported in the literature. However, next-generation probes are required to confirm this hypothesis. Both probes **P1** and **P5** were capable of labelling nuclear fraction EGFR. This has important implications as EGFR nuclear localization has been association with poor prognosis and resistance to EGFR-targeted therapies. Current knowledge of nuclear EGFR activity has

already pointed to potential combinatorial therapies, and chemical probes capable of interrogating endogenous EGFR interactions in the nucleus or other compartments would further aid the design of efficacious therapies. Various inhibitors and probes capable to be directed to specific subcellular localization have recently been reported,[35, 55] as well as EGFR proteomics analysis in other subcellular compartments.[56] Therefore, probe **P5** represents the first step in the quest for novel tools to further elucidate EGFR signaling networks with important clinical implications.

4.3 EXPERIMENTAL

4.3.1 Biosafety compliance

Biosafety approval was obtained from Macquarie University Biosafety Committee (Project ID 0340 - 520180340968).

4.3.2 Materials and equipment

HCT 116 cells were provided by Dr. Seong Beom Ahn from the Department of Biomedical Sciences (Macquarie University, Australia). MDA-MB-468 cells were purchased from ATCC. DMEM media, FBS, 200 mM glutamine, BCA kit, phosphatase inhibitor mini tablets, neutravidin beads slurry, were purchased from Thermo Fisher Scientific (Waltham, MA, USA). PBS tablets, DMSO, MeOH, $CHCl_3$, tBuOH, TCEP, TBTA, $CuSO_4$, formic acid, NH_4OH, SDS, acetonitrile, azide-PEG3-Biotin, protease inhibitor cocktail (Roche), TruPAGE pre-cast gels, 20X TruPAGE gel running buffer and 4X TruPAGE loading buffer were obtained from Sigma-Aldrich (St. Louis, MO, USA). Rhodamine-azide was purchased from Click Chemistry Tools (Scottsdale, AZ, USA). Sequencing grade modified trypsin was obtained from Promega (Madison, WI, USA). Cell lysis was performed with a Branson 450 sonifier with probe. Fluorescent gels were visualized with a Typhoon Trio Scanner. Details for MS analysis are reported in the corresponding section.

4.3.3 Cell culture

HCT 116 and MDA-MB-468 cells were grown in DMEM media supplemented with 10% FBS and 200 mM glutamine (final concentration 5 mM) at 37 °C under a humidified 5% CO2 atmosphere. Cells seeding was performed at a density of 1.7×10^4 cells/cm^2; at this density cells reached confluency after six days. Cell passage was kept in low number (<10).

4.3.4 *In situ* labeling

Cells were grown to 90% confluence in 25 mL of complete growth medium in 15-cm plates. The growth medium was removed and the cells were washed with PBS (2 x 10 mL), followed by the addition of 10 mL growth medium without FBS (probe **P5**) or PBS (probe **P1**) containing 10 µL of 1,000X DMSO stock of the corresponding probe. The plate was swirled to disperse the compound and then incubated at 37 °C for 1 h. For competed experiments, cells were pre-incubated for 1 h with the corresponding control probe. Once the incubation was completed, the media was removed and the cells washed with ice-cold PBS (3 x 10 mL). The cells were then harvested in 10 mL ice-cold PBS and isolated by centrifugation at 500xg for 5 min. The supernatant was discarded and the cell pellets stored at −80 °C.

4.3.5 Cell lysis and protein quantification

Whole cell lysate: cell pellets were thawed on ice, and 1 mL of RIPA buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 1% NP40, 0.5% sodium deoxycholate, 0.1% SDS), containing protease and phosphatase inhibitors, was added to each pellet. The cells were lysed using a probe sonicator (i.e. burst 10 times, followed by 5 min rest on ice), and the lysis procedure was repeated two more times. The crude lysates were centrifuged at $500 \times g$, 4 °C for 10 minutes to remove cellular debris. The clarified lysates were transferred to new microcentrifuge tubes.

Nuclear fractionation: each thawed cell pellet was resuspended in 1 mL of cytoplasmic/membrane buffer (20 mM HEPES pH 7.5, 10 mM KCl, 2 mM MgCl$_2$, 0.5% NP40, protease and phosphatase inhibitor) and kept on ice for 15 min. The suspension was then treated with 40 strokes in a dounce homogenizer, and the lysate centrifuged at 1,500 × g, 4 °C for 5 minutes to sediment the nuclei. The supernatant containing the cytosolic/membrane fraction was transferred to a new microcentrifuge tube. The nuclear pellet was then washed five times with cytoplasmic/membrane buffer to ensure complete removal of cytoplasmic and membrane proteins. The pellet was then resuspended in 500 µL of nuclear buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 10 mM KCl, 2 mM MgCl$_2$, 0.5% NP40, protease and phosphatase inhibitor), kept on ice for 15 min and then sonicated for 10 seconds. The lysate was then cooled on ice for 5 minutes, vortexed for 30 seconds three times and then centrifuged at 15,000 × g, 4 °C for 10 minutes. The supernatant containing the nuclear fraction was transferred to a new microcentrifuge tube.

Protein concentration was determined using the BCA protein assay on a microplate reader. Cell lysates were then normalized to a final protein concentration of 2 mg/mL in PBS.

4.3.6 Sample processing for SDS-PAGE gel

Click chemistry was performed on 50 µL of 2 mg/mL normalized lysate with a final concentration of 25 µM rhodamine-azide (1 mM stock in DMSO), 1 mM TCEP (50 mM stock in water), 100 µM TBTA (1.7 mM in tBuOH/DMSO 4:1) and 1 mM CuSO$_4$ (50 mM stock in water) in a final volume of 57 µL. The samples were incubated at 25 °C for 1 h. Excess reagents were then removed by CHCl$_3$-MeOH precipitation with 400 µL ice-cold MeOH, 100 µL CHCl$_3$ and 400 µL water. After centrifugation at 13,000 × g, 4 °C for 15 min, the top and bottom layers were carefully removed. The protein pellet was resuspended in 400 µL ice-cold MeOH followed by centrifugation at 20,000 × g, 4 °C for 5 min. The supernatant

was discarded, and the protein pellet allowed to air dry for 15 min. The protein pellet was resuspended by sonication in 45 µL 1.2% SDS in PBS and then heated at 95 °C for 5 min. The sample was then centrifuge at 6,500 × g, 25 °C for 10 minutes to remove any insoluble material. The supernatant was transferred to a new microcentrifuge tube and treated with 15 µL 4X loading buffer. The samples (30 µL) were then loaded without boiling on a 10% SDS-PAGE gel, separated and imaged using a fluorescent scanner.

4.3.7 Sample processing for analysis by LC-MS/MS

Click chemistry was performed on 500 µL of 2 mg/mL normalized lysate with a final concentration of 100 µM biotin-azide (5 mM stock in DMSO), 1 mM TCEP (50 mM stock in water), 100 µM TBTA (1.7 mM in tBuOH/DMSO 4:1) and 1 mM CuSO$_4$ (50 mM stock in water) in a final volume of 570 µL in a 1.5 mL microcentrifuge tube. After 1 h incubation at 25 °C, each sample was transferred to a 15 mL centrifuge tube, and excess reagents were removed by CHCl$_3$-MeOH precipitation with 2 mL ice-cold MeOH, 0.5 mL CHCl$_3$ and 1.5 mL water. After centrifugation at 4,000 × g, 4 °C for 15 min, the top and bottom layers were carefully removed. The protein pellet was resuspended in 600 µL ice-cold MeOH, transferred to a 1.5 mL microcentrifuge tube and centrifuged at 20,000 × g, 4 °C for 5 minutes. The supernatant was discarded and the protein pellet allowed to air dry for 15 min. The protein pellet was resuspended by sonication in 500 µL 6M urea 25 mM NH$_4$HCO$_3$ in PBS. To this solution were added 5 µL 1M DTT in water and 140 µL 10% SDS in water, and the samples were incubated at 37 °C for 30 min. The samples were cooled briefly on ice, then treated with 40 µL 500 mM iodoacetamide and incubated at room temperature for 30 min in the dark. The samples were then diluted to a final concentration of 0.2% SDS with 5.315 mL PBS. The samples were then incubated with 50 µL dry neutravidin beads (from 100 µL slurry washed three times with PBS) for 2 h at room temperature. The beads were then transferred to a 1 mL spin column and washed three times with 700 µL 1% SDS in PBS, followed by five washes

with 700 μL PBS. The beads were then transferred to a 1.5 mL microcentrifuge tube and resuspended with 200 μL 2M urea in PBS and 2 μL 100 mM $CaCl_2$ (final concentration 1 mM). On-bead digestion was performed at 37 °C overnight with 2 μg trypsin. The next morning, the digest supernatant was collected by filtration from a spin column, and the beads washed two times with 100 μL PBS to afford a final digest volume of 400 μL. The digest was acidified with 4 μL formic acid (1% final concentration) and desalted by STAGE tip over two SDV-RPS membranes, washed with 120 μL 0.1% formic acid and eluted with 120 μL 80% acetonitrile 5% NH_4OH in water. The samples were speed vacuumed to dryness and then reconstituted in 10 μL 2% acetonitrile, 0.1% formic acid in water for LC/MS/MS analysis.

4.3.8 LC-MS/MS analysis

Mass spectrometric analysis of samples was performed on an Easy-nLC liquid chromatography system coupled to a Q-Exactive Quadrupole-Orbitrap (Thermo Scientific) mass spectrometer. Approximately 1μg of proteolytic peptide was dissolved in 10 μL of sample loading buffer (2% acetonitrile, 0.1% FA) and loaded onto a 3.5 cm column with a 100μm inner diameter, packed *in-house* with Solid core Halo® (2.7μm, 160Å, ES-C18) particles (Advanced Materials Technology). Peptide samples were desalted with 0.1% FA, then eluted into a 10cm column with a 75μm inner diameter, packed *in-house* with Solid core Halo® (2.7μm, 160Å, ES-C18) particles (Advanced Materials Technology). Bound peptides were separated with a binary buffer system of 0.1% (v/v) formic acid (buffer A) and 99.9% (v/v) acetonitrile/0.1% (v/v) formic acid (buffer B), at a flow rate of 300 nL/min. Peptides were eluted with a gradient of 2% – 30% buffer B over 120 minutes.

The column eluent was directed into the ionization source of the Q Exactive Quadrupole-Orbitrap (Thermo Scientific) mass spectrometer. A 2.4 kV electrospray voltage was applied via a liquid junction upstream of the column. Peptide precursors from 350 to 1850 m/z were scanned at 70k resolution with an AGC target value of $1x10^5$. The 10 most intense ions from

the preceding survey scan were fragmented by Higher-Energy Collisional Dissociation (HCD) using normalized collision energy of 30 with an isolation width of 3 m/z. Only precursors with charge state +2 to +5 were subjected to MS/MS analysis. The MS method had a minimum signal required value of 3×10^4 for MS2 triggering, an AGC target value of 1×10^5 for MS2 and a maximum injection time of 60 ms for MS2. MS/MS scan resolution was set at 17.5k. The dynamic exclusion was set to 20 seconds.

4.3.9 Proteomic data analysis

Raw mass spectrometric data were analyzed in the Proteome Discoverer (version 2.1, Thermo Scientific) and employed SequestHT and Mascot (Matrix Science, London, UK) search engine for database search. The MS/MS spectra were matched against the human Uniprot FASTA database (Download date; 18th October 2018 with 95,106 protein sequences). Enzyme specificity was set to trypsin, and the search included cysteine carbamidomethylation as a fixed modification and N-acetylation of protein, oxidation (M), deamidation (N, Q) and PyroGlu (Q), as variable modifications. Up to two missed cleavages were allowed for protease digestion. The search results were filtered for entries having False Discovery Rate (FDR) of less than 1% at protein, peptide and PSM level estimated through Target-decoy approach.

For the identification of target proteins, a label-free quantification method based on the NSAF was used. Proteins from each probe-labelled set were compared against two blanks: a DMSO-treated sample subjected to on-bead digestion and an unlabelled lysate sample digested in solution. Probe/blank ratios of NSAF, spectral counts, unique peptides and protein coverage were obtained for common proteins. Proteins with a spectral count ratio less than two, unique peptide and protein coverage less than one or containing only one unique peptide were assigned as background. For the remaining proteins, a spectral count and NSAF ratio more

than two between the probe and the corresponding inhibitor-competed treatment were considered specific.

4.4 REFERENCES

1. Kalesh, K. A.; Shi, H.; Ge, J.; Yao, S. Q., The use of click chemistry in the emerging field of catalomics. *Org. Biomol. Chem.* **2010**, *8*, 1749-1762.
2. Su, Y.; Ge, J.; Zhu, B.; Zheng, Y.-G.; Zhu, Q.; Yao, S. Q., Target identification of biologically active small molecules via in situ methods. *Curr. Opin. Chem. Biol.* **2013**, *17*, 768-775.
3. Wright, M. H.; Sieber, S. A., Chemical proteomics approaches for identifying the cellular targets of natural products. *Nat. Prod. Rep.* **2016**, *33*, 681-708.
4. Shi, H.; Zhang, C. J.; Chen, G. Y.; Yao, S. Q., Cell-based proteome profiling of potential dasatinib targets by use of affinity-based probes. *J. Am. Chem. Soc.* **2012**, *134* (6), 3001-3014.
5. Tamura, T.; Hamachi, I., Chemistry for Covalent Modification of Endogenous/Native Proteins: From Test Tubes to Complex Biological Systems. *J. Am. Chem. Soc.* **2019**, *141*, 2782-2799.
6. Weerapana, E.; Speers, A. E.; Cravatt, B. F., Tandem orthogonal proteolysis-activity-based protein profiling (TOP-ABPP)--a general method for mapping sites of probe modification in proteomes. *Nat. Protocols* **2007**, *2* (6), 1414-1425.
7. Speers, A. E.; Cravatt, B. F., Activity-Based Protein Profiling (ABPP) and Click Chemistry (CC)-ABPP by MudPIT Mass Spectrometry. *Curr. Protoc. Chem. Biol.* **2009**, *1*, 29-41.
8. Lanning, B. R.; Whitby, L. R.; Dix, M. M.; Douhan, J.; Gilbert, A. M.; Hett, E. C., . . . Cravatt, B. F., A road map to evaluate the proteome-wide selectivity of covalent kinase inhibitors. *Nat. Chem. Biol.* **2014**, *10* (9), 760-7.
9. Kurai, J.; Chikumi, H.; Hashimoto, K.; Yamaguchi, K.; Yamasaki, A.; Sako, T., . . . Shimizu, E., Antibody-Dependent Cellular CytotoxicityMediated by Cetuximab against Lung Cancer Cell Lines. *Clin. Cancer Res.* **2007**, *13* (5), 1552-1561.
10. Niessen, S.; Dix, M. M.; Barbas, S.; Potter, Z. E.; Lu, S.; Brodsky, O., . . . Cravatt, B. F., Proteome-wide Map of Targets of T790M-EGFR-Directed Covalent Inhibitors. *Cell Chem. Biol.* **2017**, *24*, 1388-1400.
11. Arrowsmith, C. H.; Audia, J. E.; Austin, C.; Baell, J.; Bennett, J.; Blagg, J., . . . Zuercher, W. J., The promise and peril of chemical probes. *Nat. Chem. Biol.* **2015**, *11* (8), 536-541.
12. Workman, P.; Collins, I., Probing the probes: fitness factors for small molecule tools. *Chem. Biol.* **2010**, *17* (6), 561-577.
13. Blagg, J.; Workman, P., Choose and Use Your Chemical Probe Wisely to Explore Cancer Biology. *Cancer Cell Perspect.* **2017**, *32* (1), 9-25.
14. Yamaura, K.; Kuwata, K.; Tamura, T.; Kioi, Y.; Takaoka, Y.; Kiyonaka, S.; Hamachi, I., Live cell off-target identification of lapatinib using ligand-directed tosyl chemistry. *Chem. Commun.* **2014**, *50* (91), 14097-14100.
15. Tamura, T.; Tsukiji, S.; Hamachi, I., Native FKBP12 engineering by ligand-directed tosyl chemistry: labeling properties and application to photo-cross-linking of protein complexes in vitro and in living cells. *J. Am. Chem. Soc.* **2012**, *134* (4), 2216-2226.
16. Garrido-Castro, A. C.; Lin, N. U.; Polyak, K., Insights into Molecular Classifi cations of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment. *Cancer Discov.* **2019**, *9*, 176-198.
17. Chavez, K. J.; Garimella, S. V.; Lipkowitz, S., Triple Negative Breast Cancer Cell Lines: One Tool in the Search for Better Treatment of Triple Negative Breast Cancer. *Breast Dis.* **2010**, *31*, 35-48.
18. Holliday, D. L.; Speirs, V., Choosing the right cell line for breast cancer research. *Breast Cancer Res.* **2011**, *13*, 215.
19. Cailleau, R.; Olive, M.; Cruciger, J., Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. *In Vitro* **1978**, *14* (11), 911-915.
20. Lehman, B. D.; Bauer, J. A.; Chen, X.; Sanders, M. E.; Chakravarthy, A. B.; Shyr, Y.; Petenpol, J. A., Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Inv.* **2011**, *121* (7), 2750-2767.
21. Dai, X.; Cheng, H.; Bai, Z.; Li, J., Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. *J. Cancer* **2017**, *8* (16), 3131-3141.

22. Filmus, J.; Pollak, M. N.; Cailleau, R.; Buick, R. N., MDA-468, a human breast cancer cell line with a high number of epidermal growth factor (EGF) receptors, has an amplified EGF receptor gene and is growth inhibited by EGF. *Biochem Biophys Res Commun.* **1985,** *128* (2), 898-905.

23. Armstrong, D. K.; Kaufmann, S. H.; Ottaviano, Y. L.; Furuya, Y.; Buckley, J. A.; Isaacs, J. T.; Davidson, N. E., Epidermal Growth Factor-mediated Apoptosis of MDA-MB-468 Human Breast Cancer Cells. *Cancer Res.* **1994,** *54,* 5280-5283.

24. Ali, R.; Wendt, M. K., The paradoxical functions of EGFR during breast cancer progression. *Signal Transduct. Target. Ther.* **2017,** *2.*

25. Hyatt, D. C.; Ceresa, B. P., Cellular localization of the activated EGFR determines its effect on cell growth in MDA-MB-468 cells. *Exp. Cell. Res.* **2008,** *314,* 3415-3425.

26. Rush, J. S.; Quinalty, L. M.; Engelman, L.; Sherry, D. M.; Ceresa, B. P., Endosomal Accumulation of the Activated Epidermal Growth Factor Receptor (EGFR) Induces Apoptosis. *J. Biol. Chem.* **2012,** *287* (1), 712-722.

27. Ali, R.; Brown, W.; Purdy, S. C.; Davisson, V. J.; Wendt, M. K., Biased signaling downstream of epidermal growth factor receptor regulates proliferative versus apoptotic response to ligand. *Cell Death Dis.* **2018,** *9,* 976.

28. Brand, T. M.; Iida, M.; Dunn, E. F.; Luthar, N.; Kostopoulos, K. T.; Corrigan, K. L., . . . Wheeler, D. L., Nuclear epidermal growth factor receptor is a functional molecular target in triple-negative breast cancer. *Mol. Cancer Ther.* **2014,** *13* (5), 1356-68.

29. Lanning, N. J.; Castle, J. P.; Singh, S. J.; Leon, A. N.; Tovar, E. A.; Sanghera, A., . . . Graveel, C. R., Metabolic profiling of triple-negative breast cancer cells reveals metabolic vulnerabilities. *Cancer & Metabolism* **2017,** *5:6.*

30. Kaplan, O.; Jaroszewski, J. W.; Faustino, P. J.; Zugmaier, G.; Ennis, B. W.; Lippman, M.; Cohen, J. S., Toxicity and Effects of Epidermal Growth Factor on Glucose Metabolism of MDA-468 Human Breast Cancer Cells. *J. Biol. Chem.* **1990,** *266* (3), 13461-13469.

31. Lim, S.-O.; Li, C.-W.; Xia, W.; Lee, H.-H.; Chang, S.-S.; Shen, J., . . . Hung, M.-C., EGFR Signaling Enhances Aerobic Glycolysis in Triple-Negative Breast Cancer Cells to Promote Tumor Growth and Immune Escape. *Cancer Res.* **2016,** *76* (5), 1284-1296.

32. Bantscheff, M.; Lemeer, S.; Savitski, M.; Kuster, B., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem* **2012,** *404,* 939-965.

33. Zhu, W.; Smith, J. W.; Huang, C.-M., Mass Spectrometry-Based Label-Free Quantitative Proteomics. *J. Biomed. Biotechnol.* **2010,** *2010,* 840518.

34. Shigeta, K.; Hayashida, T.; Hoshiniko, Y.; Okabayashi, K.; Endo, T.; Ishii, Y., . . . Kitagawa, Y., Expression of Epidermal Growth Factor Receptor Detected by Cetuximab Indicates Its Efficacy to Inhibit In Vitro and In Vivo Proliferation of Colorectal Cancer Cells. *PloS one* **2013,** *8* (6), e66302.

35. Zhu, H.; Tamura, T.; Hamachi, I., Chemical proteomics for subcellular proteome analysis. *Curr. Opin. Chem. Biol.* **2019,** *48,* 1-7.

36. Trinkle-Mulcahy, L.; Boulon, S.; Lam, Y. W.; Urcia, R.; Boisvert, F.-M.; Vandermoere , F., . . . Lamond, A., Identifying specifi c protein interaction partners using quantitative mass spectrometry and bead proteomes. *J. Cell Biol.* **2008,** *183,* 223-239.

37. Choi, S.; Connelly, S.; Reixach, N.; Wilson, I. A.; Kelly, J. W., Chemoselective small molecules that covalently modify one lysine in a non-enzyme protein in plasma. *Nat. Chem. Biol.* **2010,** *6* (2), 133-139.

38. Dalton, S. E.; Dittus, L.; Thomas, D. A.; Convery, M. A.; Nunes, J.; Bush, J. T., . . . Campos, S., Selectively Targeting the Kinome-Conserved Lysine of PI3Kδ as a General Approach to Covalent Kinase Inhibition. *J. Am. Chem. Soc.* **2018,** *140,* 932-939.

39. Yu, C.-H.; Chou, C.-C.; Tu, H.-F.; Huang, W.-C.; Ho, Y.-Y.; Khoo, K.-H., . . . Chang, G.-D., Antibody-assisted target identification reveals afatinib, an EGFR covalent inhibitor, down-regulating ribonucleotide reductase. *Oncotarget* **2018,** *9* (30), 21512-21529.

40. Foulkes, D. M.; Byrne, D. P.; Bailey, F. P.; Ferries, S.; Eyers, C. E.; Keeshan, K., . . . Eyers, P. A., Repurposing covalent EGFR/HER2 inhibitors for on-target degradation of human Tribbles 2 (TRIB2) pseudokinase. *bioRxiv* **2018.**

41. Lee, J.-H.; Liu, R.; Li, J.; Wang, Y.; Tan, L.; Li, X.-J., . . . Lu, Z., EGFR-Phosphorylated Platelet Isoform of Phosphofructokinase 1 Promotes PI3K Activation. *Mol. Cell* **2018,** *70,* 197-210.

42. Lee, J.-H.; Liu, R.; Zhang, C.; Wang, Y.; Cai, Q.; Qian, X., . . . Lu, Z., Stabilization of phosphofructokinase 1 platelet isoform by AKT promotes tumorigenesis. *Nat. Commun.* **2017,** *8,* 949.

43. Sun, Y.; Ling, K.; Wagoner, M. P.; Anderson, R. A., Type Iγ phosphatidylinositol phosphate kinase is required for EGF-stimulated directional cell migration. *J. Cell. Biol.* **2007,** *178* (2), 297-308.

44. Ramel, D.; Lagarrigue, F.; Pons, V.; Mounier, J.; Dupuis-Coronas, S.; Chicanne, G., . . . Payrastre, B., Shigella flexneri Infection Generates the Lipid PI5P to Alter Endocytosis and Prevent Termination of EGFR Signaling. *Sci. Signal.* **2011,** *4* (191), ra61.

45. Sun, Y.; Hedman, A. C.; Tan, X.; Schill, N. J.; Anderson, R. A., Endosomal type Igamma PIP 5-kinase controls EGF receptor lysosomal sorting. *Dev. Cell* **2013,** *25* (2), 144-55.

46. Sun, M.; Cai, J.; Anderson, R. A.; Sun, Y., Type I Phosphatidylinositol Phosphate 5-Kinase i5 Controls the Ubiquitination and Degradation of the Tumor Suppressor Mitogen-inducible Gene 6. *J. Biol. Chem.* **2016,** *291* (41), 21461-21473.

47. O'Reilly, F. J.; Rappsilber, J., Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat. Struct. Mol. Biol.* **2018,** *25*, 1000-1008.

48. Saafan, H.; Foerster, S.; Parra-Guillen, Z. P.; Hammer, E.; Michaelis, M.; Cinatl, J., Jr., . . . Ritter, C. A., Utilising the EGFR interactome to identify mechanisms of drug resistance in non-small cell lung cancer - Proof of concept towards a systems pharmacology approach. *Eur. J. Pharm. Sci.* **2016,** *94*, 20-32.

49. Smith, M. A.; Hall, R.; Fisher, K.; Haake, S. M.; Khalil, F.; Schabath, M. B., . . . Haura, E. B., Annotation of human cancers with EGFR signaling–associated protein complexes using proximity ligation assays. *Sci. Signal.* **2015,** *8* (359), ra4.

50. Weihua, Z.; Tsan, R.; Huang, W.-C.; Wu, Q.; Chiu, C.-H.; Fidler, I. J.; Hung, M.-C., Survival of Cancer Cells Is Maintained by EGFR Independent of Its Kinase Activity. *Cancer Cell* **2008,** *13*, 385-393.

51. Tan, X.; Thapa, N.; Sun, Y.; Anderson, R. A., A Kinase-Independent Role for EGF Receptor in Autophagy Initiation. *Cancer Cell* **2015,** *160*, 145-160.

52. Orre, L. M.; Vesterlund, M.; Pan, Y.; Arslan, T.; Zhu, Y.; Fernandez Woodbridge, A., . . . Lehtio, J., SubCellBarCode: Proteome-wide Mapping of Protein Localization and Relocalization. *Mol. Cell.* **2019,** *73*, 166-182.

53. Itzhak, D. N.; Tyanova, S.; Cox, J.; Borner, G. H. H., Global, quantitative and dynamic mapping of protein subcellular localization. *Elife* **2016,** *5*, e16950.

54. Chen, X.; Guo, D.; Zhu, Y.; Xian, F.; Liu, S.; Lin, W.; Lou, X., Nuclear phosphoproteomics analysis reveals that CDK1/2 are involved in EGF-regulated constitutive pre-mRNA splicing in MDA-MB-468 cells. *J. Proteom.* **2016,** *141*, 77-84.

55. Bartolowits, M. D.; Brown, W.; Ali, R.; Pedley, A. M.; Chen, Q.; Harvey, K. E., . . . Davisson, V. J., Selective Inhibition of STAT3 Phosphorylation Using a NuclearTargeted Kinase Inhibitor. *ACS Chem. Biol.* **2017,** *12*, 2371-2378.

56. Gosney, J. A.; Wilkey, D. W.; Merchant, M. L.; Ceresa, B. P., Proteomics reveals novel protein associations with early endosomes in an epidermal growth factor-dependent manner. *J. Biol. Chem.* **2018,** *293* (16), 5895-5908.

CHAPTER 5

Conclusions and Future Directions

Extensive studies on EGFR have shaped our current understating of RTK cell signalling. In the clinics, this has been reflected in a number of EGFR-targeted therapies currently validated for specific types of cancer. However, the success of these first therapies is challenged by drug resistance, which has hindered their validation in other cancer cases involving aberrant EGFR signaling. A multipronged approach to EGFR studies is necessary to overcome this problem. In this context, this project sought to contribute to these efforts by the synthesis of EGFR-directed chemical probes and their use in reverse and forward chemical proteomics analysis.

As detailed in the introduction, understanding of the chemical and physiological characteristics of EGFR has helped unravelling their implications in cancer and led to current EGFR-targeted therapies as well as identification of molecular mechanisms behind drug resistance, with particular emphasis in TNBC. An overview on existing chemical probes and their use in forward and reverse chemical proteomics demonstrated their unique value to studying biological systems. Our design and synthesis of EGFR-directed chemical probes for proteomics studies as detailed in this thesis adds to the list of useful probes that can discover potential interacting partners of EGFR.

Synthetic challenges, as described in Chapter 2 rising at different stages of the route, originated a number of linkers with varied complexity; these are general modules that can be used for the synthesis of future probes. After numerous attempts, a novel EGFR-directed lysine-targeting probe **P1** was obtained for testing in forward chemical proteomics (in Chapter 4**)**.

The potential of reverse chemical proteomics as a complementary approach to the commonly used forward chemical proteomics methods in identifying EGFR inhibitors off-targets agnostically was described in Chapter 3. Using a previously obtained non-covalent EGFR-

directed probe **P3**, DNA-topoisomerase I (TOPI) was identified as a potential target of Gefitinib in various commercially available T7 cDNA tumor libraries. Despite the crosstalk between EGFR and TOPI being extensively reported in the literature, experimental evidence of a direct physical interaction between Gefitinib and TOPI has not been previously reported. Our findings support the complementarity of reverse chemical proteomics to forward chemical proteomics in identifying small-molecule/protein interactions agnostically and find potential targets that may not be discovered otherwise. However, further studies are required to validate the direct physical interaction between TOP1 and probe **P3**. These experiments could include on-phage binding assay or other methods for the identification of direct targets of small molecules such as DARTS or proteome chips for example.

Due to the intrinsic differences between forward and reverse chemical proteomics, neither EGFR nor other off-targets previously reported in forward chemical proteomics for Gefitinib were identified for probe **P3**. These results suggest that displaying the cytosolic domain of EGFR, wether in part or in full length, on the T7 phage surface may be difficult due to the choice of the restriction sites. The use of other display cloning techniques, such as the yeast three-hybrid system, could produce new libraries in which EGFR is displayed for testing with EGFR-directed probes.

To complement the reverse chemical proteomics results, in Chapter 4 the newly designed EGFR-directed lysine-targeting probe **P1** was tested and compared with the previously reported cysteine-targeting probe **P5** in forward chemical proteomics analysis of the TNBC MDA-MB-468 cells. Despite the concentration and time-dependent labelling of proteins mediated exclusively by the cysteine warhead, **P5** showed good specificity for EGFR; only a couple of proteins were enriched by the probe. Remarkably, **P5** was also labeled EGFR at much lower protein expression levels, as demonstrated for HCT 116 cells.

Although EGFR was not labelled by probe **P1**, the structural characteristics of the probe suggest that potential EGFR interactors could have been captured instead for further validation. This hypothesis is supported by the fact that none of the identified probe **P1** interactors were common to those previously reported for probe **P5**. In addition to EGFR, common **P5** reported interactors include ErbB2, PGES2 and DUS2L. Several of the **P1**-targeted proteins were metabolic enzymes that have been reported previously to crosstalk with EGFR, and evidence of direct physical interaction has been previously reported for some of them, such as PFKP, supporting this hypothesis. Both probes retained their labelling capacity when combined with nuclear fractionation, and the identified targeted proteins corresponded to their reported subcellular localization. Probe **P5** was reported to label EGFR in the nuclear fraction for the first type. This observation has important implications, as EGFR-mediated phenotypic characteristics emerging from different subcellular compartments is unique; some of these localization-specific functions have been associated with resistance to current EGFR-targeted therapies. However, without further validations at this stage is not possible to differentiate the reported targets as EGFR interactors or probes off-targets. Validation experiments might include novel probes for the determination of direct physical interaction of protein-probe, determination of K_d of potential targets, EGFR activation with EGF as well as quantitative proteomics methods.

A probe capable of capturing endogenous EGFR PPI in specific subcellular localizations would be a powerful tool to complement current efforts to characterize EGFR functions beyond its canonical enzymatic activity. This could have important clinical implications and support the validation of EGFR-targeting in cancers that remain challenging to treat such as TNBC.

To further validate the hypothesis that **P1** labels EGFR interactors, next-generation probes are required. Fine tuning electrophilic warheads will likely improve labelling specificity,

potentially capturing residues located further away from the ligand binding site. The proposed structure for probe **P2** might help elucidate probe targets as EGFR interactors or drugs off-targets as discussed in chapter 4. Other probes could be designed that accumulate specifically in certain subcellular compartments. The modular approach to probe synthesis is expected to allow rapid access to these probes to continue our efforts of elucidating EGFR signalling networks, using the chemical proteomics analysis platform established in this work, in order to help develop more efficacious cancer therapies.

APPENDICES

A.1 NMR SPECTRA OF NEW COMPOUNDS

157

PMBO

1.7

¹H NMR: 7.284, 7.264, 6.904, 6.882, 4.491, 4.344, 4.338, 3.889, 3.824, 3.722, 3.706, 3.690, 3.600, 3.559, 2.510, 2.494, 2.477, 2.427, 2.421, 2.415, 1.465

Integrations: 2.38, 2.06, 1.82, 1.72, 0.90, 3.00, 1.84, 3.54, 1.95, 0.82, 9.51

7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 ppm

PMBO

1.7

¹³C NMR: 170.98, 159.31, 130.41, 129.40, 113.88, 80.65, 80.37, 76.72, 74.25, 73.19, 71.29, 71.00, 70.00, 67.26, 57.79, 55.39, 36.42, 28.22

170 160 150 140 130 120 110 100 90 80 70 60 50 40 ppm

158

L1A

L1A

L1B

L1B

1.8

1.8

P1 (¹H NMR)

9.522

8.944
8.705
8.274, 8.267, 8.257, 8.250
7.868, 7.828
7.488, 7.366, 7.346, 7.323

4.560, 4.551, 4.534, 4.525, 4.445, 4.429, 4.424, 4.418, 4.179, 4.174, 3.916, 3.901, 3.886, 3.794, 3.788, 3.782, 3.774

2.893

1.00 | 1.34 | 1.30 | 1.28 | 1.37 | 1.25 | 1.33 | 1.05 | 3.36 | 0.94 | 2.17 | 1.47 | 5.04

9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 ppm

P1 (¹³C NMR)

170.90, 170.52, 168.06
159.04, 158.31, 155.22, 154.69
140.48, 137.03
124.83, 123.18, 123.11, 121.87, 120.70, 120.52, 117.43, 117.21, 111.78, 109.19
81.02, 76.44, 75.88, 71.07, 70.64, 67.05
58.38
32.78, 26.29

170 160 150 140 130 120 110 100 90 80 70 60 50 40 ppm

163

PMBO⌒⌒OTBS
     |
     O
    4.5

7.271
7.260
7.250
6.881
6.859

4.515
4.486
4.477
4.466
4.447
3.802
3.674
3.661
3.586
3.576
3.561
3.551
3.492
3.479
3.453
3.426
3.413
3.403
3.400
3.389
3.376

0.878

0.049
0.044

3.30 2.00   1.93 3.02 1.98 0.94 3.88 1.13   9.79   6.03

7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 ppm

PMBO⌒⌒OTBS
     |
     O
    4.5

159.29

130.58
129.43

113.87

81.23
73.22
69.45
62.59
58.25
55.41

26.03
18.42

-5.24
-5.28

160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

164

PMBO￼OH
4.6
O—

7.262
7.260
7.241
6.889
6.867
4.508
4.479
4.474
4.445
3.803
3.772
3.720
3.663
3.622
3.584
3.572
3.559
3.552
3.547
3.540
3.527
3.515
3.453

2.70
2.00
2.03
3.00
1.23
1.00
1.97
4.01

7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 ppm

PMBO￼OH
4.6
O—

159.43
130.12
129.48
113.97
80.13
73.36
69.47
62.69
57.92
55.41

165 160 155 150 145 140 135 130 125 120 115 110 105 100 95 90 85 80 75 70 65 60 ppm

166

3.750
3.679
3.650
3.636
3.620
3.607
3.594
3.585
3.582
3.573
3.446
3.437
3.425
3.412
3.400

2.506
2.501
2.490
2.485
2.475
2.470

1.447

L4A

3.31
1.17
1.91
2.81
1.14

1.98

9.00

4.0   3.5   3.0   2.5   2.0   1.5   ppm

171.21

80.94
79.91
77.36
70.55
67.33
62.32
57.88

36.28
28.21

L4A

170 160 150 140 130 120 110 100 90   80   70   60   50   40   30   ppm

L4B

L4B

¹H NMR peaks: 9.176, 8.976, 8.129, 8.122, 8.112, 8.105, 7.948, 7.834, 7.829, 7.456, 7.434, 7.411, 4.567, 4.559, 4.541, 4.532, 4.429, 4.414, 4.403, 4.388, 3.826, 3.822, 3.788, 3.773, 3.758, 3.691, 3.679, 3.676, 3.468, 2.564, 2.548, 2.533

Integrations: 1.13, 1.14, 1.06, 0.92, 1.04, 1.04, 0.71, 0.59, 0.73, 1.86, 2.00, 3.00, 2.11

4.9

¹³C NMR peaks: 173.13, 160.62, 160.02, 158.09, 157.46, 155.63, 154.85, 145.65, 141.45, 134.71, 134.68, 126.98, 125.35, 125.28, 123.34, 121.24, 121.05, 117.90, 117.68, 107.68, 106.07, 78.80, 71.83, 69.90, 67.80, 58.39, 35.34

4.9

PMBO[structure 2.4]

PMBO[structure 2.4]

PMBO O OH
OH
2.5

7.260
7.240
6.887
6.865
4.429
3.824
3.799
3.696
3.686
3.667
3.658
3.620
3.607
3.592
3.576
3.569
3.560
3.553
3.537
3.522
3.507
3.497
3.482
1.894
1.878
1.863
1.847
1.832

2.34   2.00   2.04   1.01 3.08 1.09 3.02 4.06   2.05

PMBO O OH
OH
2.5

159.32
130.52
129.45
113.92
77.47
77.15
76.84
72.77
72.63
70.57
68.90
66.97
64.27
55.41
29.99

^{1}H NMR spectrum

PMBO — O — OTBS
          OH
2.6

7.263
7.260
7.242
6.885
6.864
4.429
3.801
3.637
3.625
3.612
3.578
3.562
3.546
3.542
3.527
3.511
3.467
3.454
3.452
3.437
2.497
2.485
1.905
1.889
1.873
1.857
1.842
0.897
0.067

2.53 2.00 2.05 4.03 1.79 4.37 1.96 0.98 2.02 9.19 6.11

7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 ppm

PMBO — O — OTBS
          OH
2.6

159.29
130.70
129.38
113.91
72.78
71.73
70.76
68.66
67.12
64.18
55.41
30.16
26.01
18.43
-5.28

160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

174

7.265
7.260
7.243
6.885
6.863

4.429
4.319
4.317
4.313
4.311
3.803
3.735
3.724
3.709
3.684
3.673
3.667
3.572
3.569
3.545
3.538
3.529
3.513
3.494
3.481
3.468
2.400
2.397
2.391
2.385
1.904
1.888
1.872
1.856
1.841
0.892
0.885
0.079
0.073
0.060
0.058

PMBO—O—OTBS

O

2.7

2.89 2.00 1.99 1.45 2.97 2.34 6.61 0.94 2.05 9.00 5.96

7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 ppm

159.27

130.80
129.37

113.91

80.53
78.46
74.09
72.78
70.91
68.66
67.22
63.19
57.86
55.42

30.28
30.23
26.03
25.99

18.41

-4.56
-4.58
-5.23
-5.29

PMBO—O—OTBS

O

2.7

160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

PMBO~~~O~~~OMs

L2B

PMBO~~~O~~~OMs

L2B

1H NMR peaks: 10.150, 9.223, 8.668, 8.163, 8.156, 8.146, 8.139, 7.814, 7.774, 7.480, 7.477, 7.458, 7.435, 7.170, 7.148, 6.846, 6.824, 4.437, 4.428, 4.411, 4.401, 4.357, 4.351, 4.337, 4.330, 4.324, 4.293, 4.024, 4.015, 4.011, 4.001, 3.697, 3.594, 3.581, 3.521, 3.505, 3.489, 3.433, 3.417, 3.412, 3.406, 3.401, 3.180, 1.785, 1.769, 1.753

Integrations: 0.96, 0.96, 1.00, 1.13, 1.14, 2.17, 2.16, 2.04, 0.99, 3.05, 1.98, 1.03, 3.07, 1.90, 2.15, 3.57, 2.84, 2.29

13C NMR peaks: 157.85, 157.41, 154.83, 153.68, 153.29, 152.40, 138.73, 130.89, 130.41, 128.92, 123.82, 122.62, 122.55, 121.99, 118.97, 118.79, 116.76, 116.54, 113.48, 110.32, 108.05, 80.33, 77.07, 74.93, 71.49, 69.15, 69.05, 67.74, 66.24, 56.95, 54.93, 39.49, 29.48

1.10

1.10

¹H NMR

Peak labels:
10.142

9.202

8.656, 8.161, 8.155, 8.144, 8.138, 7.810, 7.770, 7.478, 7.464, 7.455, 7.432

4.635, 4.622, 4.483, 4.469, 4.455, 4.374, 4.358, 4.343, 3.592, 3.576, 3.561, 3.434, 3.409, 3.398, 3.340, 3.325, 3.298, 2.500, 2.031, 2.016, 2.000

Integrations: 0.92, 1.00, 0.98, 0.98, 0.99, 2.01, 1.00, 1.01, 2.03, 3.10, 1.03, 5.28, 2.08

1.11

x-axis: 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 ppm

¹³C NMR

Peak labels:
157.83, 157.33, 154.81, 153.83, 153.25, 152.39

138.86, 135.93

123.81, 122.62, 122.55, 121.78, 118.97, 118.79, 116.76, 116.54, 110.18, 107.89

72.42, 70.47, 66.79, 66.66, 63.03

39.50

28.60

1.11

x-axis: 160 150 140 130 120 110 100 90 80 70 60 50 40 ppm

A.2 DNA SEQUENCES OF SINGLE PLAQUES

All the cDNA inserts of individual plaques were analysed by a BLASTn search (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with the following parameters: **Database** –Reference RNA sequences (refseq_rna); **Organism** –Homo sapiends (taxid:9606); **Program selection** – Highly similar sequences (megablast). DNA sequence corresponds to single phage particles shown in **Fig. 3.5** (Chapter 3, Pag. 103).

Lung tumour, plaque E3

```
AGGGGGGGGCAGGGGTGGGTCTTCGCCCAGAAGCTGCAGGAGCTGTCGTATTCCAGTCAGGTGTGATGCTCGGGG
ATCCGAATTCAAGCCGCGGGTGGAAGAAGATGGCGTCGGGTGGTGGTGGCTGTAGCGCTTCGGAGAGACTGCCTC
CGCCCTTCCCCGGCCTGGAGCCGGAGTCCGAGGGGGCGGCCGGGGGATCAGAACCCGAGGCTGGGGACAGCGACA
CCGAGGGGGAGGACATTTTCACCGGCGCCGCGGTGGTCAGTAAACATCAGTCTCCAAAGATAACTACATCCCTTC
TTCCCATCAACAATGGCTCCAAAGAAAATGGGATCCATGAAGAACAAGACCAAGAGCCACAGGATCTCTTTGCAG
ATGCCACAGTGGAGCTATCCTTGGACAGCACACAAAATAATCAGAAGAAGGTGCTAGCCAAAACACTCATTTCTC
TTCCTCCTCAGGAAGCCACAAATTCTTCGAAGCCCCAGCCAACCTATGAGGAGCTAGAGGAAGAAGAACAGGAGG
ATCAATTTGATTTGACAGTCGGTATAACTGATCCTGAGAAGATAGGGGATGGTATGAATGCATATGTAGCCTACA
AAGTTACAACACAGACAAGCTTACCATTGTTCAGAAGCAAACAGTTTGCAGTAAAAAGAAGATTTAGTGACTTTC
TGGGTCTTTATGAGAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGCCTTGGGG
GGGGGTAAACTGGCCCTAATTGCCCGGGGGC (forward primer)
```

```
LOCUS       NM_003099      8332 bp     mRNA       linear     PRI 25-JUN-2018
DEFINITION  Homo sapiens sorting nexin 1 (SNX1), transcript variant 1, mRNA.
ACCESSION   NM_003099      VERSION        NM_003099.4
```

```
CDS 131-1699
   1 aggtctccgc ccctcggatc ccacggggtc ccttgcggcc ctcccactcc tcgcaccgtt
  61 ggatcgcttt gctcacggcg ctatctctcg ataaagttgt tgttgcggct tccgccgcgg
 121 gtggaagaag atggcgtcgg gtggtggtgg ctgtagcgct tcggagagac tgcctccgcc
 181 cttccccggc ctggagccgg agtccgaggg ggcggccggg ggatcagaac ccgaggctgg
 241 ggacagcgac accgaggggg aggacatttt caccggcgcc gcggtggtca gtaaacatca
 301 gtctccaaag ataactacat cccttcttcc catcaacaat ggctccaaag aaaatgggat
 361 ccatgaagaa caagaccaag agccacagga tctctttgca gatgccacag tggagctatc
 421 cttggacagc acacaaaata atcagaagaa ggtgctagcc aaaacactca tttctcttcc
 481 tcctcaggaa gccacaaatt cttcgaagcc ccagccaacc tatgaggagc tagaggaaga
 541 agaacaggag gatcaatttg atttgacagt cggtataact gatcctgaga agatagggga
 601 tggtatgaat gcatatgtag cctacaaagt tacaacacag acaagcttac cattgttcag
 661 aagcaaacag tttgcagtaa aaagaagatt tagtgacttt ctgggtcttt atgagaagct
 721 ttccgagaag cactctcaga atggcttcat tgtccctccg cccccggaga gagcctcat
   //
1681 ggcaaaggcc atctcctaat ggaccaagga ccccagagcc cacctgtgtg acgctgcctt
   //
8281 aatcaaataa agctattttc attatgggaa aacaaaaaaa aaaaaaaaa aa
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, <u>underlined</u> = expressed by CDS)

Frame 1:
<u>NSSRGWKKMASGGGGCSASERLPPPFPGLEPESEGAAGGSEPEAGDSDTEGEDIFTGAAVVSKHQSPKITTSLLP</u>
<u>INNGSKENGIHEEQDQEPQDLFADATVELSLDSTQNNQKKVLAKTLISLPPQEATNSSKPQPTYEELEEEEQEDQ</u>
<u>FDLTVGITDPEKIGDGMNAYVAY</u>

Frame 2:
IQAAGGRRWRRVVVAVALRRDCLRPSPAWSRSPRGRPGDQNPRLGTATPRGRTFSPAPRWSVNISLQR

Frame 3:
FKPRVEEDGVGWWWL

<u>Lung tumour, plaque E4</u>

GCAAGTGGGGGCCCGGGTGGGTCTTCGCCCAGAAGCTGCAGGAGCTGTCGTATTCCAGTCAGGTGTGATGCTCGG

GGATCC**<u>GAATTCA</u>**AGCCGAAAAGAGGAAAAGGTTCGAGCCTCTGGGGATGCAAAAATAAAGAAGGAGAAGGAAAA

TGGCTTCTCTAGTCCACCACAAATTAAAGATGAACCTGAAGATGATGGCTATTTTGTTCCTCCTAAAGAGGATAT

AAAGCCATTAAAGAGACCTCGAGATGAGGATGATGCTGATTATAAACCTAAGAAAATTAAAACAGAAGATACCAA

GAAGGAGAAGAAAAGAAAACTAGAAGAAGAAGAGGATGGTAAATTGAAAAAACCCAAGAATAAAGATAAAGATAA

AAAAGTTCCTGAGCCAGATAACAAGAAAAAGAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCC

TCTAAACGGTTTTTGGAGGGGTAAAAAATATGAGCTATTTTGTTCCTCCTAAAGAGGATATAAAGCCATTAAAGA

GACCTCGAGATGAGGATGATGCTGATTATAAACCTAAGAAAATTAAAACAGAAGATACCAAGAAGGAGAAGAAAA

GAAAACTAGAAGAAGAAGAGGATGGTAAATTGAAAAAACCCAAGAATAAAGATAAAGATAAAAAAGTTCCTGAGC

CAGATAACAAGAAAAAGAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGTCTT

GGAGGGGTAGTTATTATTTTAGGAGCGGGGAAG (forward primer)

LOCUS   NM_003286   3750 bp   mRNA   linear   PRI 01-JUL-2018
DEFINITION **Homo sapiens DNA topoisomerase I (TOP1), mRNA.**
ACCESSION   NM_003286   VERSION   NM_003286.3

CDS 251-2548

```
  1 agcccaaatg cgaacttagg ctgttacaca actgctgggg tctgttctcg ccgcccgccc
 61 ggcagtcagg cagcgtcgcc gccgtggtag cagcctcagc cgtttctgga gtctcgggcc
121 cacagtcacc gccgcttacc tgcgcctcct cgagcctccg gagtccccgt ccgcccgcac
181 aggccggttc gccgtctgcg tctcccccac gccgcctcgc ctgccgccgc gctcgtccct
241 ccgggccgac atgagtgggg accacctcca caacgattcc cagatcgaag cggatttccg
301 attgaatgat tctcataaac acaaagataa acacaaagat cgagaacacc ggcacaaaga
361 acacaagaag gagaaggacc gggaaaagtc caagcatagc aacagtgaac ataaagattc
421 tgaaaagaaa cacaaagaga aggagaagac caaacacaaa gatggaagct cagaaaagca
481 taaagacaaa cataaagaca gagacaagga aaaacgaaaa gaggaaaagg ttcgagcctc
541 tggggatgca aaaataaaga aggagaagga aaatggcttc tctagtccac cacaaattaa
601 agatgaacct gaagatgatg ctatttttgt tcctcctaaa gaggatataa agccattaaa
661 gagacctcga gatgaggatg atgctgatta taaacctaag aaaattaaaa cagaagatac
721 caagaaggag aagaaaagaa aactagaaga agaagaggat ggtaaattga aaaaacccaa
781 gaataaagat aaagataaaa aagttcctga gccagataac aagaaaaaga agccgaagaa
   //
2521 catggctgat gaagactatg agtttttag cc agtctcaaga ggcagagttc tgtgaagagg
```

186

```
  //
3721 cctacaactt aatggaaaaa aaaaaaaaa
```

Bold = Gene coding sequence
<u>Underlined</u> = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, <u>underlined</u> = expressed by CDS)

Frame 1:

<u>NSS</u>RKEEKVRASGDAKIKKEKENGFSSPPQIKDEPEDDGYFVPPKEDIKPLKRPRDEDDADYKPKKIKTEDTKK
EKKRKLEEEEDGKLKKPKNKDKDKKVPEPDNKKKKLAAALE

Frame 2:
IQAEKRKRFEPLGMQK

Frame 3:
FKPKRGKGSSLWGCKNKEGEGKWLL

Lung tumour, plaque E5

AGGTGGGGCAGGGTGGTCTTCGCCCAGAAGCTGCAGGAGCTGTCGTATTCCAGTCAGGTGTGATGCTCGGGGATC
C**<u>GAATTC</u>**TTCAAATAAGCCAGCTGTCACCACCAAGTCACCTGCAGTGAAGCCAGCTGCAGCCCCCAAGCAACCTG
TGGGCGGTGGCCAGAAGCTTCTGACGAGAAAGGCTGACAGCAGCTCCAGTGAGGAAGAGAGCAGCTCCAGTGAGG
AGGAGAAGACAAAGAAGATGGTGGCCACCACTAAGCCCAAGGCGACTGCCAAAGCAGCTCTATCTCTGCCTGCCA
AGCAGGCTCCTCAGGGTAGTAGGGACAGCAGCTCTGATTCAGACAGCTCCAGCAGTGAGGAGGAGGAAGAGAAGA
CATCTAAGTCTGCAGTTAAGAAGAAGCCACAGAAGGTAGCAGGAGGTGCAGCCCCTTCCAAGCCAGCCTCTGCAA
AGAAAGGAAAGGCTGAGAGCAGCAACAGTTCTTCTTCTGATGACTCCAGTGAGGAAGAGGAAGAGAAGCTCAAGG
GCAAGGGCTCTCCAAGACCACAAGCCCCCAAGGCCAATGGCACCTCTGCACTGACTGCCCAGAATGGAAAAGCAG
CTAAGAACAGTGAGGAGGAGGAAGAAGAAAAGAAAAAGGCGGCAGTGGTAGTTTCCAAATCAGGTTCATTAAAGA
AGCGGAAGCAGAATGAGGCTGCCAAGGAGGCAGAGACTCCTCAGGCCAAGAAGATAAAGCTTGCGGCCGCACTCG
AGTAACTAGTTAACCCCTTGGGGCCTCTAACGGCTTTGGGGGGGGGGTAAAGGCCCCGGGGGGGGG (forward
primer)

```
LOCUS    NM_001284388    3977 bp    mRNA    linear   PRI 30-JUN-2018
DEFINITION  Homo sapiens nucleolar and coiled-body phosphoprotein 1 (NOLC1),
            transcript variant 1, mRNA.
ACCESSION  NM_001284388 XM_005270274     VERSION     NM_001284388.1

CDS 236-2365
   1 gcggccggtg ggctccgccc ttaaccaaga tggcgatacg cgtgggaccg gaaagagttt
  61 atagatttcc cgtctaccct acctctgagg tgaaggtggg actgccctgt ggagcccacc
 121 ctttccgtta tgcgcccgcg cggcgcaatg acgtaacaca ggcccgccca ctgcccctgt
 181 tgggttcctg agtcgtgctg cgtcgacaac ggtagtgacg cgtattgcct ggagg atggc
 241 ggacgccggc attcgccgcg tggttcccag cgacctgtat cccctcgtgc tcggcttcct
 //
1381 ggatgatgaa gctccttcta agccagctgg taccaccaag aattcttcaa ataagccagc
1441 tgtcaccacc aagtcacctg cagtgaagcc agctgcagcc cccaagcaac ctgtgggcgg
1501 tggccagaag cttctgacga gaaaggctga cagcagctcc agtgaggaag agagcagctc
1561 cagtgaggag gagaagacaa agaagatggt ggccaccact aagcccaagg cgactgccaa
1621 agcagctcta tctctgcctg ccaagcaggc tcctcagggt agtagggaca gcagctctga
1681 ttcagacagc tccagcagtg aggaggagga agagaagaca tctaagtctg cagttaagaa
```

```
1741 gaagccacag aaggtagcag gaggtgcagc cccttccaag ccagcctctg caaagaaagg
1801 aaaggctgag agcagcaaca gttcttcttc tgatgactcc agtgaggaag aggaagagaa
1861 gctcaagggc aagggctctc caagaccaca agcccccaag gccaatggca cctctgcact
1921 gactgcccag aatggaaaag cagctaagaa cagtgaggag gaggaagaag aaaagaaaaa
1981 ggcggcagtg gtagtttcca atcaggttc attaaagaag cggaagcaga atgaggctgc
2041 caaggaggca gagactcctc aggccaagaa gataaagctt cagacccta acacatttcc
2101 aaaaaggaag aaaggagaaa aaagggcatc atccccattc cgaagggtca gggaggagga
2161 aattgaggtg gattcacgag ttgcggacaa ctcctttgat gccaagcgag gtgcagccgg
2221 agactgggga gagcgagcca atcaggtttt gaagttcacc aaaggcaagt cctttcggca
2281 tgagaaaacc aagaagaagc ggggcagcta ccggggaggc tcaatctctg tccaggtcaa
2341 ttctattaag tttgacagcg agtgacctga ggccatcttc ggtgaagcaa gggtgatgat
  //
3961 aatgtgtaaa aaaaaaa
```

Bold = Gene coding sequence
<u>Underlined</u> = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, <u>underlined</u> = expressed by CDS)

Frame 1:
<u>NSSNKPAVTTKSPAVKPAAAPKQPVGGGQKLLTRKADSSSSEEESSSSEEEKTKKMVATTKPKATAKAALSLPAK</u>
<u>QAPQGSRDSSSDSDSSSSEEEEEKTSKSAVKKKPQKVAGGAAPSKPASAKKGKAESSNSSSSDDSSEEEEEKLKG</u>
<u>KGSPRPQAPKANGTSALTAQNGKAAKNSEEEEEEKKKAAVVVSKSGSLKKRKQNEAAKEAETPQAKKIKLAAALE</u>

Frame 2:
ILQISQLSPPSHLQ

Frame 3:
FFK

Lung tumour, plaque E7

ACAGTGGGGGCCGGTGGTCTTCGCCCAGAAGCTGCAGGAGCTGTCGTATTCCAGTCAGGTGTGATGCTCGGGGAT
CC**<u>GAATTC</u>**AATTCCGTCTTCCTGATGGTTCTTCCTTTACAAATCAGTTCCCTTCTGATGCTCCTCTAGAAGAGGC
AAGGCAGTTTGCTGCACAGACTGTTGGCAACACTTACGGTAATTTTTCGTTAGCAACCATGTTTCCCAGGAGGGA
ATTTACCAAAGAAGATTATAAAAAGAAGTTACTGGATTTGGAACTTGCCCCAAGCGCTTCGGTGGTACTGTTGCC
AGCAGGAAGACCAACTGCATCCATTGTACACTCTTCCAGCGGAGACATTTGGACCTTGTTGGGAACAGTGCTTTA
TCCATTCCTTGCCATCTGGAGATTAATTAGCAATTTCTTGTTTAGTAATCCGCCTCCCACACAGACTTCAGTGAG
AGTAACATCGTCAGAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGGCTTTGGA
AGGGGTAAAAAGGAGGTTAAGAAAAGGGGG (forward primer)

LOCUS   NM_014607   4018 bp   mRNA   linear   PRI 23-JUN-2018
DEFINITION **Homo sapiens UBX domain protein 4 (UBXN4), mRNA.**
ACCESSION NM_014607   VERSION NM_014607.3

CDS 312-1838
```
   1 ttcctccact tcctggactg cgcgaccgga ggcctagcgg gcgcgcgccc gcaccatcga
  61 ctcgccaacg agagaaggtc ctggggcacg gacaccgacg ggttgcgact gtgacgtgag
 121 gtgttctcgc gcgcgctagc gcgcgtctcc gggtgccgct gacgggcgtg cgcgcttgtg
 181 cggagccgga ggtggggggcc gaaccagcca aggttgcggg ggccgcagag ccggacgaag
 241 acgagggcg gagccggctt cgggactgcg gagactacac accgagcgag cgcctggggcc
```

188

```
 301 cgaagggagc gatgctgtgg ttccagggcg ccattccggc cgccatcgcg acggccaaaa
  //
1201 ccttgctagc aaaacaggca gaaatggaag tcaagaggga atcttatgca agagaaagaa
1261 gcactgttgc aagaattcaa ttccgtcttc ctgatggttc ttcctttaca aatcagttcc
1321 cttctgatgc tcctctagaa gaggcaaggc agtttgctgc acagactgtt ggcaacactt
1381 acggtaattt ttcgttagca accatgtttc ccaggaggga atttaccaaa gaagattata
1441 aaaagaagtt actggatttg gaacttgccc caagcgcttc ggtggtactg ttgccagcag
1501 gaagaccaac tgcatccatt gtacactctt ccagcggaga catttggacc ttgttgggaa
1561 cagtgcttta tccattcctt gccatctgga gattaattag caatttcttg tttagtaatc
1621 cgcctcccac acagacttca gtgagagtaa catcgtcaga accccaaac cctgcatcat
1681 ctagcaaatc agaaaaaagg gaaccagtga gaaaaagagt gctggaaaaa cgtggagacg
1741 actttaaaaa ggaggggaaa atttatagat taaggactca agatgatggt gaagatgaaa
1801 acaacacttg gaatggaaat tccactcaac agatgtagtg tgacaagtat aatatgtgca
  //
3961 aagcaaatat gtaatcttta ttttttaaat aaatgggatc atattatata ttctaaaa
```

Bold = Gene coding sequence
<u>Underlined</u> = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, <u>underlined</u> =
expressed by CDS)

Frame 1:
NSIPSS

Frame 2:
<u>**IQFRLPDGSSFTNQFPSDAPLEEARQFAAQTVGNTYGNFSLATMFPRREFTKEDYKKKLLDLELAPSASVVLLPA
GRPTASIVHSSSGDIWTLLGTVLYPFLAIWRLISNFLFSNPPPTQTSVRVTSSE**</u>ACGRTRVTS

Frame 3:
FNSVFLMVLPLQISSLLMLL

Breast tumour, plaque A1

GAAAGGGGGGCGGGGTGGGTCTTCGCCCAGAAGCTGCAGGAGCTGTCGTATTCCAGTCAGGTGTGATGCTCGGGG
ATCC**<u>GAATTCA</u>**AGCCGAGAACACCGGCACAAAGAACACAAGAAGGAGAAGGACCGGGAAAAGTCCAAGCATAGCA
ACAGTGAACATAAAGATTCTGAAAAGAAACACAAAGAGAAGGAGAAGACCAAACACAAAGATGGAAGCTCAGAAA
AGCATAAAGACAAACATAAAGACAGAGACAAGGAAAAACGAAAAGAGGAAAAGGTTCGAGCCTCTGGGGATGCAA
AAATAAAGAAGGAGAAGGAAATGGCTTCTCTAGTCCACCACAAATTAAAGATGAACCTGAAGATGATGGCTATT
TTGTTCCTCCTAAAGAGGATATAAAGCCATTAAAGAGACCTCGAGATGAGGATGATGCTGATTATAAACCTAAGA
AAATTAAAACAGAAGATACCAAGAAGGAGAAGAAAGAAACTAGAAGAAGAAGAGAAATGAATGGAATAAATGC
AGAGGATCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCGG
GGGGCCCCCCCCTAAAAAAAATTTTTTCCCCCGGGGGGGGCCCCAAGGTATGGGGGGGAGGGGGGGGGGGGGGAACG
AACACCCCCAAAAAAAGCAAAAAAAGCCGGGCAAAAGGGGAA (forward primer).


```
LOCUS NM_003286      3750 bp    mRNA    linear    PRI 01-JUL-2018
DEFINITION  Homo sapiens DNA topoisomerase I (TOP1), mRNA.
ACCESSION NM_003286    VERSION NM_003286.3

CDS 251-2548
   1 agcccaaatg cgaacttagg ctgttacaca actgctgggg tctgttctcg ccgcccgccc
```

```
  61 ggcagtcagg cagcgtcgcc gccgtggtag cagcctcagc cgtttctgga gtctcgggcc
 121 cacagtcacc gccgcttacc tgcgcctcct cgagcctccg gagtccccgt ccgcccgcac
 181 aggccggttc gccgtctgcg tctcccccac gccgcctcgc ctgccgccgc gctcgtccct
 241 ccgggccgac atgagtgggg accacctcca caacgattcc cagatcgaag cggatttccg
 301 attgaatgat tctcataaac acaaagataa acacaaagat cgagaacacc ggcacaaaga
 361 acacaagaag gagaaggacc gggaaaagtc caagcatagc aacagtgaac ataaagattc
 421 tgaaaagaaa cacaaagaga aggagaagac caaacacaaa gatggaagct cagaaaagca
 481 taaagacaaa cataaagaca gagacaagga aaaacgaaaa gaggaaaagg ttcgagcctc
 541 tggggatgca aaaataaaga aggagaagga aaatggcttc tctagtccac cacaaattaa
 601 agatgaacct gaagatgatg ctattttgt cctcctaaa gaggatataa agccattaaa
 661 gagacctcga gatgaggatg atgctgatta taaacctaag aaaattaaaa cagaagatac
 721 caagaaggag aagaaaagaa aactagaaga agaagaggat ggtaaattga aaaaacccaa
 781 gaataaagat aaagataaaa aagttcctga gccagataac aagaaaaaga agccgaagaa
  //
2461 tgtcccaatt gagaagattt acaacaaaac ccagcgggag aagtttgcct gggccattga
2521 catggctgat gaagactatg agtttttagcc agtctcaaga ggcagagttc tgtgaagagg
  //
3661 tatcttctat ttttatcatg aattcccttt taatcaactg taggttattt aaaataaatt
3721 cctacaactt aatggaaaaa aaaaaaaaaa //
```

Bold = Gene coding sequence
Underlined = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, underlined = expressed by CDS)

Frame 1:

**NSSREHRHKEHKKEKDREKSKHSNSEHKDSEKKHKEKEKTKHKDGSSEKHKDKHKDRDKEKRKEEKVRASGDAKI
KKEKENGFSSPPQIKDEPEDDGYFVPPKEDIKPLKRPRDEDDADYKPKKIKTEDTKKEKKRKLEEEEK**

Frame 2:
IQAENTGTKNTRRRRTGKSPSIATVNIKILKRNTKRRRRPNTKMEAQKSIKTNIKTETRKNEKRKRFEPLGMQK

Frame 3:
FKPRTPAQRTQEGEGPGKVQA

Breast tumour, plaque A5

AGGGGGGGGGCGGGGTGGGTCTTCGCCCAGAAGCTGCAGGAGCTGTCGTATTCCAGTCAGGTGTGATGCTCGGGGA
TCC**GAATTCA**AGCGAGGCGGCAGGCCCGGGTCAGGGGCCTCGAGATCGGGCTTGGGCCCAGAGCATGTTCCAGAT
CCCAGAGTTTGAGCCGAGTGAGCAGGAAGACTCCAGCTCTGCAGAGAGGGGCCTGGGCCCCAGCCCCGCAGGGGA
CGGGCCCTCAGGCTCCGGCAAGCATCATCGCCAGGCCCCAGGCCTCCTGTGGGACGCCAGTCACCAGCAGGAGCA
GCCAACCAGCAGCAGCCATCATGGAGGCGCTGGGGCTGTGGAGATCCGGAGTCGCCACAGCTCCTACCCCGCGGG
GACGGAGGACGACGAAGGGATGGGGGAGGAGCCCAGCCCCTTTCGGGGCCGCTCGCGCTCGGCGCCCCCCAACCT
CTGGGCAGCACAGCGCTATGGCCGCGAGCTCCGGAGGATGAGTGACGAGTTTGTGGACTCCTTTAAGAAGGGACT
TCCTCGCCCGAAGAGCGCGGGCACAGCAACGCAGATGCGGCAAAGCTCCAGCTGGACGCGAGTCTTCCAGTCCTG
GTGGGATCGGAACTTGGGCAGGGGAAGCTCCGCCCCCTCCCAGTGACCTTCGCTCCACATCCCGAAAGCTTGCGG
CCGCACTCGAGTAACTAGTTAACCCCTTGGGGCCTCTAAACGGTCTTTGGAGGGGGTAAGGGGGCCGCGGGCCGG
G

LOCUS   NM_032989   986 bp   mRNA   linear   PRI 05-AUG-2018
DEFINITION **Homo sapiens BCL2 associated agonist of cell death (BAD),
            transcript variant 2, mRNA.**

190

ACCESSION NM_032989   VERSION NM_032989.2

CDS 83-589
```
  1 aactagggcc cggagcccgg ggtgctggag ggaggcggca ggcccgggtc aggggcctcg
 61 agatcgggct tgggcccaga gcatgttcca gatcccagag tttgagccga gtgagcagga
121 agactccagc tctgcagaga ggggcctggg ccccagcccc gcaggggacg ggccctcagg
181 ctccggcaag catcatcgcc aggccccagg cctcctgtgg gacgccagtc accagcagga
241 gcagccaacc agcagcagcc atcatggagg cgctggggct gtggagatcc ggagtcgcca
301 cagctcctac cccgcggggga cggaggacga cgaagggatg ggggaggagc ccagcccctt
361 tcggggccgc tcgcgctcgg cgccccccaa cctctgggca gcacagcgct atggccgcga
421 gctccggagg atgagtgacg agtttgtgga ctcctttaag aagggacttc ctcgcccgaa
481 gagcgcgggc acagcaacgc agatgcggca aagctccagc tggacgcgag tcttccagtc
541 ctggtgggat cggaacttgg cagggggaag ctccgccccc tcccagtgac cttcgctcca
601 catcccgaaa ctccacccgt tcccactgcc ctgggcagcc atcttgaata tgggcggaag
661 tacttccctc aggcctatgc aaaaagagga tccgtgctgt ctcctttgga gggagggctg
721 acccagattc ccttccggtg cgtgtgaagc cacggaaggc ttggtcccat cggaagtttt
781 gggttttccg cccacagccg ccggaagtgg ctccgtggcc ccgccctcag gctccgggct
841 ttcccccagg cgcctgcgct aagtcgcgag ccaggtttaa ccgttgcgtc accgggaccc
901 gagcccccgc gatgccctgg gggccgtgct cactaccaaa tgttaataaa gcccgcgtct
961 gtgccgccga aaaaaaaaaa aaaaaa
```

Bold = Gene coding sequence
<u>Underlined</u> = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, <u>underlined</u> = expressed by CDS)

Frame 1:
NSSEAAGPGQGPRDRAWAQS<u>MFQIPEFEPSEQEDSSSAERGLGPSPAGDGPSGSGKHHRQAPGLLWDASHQQEQP</u>
<u>TSSSHHGGAGAVEIRSRHSSYPAGTEDDEGMGEEPSPFRGRSRSAPPNLWAAQRYGRELRRMSDEFVDSFKKGLP</u>
<u>RPKSAGTATQMRQSSSWTRVFQSWWDRNLGRGSSAPSQ</u>

Frame 2:
IQARRQARVRGLEIGLGPRACSRSQSLSRVSRKTPALQRGAWAPAPQGTGPQAPASIIARPQASCGTPVTSRSSQ
PAAAIMEALGLWRSGVATAPTPRGRRTTKGWGRSPAPFGAARARRPPTSGQHSAMAASSGG

Frame 3:
FKRGGRPGSGASRSGLGPEHVPDPRV

<u>Breast tumour, plaque A9</u>

CAGGGGGCAGGTGGTCTTCGCCCAGAAGCTGCAGGAGCTGTCGTATTCCAGTCAGGTGTGATGCTCGGGGATCC**G**
AATTCAAGCCAAATATTCAGCAGAGGAAGGTGCAGAGGAAGAAGAAGAGAAGATGCAGAATGGGAAAGACAGAGG
TTAGCAGAGGAGCCGGACACAGAGGAAGCTCAGGCACAGAGGACGAGGAGCAAGCTGGCGCCGACATGGCGAAGG
CAAGGTCTTCCCCCAGAGGCACATTCCTCTCCATCTTTCCACCGCACACCTGGACCAGGCTTGCAGGCTGCCAGA
CGTCACTCCACCCGCCAGGGAGAGGGGAGCCAGAGCCGGTGGGAAGCGGGGAGGGGCTGCGTGGCACAGCTAGTG
GGCCTCCCCCTGCACAGCCCTGCATGTACTAGCACCTTCATCACTCCCCTCAGGGCATGGTCTCATCTCCGCATC
AGGAATTCACCTGGAGGTTGAAAAGAGAAAAGAAAGCTTGCGGCCGCACTCGAGTAACTAGTTAACCCCTTGGG
GCCTCTAAACGGGTTTGGGAGGGGGTAAGGGGGAGTCCTCTTGCGGGGGCGGCAGGAGTGCGGTTCTGGGTCCAA
CGGCTGGGGGAAATCGGGATGCGGAAGATTAGAACAGGCCCTGAGGGGGAGTGGAGGAACGAGGCAGGTAAGAGT
GGGGCGGTGCAGGGGGGAGGGGCAGTGACGGGGCCAGGCACCGGTCGGAGCGGCGGGGGGGGGGGGGGGGGGGGGG

GGGGGGGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGAGGGGGGGGGGG
GGGGGGGGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
(forward primer)

LOCUS   NM_001351381   2311 bp   mRNA   linear   PRI 01-JUL-2018
DEFINITION **Homo sapiens dystrobrevin beta (DTNB), transcript variant 15,
            mRNA.**
ACCESSION   NM_001351381 XM_017003540   VERSION   NM_001351381.1

CDS 619-1971
```
   1 gcgcgctcgc gctgggcgcg cgcccgccca tccgccctcc gccctccgcc ctcgcgccct
  61 cgcgccgggg gcgggcgctg cgcgtgccgg gcggagaagc ggcgcgagcc gggcgctgcg
 121 aacgttcgcc gcgggggtgg ctccggggcc tgagtaggcg ctgccgctgc ctcagccgag
 181 ggggctgggc cggagcgtgc ggaggagtga ggccgcagga gaccttcccg acgacccctg
 241 ctccggcggg gaagtgagca aggatgattg aggaaagtgg gaacaagcgg aagaccatgg
 301 cagagaagag gcagctgttc atagaaatgc gtgctcagaa ttttgatgtc atacgactat
 361 caacttacag aacagcctgc aaattacgat ttgtacaaaa acgatgcaac cttcatcttg
 421 ttgatatctg gaacatgatt gaagccttcc gagacaatgg ccttaataca ctggaccata
 481 ccaccgagat cagtgtgtcc cgcctcgaaa ctgtcatctc tccatctac tatcagttga
 541 acaagcgcct tccttctact caccaaatta gtgtggaaca atctatcagc ctcctcctca
 601 actttatgat tgctgcat**at gacagatgtt ttctcccaga tgtcagattc caatggctta
 661 atgatattta gcaagtttga ccagtttctg aaggaagttc tgaagctccc aacagctgtc
 721 tttgaagggc catcttttgg ttacacagag cactcagtcc gcacctgttt ccacagcag
 781 agaaagataa tgctaaatat gtttttagac acaatgatgg ctgaccctcc tccccagtgc
 841 cttgtctggc tacctctcat gcacaggctt gcccatgttg agaatgtctt ccatcccgtg
 901 gagtgctcct actgccgatg tgagagtatg atgggtttcc ggtaccgatg ccagcagtgc
 961 cacaactatc agctctgcca gaattgcttt tggcgtggcc atgccggcgg ccctcacagc
1021 aaccagcacc agatgaagga gcattcctct tggaaatctc ctgcaaagaa gctgagccat
1081 gcaattagta aatctttggg gtgtgtaccc acgagagaac ccccgcatcc tgtttttcct
1141 gagcaaccag agaaaccact tgaccttgca catatagttc ctcctcgccc tctgactaat
1201 atgaatgaca ccatggttag ccacatgtcc tctggagtgc ccactcccac caagagtgtt
1261 ctggacagtc ctagccgact ggatgaggaa caccgtctta tagctcgcta tgctgcccgg
1321 ctggctgcag aagcaggaaa cgtgactcgt cctcccactg acttgagctt taactttgat
1381 gccaacaaac aacaaagaca gcttattgca gaactggaaa acaaaaacag agagatcct**g
1441 **caggagattc agcgtctccg cctggaacac gagcaggcct cccagcccac ccctgagaag
1501 gcacagcaga accccacgct gctggcagag ctgcggctgc tgaggcaaag gaaggatgaa
1561 ctggagcaga ggatgtcggc cctgcaggag agcaggcggg agctgatggt ccagctggaa
1621 gagctgatga agttgctgaa ggaggaagag caaaagcagg cagctcaggc cacagggtca
1681 ccacatacat cgcccaccca tggaggcggc cggccaatgc ccatgccagt gcgctccacg
1741 tctgccggct ccacccccac ccactgtccg caggactcgc tgagcggagt cgggggagac
1801 gtgcaggagg ccttcgcaca aggtacgagg agaaacctcc gcaatgacct gctggtggca
1861 gctgactcca tcaccaacac catgtcatcc ctggtgaagg agctcc**attc agcagaggaa
1921 **ggtgcagagg aagaagaaga gaagatgcag aatgggaaag acagaggtta g**cagaggagc
1981 cggacacaga ggaagctcag gcacagagga cgaggagcaa gctggcgccg acatggcgaa
2041 ggcaaggtct cccccagag gcacattcct ctccatcttt ccaccgcaca cctggaccag
2101 gcttgcaggc tgccagacgt cactccaccc gccagggaga ggggagccag agccggtggg
2161 aagcggggag gggctgcgtg gcacagctag tgggcctccc cctgcacagc cctgcatgta
2221 ctagcacctt catcactccc ctcagggcat ggtctcatct ccgcatcagg aattcacctg
2281 gaggttgaaa agagaaaaga aaaagcacca a
```

Bold = Gene coding sequence
<u>Underlined</u> = Portion of gene incorporated in phage

Peptide encoded by Phage DNA (**Bold** = correct reading frame, <u>underlined</u> = expressed by CDS)

Frame 1:
NSSQIFSRGRCRGRRREDAEWERQRLAEEPDTEEAQAQRTRSKLAPTWRRQGLPPEAHSSPSFHRTPGPGLQAAR
RHSTRQGEGSQSRWEAGRGCVAQLVGLPLHSPACTSTFITPLRAWSHLRIRNSPGG

Frame 2:
IQAKYSAEEGAEEEEEKMQNGKDRG

Frame 3:
FKPNIQQRKVQRKKKRRCRMG<u>KTEVSRGAGHRGSSGTEDEEQAGADMAKARSSPRGTFLSIFPPHTWTRLAGCQ</u>
<u>TSLHPPGRGEPEPVGSGEGLRGTASGPPPAQPCMY</u>

A.3 Full list of LC-MS/MS identified targets

Table A.3.1 Targets of probe **P5**. Comparison with DMSO in solution digestion.
5 µM, MDA-MB-468 cells, whole cell lysate

a = % of protein sequence covered; b = PSM (spectral count); c = number of unique peptides; d = NSAF

UniProt ID	Protein Name	Probe P5				DMSO				Probe/DMSO Ratio			
		a	b	c	d	a	b	c	d	a	b	c	d
P00533	Epidermal growth factor receptor	46	416	43	0.0416	38	123	37	0.0009	1.2	3.4	1.2	47.0
Q9H7Z7	Prostaglandin E synthase 2	33	24	8	0.0077	5	1	1	0.0000	6.6	24.0	8.0	234.4
K7EKP8	Cytosolic acyl coenzyme A thioester hydrolase (Fragment)	21	13	5	0.0056	10	3	3	0.0001	2.1	4.3	1.7	60.2
Q8N2G8	GH3 domain-containing protein	10	6	5	0.0014	2	1	1	0.0000	5.0	6.0	5.0	83.4
Q03518	Antigen peptide transporter 1	12	9	6	0.0013	12	2	2	0.0001	1.0	4.5	3.0	25.0
P04150	Glucocorticoid receptor	10	6	6	0.0009	9	1	1	0.0001	1.1	6.0	6.0	15.4
Q02318	Sterol 26-hydroxylase, mitochondrial	10	4	4	0.0009	3	1	1	0.0000	3.3	4.0	4.0	55.6
Q8TEQ8	GPI ethanolamine phosphate transferase 3	2	7	2	0.0008	1	1	1	0.0000	2.0	7.0	2.0	97.2
H0Y8P3	LIM and calponin homology domains-containing protein 1 (Fragment)	6	5	4	0.0007	1	1	1	0.0000	6.0	5.0	4.0	69.5
H0Y742	SUN domain-containing protein 1 (Fragment)	4	3	2	0.0005	2	1	1	0.0000	2.0	3.0	2.0	41.7

Table A.3.2 Targets of probe **P5**. Comparison with inhibitor **I5** competed sample.
5 µM, MDA-MB-468 cells, whole cell lysate

a = % of protein sequence covered; b = PSM (spectral count); c = number of unique peptides; d = NSAF

UniProt ID	Protein Name	Probe P5				Inhibitor I5				Probe/Inhibitor Ratio			
		a	b	c	d	a	b	c	d	a	b	c	d
P00533	Epidermal growth factor receptor	46	416	43	0.0416	10	24	8	0.0038	4.6	17.3	5.4	10.9
Q9H7Z7	Prostaglandin E synthase 2	33	24	8	0.0077	11	8	3	0.0041	3.0	3.0	2.7	1.9
K7EKP8	Cytosolic acyl coenzyme A thioester hydrolase (Fragment)	21	13	5	0.0056	18	7	4	0.0048	1.2	1.9	1.3	1.2
Q8N2G8	GH3 domain-containing protein	10	6	5	0.0014								
Q03518	Antigen peptide transporter 1	12	9	6	0.0013	3	5	2	0.0012	4.0	1.8	3.0	1.1
P04150	Glucocorticoid receptor	10	6	6	0.0009	12	1	1	0.0013	0.8	6.0	6.0	0.7
Q02318	Sterol 26-hydroxylase, mitochondrial	10	4	4	0.0009								
Q8TEQ8	GPI ethanolamine phosphate transferase 3	2	7	2	0.0008	1	5	1	0.0009	2.0	1.4	2.0	0.9
H0Y8P3	LIM and calponin homology domains-containing protein 1 (Fragment)	6	5	4	0.0007	1	1	1	0.0002	6.0	5.0	4.0	3.2
H0Y742	SUN domain-containing protein 1 (Fragment)	4	3	2	0.0005	4	3	2	0.0008	1.0	1.0	1.0	0.6

Table A.3.3 Targets of probe **P1**. Comparison with DMSO in solution digestion.
5 µM, MDA-MB-468 cells, whole cell lysate

a = % of protein sequence covered; b = PSM (spectral count); c = number of unique peptides; d = NSAF

UniProt ID	Protein name	Probe P1				DMSO				Probe/DMSO Ratio			
		a	b	c	d	a	b	c	d	a	b	c	d
P40939	Trifunctional enzyme subunit alpha, mitochondrial	39	285	21	0.0154	30	41	20	0.0005	1.3	7.0	1.1	34.1
P55084	Trifunctional enzyme subunit beta, mitochondrial	37	170	13	0.0148	38	28	16	0.0005	1.0	6.1	0.8	28.7
P17858	ATP-dependent 6-phosphofructokinase, liver type	33	183	14	0.0097	16	17	8	0.0002	2.1	10.8	1.8	50.8
Q01813	ATP-dependent 6-phosphofructokinase, platelet type	32	180	15	0.0094	11	13	7	0.0001	2.9	13.8	2.1	65.4
Q9NRV9	Heme-binding protein 1	46	33	6	0.0072	22	4	3	0.0002	2.1	8.3	2.0	39.0
Q9BW60	Elongation of very long chain fatty acids protein 1	14	48	3	0.0071	9	7	2	0.0002	1.6	6.9	1.5	32.4
Q9Y6C9	Mitochondrial carrier homolog 2	32	35	3	0.0048	14	5	3	0.0001	2.3	7.0	1.0	33.1
P37268	Squalene synthase	22	45	8	0.0044	17	11	6	0.0002	1.3	4.1	1.3	19.3
Q53GQ0	Very-long-chain 3-oxoacyl-CoA reductase	33	30	8	0.0040	19	10	6	0.0003	1.7	3.0	1.3	14.2
Q8TBX8	Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma	22	39	7	0.0038	8	1	1	0.0001	2.7	39.0	7.0	74.8
Q02978	Mitochondrial 2-oxoglutarate/malate carrier protein	18	29	5	0.0038	8	3	2	0.0001	2.3	9.7	2.5	45.7
Q00765	Receptor expression-enhancing protein 5	16	16	4	0.0035	16	6	3	0.0003	1.0	2.7	1.3	12.6
Q9UM00	Calcium load-activated calcium channel	12	15	2	0.0033	6	3	1	0.0001	2.0	5.0	2.0	30.0
O75844	CAAX prenyl protease 1 homolog	9	34	6	0.0029	9	12	5	0.0002	1.0	2.8	1.2	13.4
P23786	Carnitine O-palmitoyltransferase 2, mitochondrial	22	38	11	0.0024	11	9	6	0.0001	2.0	4.2	1.8	19.9
O75431	Metaxin-2	27	15	4	0.0023	8	2	1	0.0001	3.4	7.5	4.0	30.8
O95573	Long-chain-fatty-acid--CoA ligase 3	17	36	8	0.0021	14	12	9	0.0001	1.2	3.0	0.9	14.2
P28288	ATP-binding cassette sub-family D member 3	11	26	6	0.0016	7	7	5	0.0001	1.5	3.7	1.2	17.5
Q3ZCQ8	Mitochondrial import inner membrane translocase subunit TIM50	13	12	3	0.0014	12	5	3	0.0001	1.1	2.4	1.0	11.3
Q8N118	Cytochrome P450 4X1	7	17	3	0.0014	6	4	3	0.0001	1.2	4.3	1.0	20.1
Q9UJS0	Calcium-binding mitochondrial carrier protein Aralar2	8	21	3	0.0013	5	5	2	0.0001	1.6	4.2	1.5	19.8
O95470	Sphingosine-1-phosphate lyase 1	11	16	5	0.0012	11	1	1	0.0001	1.0	16.0	5.0	18.0
Q9Y6I9	Testis-expressed protein 264	12	8	3	0.0011	11	2	2	0.0001	1.1	4.0	1.5	14.4
O00178	GTP-binding protein 1	10	17	5	0.0010	4	2	2	0.0000	2.6	8.5	2.5	40.1
O75208	Ubiquinone biosynthesis protein COQ9, mitochondrial	12	8	3	0.0010	9	3	2	0.0001	1.4	2.7	1.5	12.0
Q6NUK1	Calcium-binding mitochondrial carrier protein SCaMC-1	13	12	4	0.0010	6	5	3	0.0001	2.2	2.4	1.3	11.3
Q16850	Lanosterol 14-alpha demethylase	10	12	4	0.0010	2	1	1	0.0000	5.1	12.0	4.0	57.4
O95299	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial	9	8	3	0.0009	7	3	2	0.0001	1.2	2.7	1.5	12.6
O15533	Tapasin	5	10	2	0.0009	4	3	2	0.0001	1.3	3.3	1.0	16.4

UniProt ID	Protein name	a	b	c	d	a	b	c	d	a	b	c	d
Q9NPL8	Complex I assembly factor TIMMDC1, mitochondrial	10	6	3	0.0009	4	1	1	0.0000	2.5	6.0	3.0	28.3
Q96S66	Chloride channel CLIC-like protein 1	11	10	3	0.0007	8	3	3	0.0000	1.4	3.3	1.0	15.7
Q8WWC4	m-AAA protease-interacting protein 1, mitochondrial	7	5	2	0.0007	4	2	1	0.0001	1.7	2.5	2.0	11.8
Q96N66	Lysophospholipid acyltransferase 7	9	8	3	0.0007	9	3	3	0.0001	1.0	2.7	1.0	12.6
Q92506	Estradiol 17-beta-dehydrogenase 8	16	4	3	0.0006	5	1	1	0.0000	3.2	4.0	3.0	18.9
Q9NPF4	Probable tRNA N6-adenosine threonylcarbamoyltransferase	13	5	3	0.0006	8	2	2	0.0001	1.6	2.5	1.5	11.8
Q9UQB8	Brain-specific angiogenesis inhibitor 1-associated protein 2	6	8	3	0.0006	5	3	3	0.0000	1.3	2.7	1.0	12.6
Q6NUM9	All-trans-retinol 13,14-reductase	7	7	3	0.0005	3	2	2	0.0000	2.2	3.5	1.5	16.5
Q02318	Sterol 26-hydroxylase, mitochondrial	8	6	3	0.0005	3	1	1	0.0000	2.5	6.0	3.0	28.3
Q8NFQ8	Torsin-1A-interacting protein 2	6	5	2	0.0004	5	2	1	0.0000	1.3	2.5	2.0	11.8
Q9NQC3	Reticulon-4	3	12	3	0.0004	2	4	2	0.0000	1.6	3.0	1.5	14.2
Q9Y512	Sorting and assembly machinery component 50 homolog	5	4	2	0.0004	2	1	1	0.0000	2.6	4.0	2.0	18.9
Q8TEQ8	GPI ethanolamine phosphate transferase 3	2	6	2	0.0002	1	1	1	0.0000	2.0	6.0	2.0	28.3
P33897	ATP-binding cassette sub-family D member 1	5	4	2	0.0002	2	1	1	0.0000	2.3	4.0	2.0	18.9
Q6P2E9	Enhancer of mRNA-decapping protein 4	1	4	2	0.0001	1	1	1	0.0000	1.4	4.0	2.0	18.9

Table A.3.4 Targets of probe **P1**. Comparison with inhibitor **I1** sample
5 μM, MDA-MB-468 cells, whole cell lysate
a = % of protein sequence covered; b = PSM (spectral count); c = number of unique peptides; d = NSAF

UniProt ID	Protein name	Probe P1				Inhibitor I1				Probe/DMSO Ratio			
		a	b	c	d	a	b	c	d	a	b	c	d
P40939	Trifunctional enzyme subunit alpha, mitochondrial	39	285	21	0.0154	41	450	22	0.0175	0.9	0.6	1.0	0.9
P55084	Trifunctional enzyme subunit beta, mitochondrial	37	170	13	0.0148	31	202	13	0.0127	1.2	0.8	1.0	1.2
P17858	ATP-dependent 6-phosphofructokinase, liver type	33	183	14	0.0097	8.8	54	3	0.0021	3.7	3.4	4.7	4.7
Q01813	ATP-dependent 6-phosphofructokinase, platelet type	32	180	15	0.0094	20	100	8	0.0038	1.6	1.8	1.9	2.5
Q9NRV9	Heme-binding protein 1	46	33	6	0.0072	24	19	4	0.0030	1.9	1.7	1.5	2.4
Q9BW60	Elongation of very long chain fatty acids protein 1	14	48	3	0.0071	14	42	3	0.0045	1.0	1.1	1.0	1.6
Q9Y6C9	Mitochondrial carrier homolog 2	32	35	3	0.0048	25	29	3	0.0028	1.3	1.2	1.0	1.7
P37268	Squalene synthase	22	45	8	0.0044	16	21	5	0.0015	1.4	2.1	1.6	3.0
Q53GQ0	Very-long-chain 3-oxoacyl-CoA reductase	33	30	8	0.0040	9.9	36	4	0.0034	3.3	0.8	2.0	1.2
Q8TBX8	Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma	22	39	7	0.0038	12	17	3	0.0012	1.8	2.3	2.3	3.2

ID	Protein											
Q02978	Mitochondrial 2-oxoglutarate/malate carrier protein	18	29	5	0.0038	18	38	5	0.0036	1.0	0.8	1.0	1.1
Q00765	Receptor expression-enhancing protein 5	16	16	4	0.0035	17	14	5	0.0022	0.9	1.1	0.8	1.6
Q9UM00	Calcium load-activated calcium channel	12	15	2	0.0033	12	24	2	0.0038	1.0	0.6	1.0	0.9
O75844	CAAX prenyl protease 1 homolog	9	34	6	0.0029	13	43	6	0.0027	0.7	0.8	1.0	1.1
P23786	Carnitine O-palmitoyltransferase 2, mitochondrial	22	38	11	0.0024	17	42	8	0.0019	1.3	0.9	1.4	1.3
O75431	Metaxin-2	27	15	4	0.0023	27	24	4	0.0027	1.0	0.6	1.0	0.9
O95573	Long-chain-fatty-acid--CoA ligase 3	17	36	8	0.0021	9.7	35	5	0.0014	1.7	1.0	1.6	1.4
P28288	ATP-binding cassette sub-family D member 3	11	26	6	0.0016								
Q3ZCQ8	Mitochondrial import inner membrane translocase subunit TIM50	13	12	3	0.0014	13	10	3	0.0008	1.0	1.2	1.0	1.7
Q8N118	Cytochrome P450 4X1	7	17	3	0.0014	7.3	23	3	0.0013	1.0	0.7	1.0	1.0
Q9UJS0	Calcium-binding mitochondrial carrier protein Aralar2	8	21	3	0.0013	11	24	5	0.0011	0.8	0.9	0.6	1.2
O95470	Sphingosine-1-phosphate lyase 1	11	16	5	0.0012	6.5	7	3	0.0004	1.8	2.3	1.7	3.2
Q9Y6I9	Testis-expressed protein 264	12	8	3	0.0011	8.3	6	2	0.0006	1.4	1.3	1.5	1.8
O00178	GTP-binding protein 1	10	17	5	0.0010	4	7	2	0.0003	2.6	2.4	2.5	3.4
Q6NUK1	Calcium-binding mitochondrial carrier protein SCaMC-1	13	12	4	0.0010	2.5	6	1	0.0004	5.2	2.0	4.0	2.8
O75208	Ubiquinone biosynthesis protein COQ9, mitochondrial	12	8	3	0.0010	3.8	2	1	0.0002	3.2	4.0	3.0	5.5
Q16850	Lanosterol 14-alpha demethylase	10	12	4	0.0010	7.8	7	3	0.0004	1.3	1.7	1.3	2.4
O95299	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial	9	8	3	0.0009								
O15533	Tapasin	5	10	2	0.0009	5.4	10	2	0.0007	1.0	1.0	1.0	1.4
Q9NPL8	Complex I assembly factor TIMMDC1, mitochondrial	10	6	3	0.0009								
Q96S66	Chloride channel CLIC-like protein 1	11	10	3	0.0007	11	14	3	0.0008	1.0	0.7	1.0	1.0
Q8WWC4	m-AAA protease-interacting protein 1, mitochondrial	7	5	2	0.0007	7.9	10	2	0.0010	0.9	0.5	1.0	0.7
Q96N66	Lysophospholipid acyltransferase 7	9	8	3	0.0007	8.7	14	3	0.0009	1.0	0.6	1.0	0.8
Q92506	Estradiol 17-beta-dehydrogenase 8	16	4	3	0.0006	11	6	2	0.0007	1.5	0.7	1.5	0.9
Q9NPF4	Probable tRNA N6-adenosine threonylcarbamoyltransferase	13	5	3	0.0006								
Q9UQB8	Brain-specific angiogenesis inhibitor 1-associated protein 2	6	8	3	0.0006	6.3	6	3	0.0003	1.0	1.3	1.0	1.8
Q6NUM9	All-trans-retinol 13,14-reductase	7	7	3	0.0005	6.6	10	3	0.0005	1.0	0.7	1.0	1.0
Q02318	Sterol 26-hydroxylase, mitochondrial	8	6	3	0.0005	5.1	6	2	0.0003	1.5	1.0	1.5	1.4
Q8NFQ8	Torsin-1A-interacting protein 2	6	5	2	0.0004	4.5	2	1	0.0001	1.4	2.5	2.0	3.5
Q9NQC3	Reticulon-4	3	12	3	0.0004	4.3	25	4	0.0006	0.7	0.5	0.8	0.7
Q9Y512	Sorting and assembly machinery component 50 homolog	5	4	2	0.0004	2.6	2	1	0.0001	2.0	2.0	2.0	2.8
Q8TEQ8	GPI ethanolamine phosphate transferase 3	2	6	2	0.0002	2	8	2	0.0002	1.0	0.8	1.0	1.0
P33897	ATP-binding cassette sub-family D member 1	5	4	2	0.0002								
Q6P2E9	Enhancer of mRNA-decapping protein 4	1	4	2	0.0001	0.6	1	1	0.0000	2.2	4.0	2.0	5.5

198

Table A.3.5 Targets of probe **P5**. Comparison with DMSO in solution digestion.
5 μM, MDA-MB-468 cells, nuclear fraction.
a = % of protein sequence covered; b = PSM (spectral count); c = number of unique peptides; d = NSAF

UniProt ID	Protein name	Probe P5				DMSO				Probe/DMSO Ratio			
		a	b	c	d	a	b	c	d	a	b	c	d
P00533	Epidermal growth factor receptor	46	611	45	0.0469	21	72	17	0.0008	2.1	8.5	2.6	56.3
Q9H7Z7	Prostaglandin E synthase 2	34	56	10	0.0138	23	8	3	0.0007	1.5	7.0	3.3	19.7
P16615	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	24	64	20	0.0057	12	23	11	0.0003	2.0	2.8	1.8	18.5
Q9NX74	tRNA-dihydrouridine(20) synthase [NAD(P)+]-like	34	28	14	0.0053								
Q96F25	UDP-N-acetylglucosamine transferase subunit ALG14 homolog	25	12	5	0.0052								
Q99594	Transcriptional enhancer factor TEF-5	18	24	9	0.0051								
P55265	Double-stranded RNA-specific adenosine deaminase	23	66	29	0.0050	16	26	16	0.0003	1.4	2.5	1.8	16.8
O00483	Cytochrome c oxidase subunit NDUFA4	21	4	2	0.0046								
P22695	Cytochrome b-c1 complex subunit 2, mitochondrial	17	18	7	0.0037	7	7	2	0.0002	2.6	2.6	3.5	17.1
Q9HC21	Mitochondrial thiamine pyrophosphate carrier	24	12	8	0.0035								
Q5VV42	Threonylcarbamoyladenosine tRNA methylthiotransferase	14	20	6	0.0032								
Q9UJG1	Motile sperm domain-containing protein 1	20	7	4	0.0031								
Q6NUM9	All-trans-retinol 13,14-reductase	11	20	6	0.0030								
Q03518	Antigen peptide transporter 1	10	25	7	0.0029								
Q8WV74	Nucleoside diphosphate-linked moiety X motif 8	14	7	2	0.0028								
P10599	Thioredoxin	10	3	2	0.0027								
Q9Y6I9	Testis-expressed protein 264	11	8	3	0.0024								
Q9UJX3	Anaphase-promoting complex subunit 7	13	15	7	0.0023								
Q16850	Lanosterol 14-alpha demethylase	8	12	4	0.0022								
O94901	SUN domain-containing protein 1	14	19	9	0.0022								
Q8N2G8	GH3 domain-containing protein	12	11	7	0.0019								
O15228	Dihydroxyacetone phosphate acyltransferase	9	14	5	0.0019								
O95197	Reticulon-3	6	21	4	0.0019								
Q3KQZ1	Solute carrier family 25 member 35	9	6	3	0.0019								
P08243	Asparagine synthetase [glutamine-hydrolyzing]	16	11	7	0.0018								
Q9H4G4	Golgi-associated plant pathogenesis-related protein 1	18	3	2	0.0018								
Q96CU9	FAD-dependent oxidoreductase domain-containing protein 1	13	9	6	0.0017								
Q3SXM5	Inactive hydroxysteroid dehydrogenase-like protein 1	13	6	4	0.0017								
O15533	Tapasin	7	8	3	0.0017								

Q5TFE4	5'-nucleotidase domain-containing protein 1	12	8	4	0.0016
Q8TEQ8	GPI ethanolamine phosphate transferase 3	8	18	7	0.0015
Q8N2K0	Monoacylglycerol lipase ABHD12	5	6	2	0.0014
Q7L5N7	Lysophosphatidylcholine acyltransferase 2	12	8	5	0.0014
Q9Y6K0	Choline/ethanolaminephosphotransferase 1	5	6	2	0.0013
Q9UPQ0	LIM and calponin homology domains-containing protein 1	8	14	7	0.0012
Q96N66	Lysophospholipid acyltransferase 7	7	6	3	0.0012
O75439	Mitochondrial-processing peptidase subunit beta	11	6	3	0.0011
Q9Y679	Ancient ubiquitous protein 1	6	5	2	0.0011
Q9H857	5'-nucleotidase domain-containing protein 2	3	6	2	0.0011
Q9NQC3	Reticulon-4	3	11	3	0.0009
P53611	Geranylgeranyl transferase type-2 subunit beta	6	3	2	0.0008
Q96GQ5	RUS1 family protein C16orf58	6	4	3	0.0008
Q9H6T0	Epithelial splicing regulatory protein 2	5	6	3	0.0008
Q9Y4W2	Ribosomal biogenesis protein LAS1L	7	6	4	0.0008
Q86X29	Lipolysis-stimulated lipoprotein receptor	5	5	3	0.0007
Q9H0W8	Protein SMG9	8	4	3	0.0007
Q96SI9	Spermatid perinuclear RNA-binding protein	6	5	4	0.0007
Q96HA1	Nuclear envelope pore membrane protein POM 121	5	9	7	0.0007
Q9BU23	Lipase maturation factor 2	6	5	3	0.0007
O95470	Sphingosine-1-phosphate lyase 1	6	4	3	0.0007
O43272	Proline dehydrogenase 1, mitochondrial	6	4	3	0.0006
Q14145	Kelch-like ECH-associated protein 1	5	4	3	0.0006
Q9BTE3	Mini-chromosome maintenance complex-binding protein	6	4	3	0.0006
Q9BRR6	ADP-dependent glucokinase	5	3	2	0.0006
A6NDB9	Paralemmin-3	3	4	2	0.0006
Q96GM8	Target of EGR1 protein 1	7	3	2	0.0005
Q03519	Antigen peptide transporter 2	3	4	2	0.0005
Q8IZL9	Cyclin-dependent kinase 20	9	2	2	0.0005
Q9Y223	Bifunctional UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase	4	4	3	0.0005
P04150	Glucocorticoid receptor	4	4	3	0.0005
Q8IY17	Neuropathy target esterase	3	7	5	0.0005
P50895	Basal cell adhesion molecule	5	3	2	0.0004
Q7Z3C6	Autophagy-related protein 9A	3	4	3	0.0004
Q9UPN9	E3 ubiquitin-protein ligase TRIM33	3	5	3	0.0004
A0FGR8	Extended synaptotagmin-2	5	4	2	0.0004
Q5JTH9	RRP12-like protein	2	5	3	0.0004

P29590	Protein PML	2	3	2	0.0003								
P42695	Condensin-2 complex subunit D3	3	5	4	0.0003								
Q8IZX4	Transcription initiation factor TFIID subunit 1-like	1	6	2	0.0003								
Q86UT6	NLR family member X1	2	3	2	0.0003								
Q8WWM7	Ataxin-2-like protein	2	3	2	0.0003								
Q14669	E3 ubiquitin-protein ligase TRIP12	2	5	4	0.0002								
O95785	Protein Wiz	2	4	3	0.0002								
A0FGR8-2	Isoform 2 of Extended synaptotagmin-2	4	2	3	0.0002								

Table A.3.6 Targets of probe **P5**. Comparison with DMSO in solution digestion.
5 µM, MDA-MB-468 cells, cytosolic/membrane fraction.
a = % of protein sequence covered; b = PSM (spectral count); c = number of unique peptides; d = NSAF

UniProt ID	Protein Name	Probe 5				DMSO				Probe/DMSO Ratio			
		a	b	c	d	a	b	c	d	a	b	c	d
P00533	Epidermal growth factor receptor	46	606	47	0.0407	24	83	21	0.0008	1.9	7.3	2.2	51.6
Q9NX74	tRNA-dihydrouridine(20) synthase [NAD(P)+]-like	30	34	13	0.0056								
P40938	Replication factor C subunit 3	8	22	2	0.0050								
P08243	Asparagine synthetase [glutamine-hydrolyzing]	30	34	14	0.0049								
Q5TFE4	5'-nucleotidase domain-containing protein 1	14	27	5	0.0048								
O00154	Cytosolic acyl coenzyme A thioester hydrolase	23	22	8	0.0047								
P04279	Semenogelin-1	24	26	7	0.0046								
P09382	Galectin-1	28	6	3	0.0036								
P09110	3-ketoacyl-CoA thiolase, peroxisomal	10	13	2	0.0025								
Q9BTE3	Mini-chromosome maintenance complex-binding protein	11	15	5	0.0019								
O95470	Sphingosine-1-phosphate lyase 1	15	12	6	0.0017								
Q9UPQ0	LIM and calponin homology domains-containing protein 1	11	17	9	0.0013								
Q9H0W8	Protein SMG9	11	8	4	0.0013								
P51570	Galactokinase	10	6	4	0.0012								
Q14145	Kelch-like ECH-associated protein 1	9	9	5	0.0012								
P04150	Glucocorticoid receptor	15	11	8	0.0012								
Q6P996	Pyridoxal-dependent decarboxylase domain-containing protein 1	10	11	6	0.0011								
Q9NQC3	Reticulon-4	6	16	3	0.0011	4	6	2	0.0001	1.6	2.7	1.5	18.9
Q8TCG1	Protein CIP2A	6	12	6	0.0011								
Q02383	Semenogelin-2	7	6	2	0.0008								
O95070	Protein YIF1A	9	3	2	0.0008								
Q9UJX3	Anaphase-promoting complex subunit 7	5	6	3	0.0008								
Q6NXE6	Armadillo repeat-containing protein 6	5	5	3	0.0008								

UniProt ID	Protein name				
P52306	Rap1 GTPase-GDP dissociation stimulator 1	5	6	3	0.0008
Q16831	Uridine phosphorylase 1	7	3	2	0.0008
Q9NYL2	Mitogen-activated protein kinase kinase kinase 20	8	7	6	0.0007
Q9BT22	Chitobiosyldiphosphodolichol beta-mannosyltransferase	6	4	3	0.0007
Q03518	Antigen peptide transporter 1	4	6	3	0.0006
Q8TEQ8	GPI ethanolamine phosphate transferase 3	4	8	4	0.0006
Q6PJG6	BRCA1-associated ATM activator 1	5	6	4	0.0006
Q86WR7	Proline and serine-rich protein 2	8	3	2	0.0006
Q9NYL2-2	Isoform 2 of Mitogen-activated protein kinase kinase kinase 20	4	3	3	0.0005
Q9NZB2	Constitutive coactivator of PPAR-gamma-like protein 1	5	6	4	0.0004
O43310-2	Isoform 2 of CBP80/20-dependent translation initiation factor	3	3	3	0.0004
P33981	Dual specificity protein kinase TTK	5	4	3	0.0004
O15228	Dihydroxyacetone phosphate acyltransferase	3	3	2	0.0004
Q9BTW9	Tubulin-specific chaperone D	3	4	3	0.0003
Q6P2E9	Enhancer of mRNA-decapping protein 4	3	4	3	0.0002
Q14207	Protein NPAT	3	4	3	0.0002
Q9Y6K5	2'-5'-oligoadenylate synthase 3	2	3	2	0.0002
Q86XA9	HEAT repeat-containing protein 5A	2	4	3	0.0002
O14981	TATA-binding protein-associated factor 172	1	3	2	0.0001
Q99973	Telomerase protein component 1	1	3	2	0.0001

Table A.3.7 Targets of probe **P1**. Comparison with DMSO in solution digestion.
5 µM, MDA-MB-468 cells, nuclear fraction.
a = % of protein sequence covered; b = PSM (spectral count); c = number of unique peptides; d = NSAF

UniProt ID	Protein name	Probe P1				DMSO				Probe/DMSO Ratio			
		a	b	c	d	a	b	c	d	a	b	c	d
P40939	Trifunctional enzyme subunit alpha, mitochondrial	50	249	34	0.0176	34	80	19	0.0015	1.5	3.1	1.8	12.0
P55084	Trifunctional enzyme subunit beta, mitochondrial	43	114	19	0.0129	30	38	10	0.0011	1.5	3.0	1.9	11.5
Q9BW60	Elongation of very long chain fatty acids protein 1	14	67	3	0.0129								
P17858	ATP-dependent 6-phosphofructokinase, liver type	37	185	20	0.0128								
Q9BRX8	Peroxiredoxin-like 2A	42	49	10	0.0115								
Q01813	ATP-dependent 6-phosphofructokinase, platelet type	33	98	17	0.0067								

P49748	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	42	75	21	0.0062	20	20	8	0.0004	2.1	3.8	2.6	14.4
Q15067	Peroxisomal acyl-coenzyme A oxidase 1	26	64	13	0.0052	18	14	7	0.0003	1.5	4.6	1.9	17.6
Q02978	Mitochondrial 2-oxoglutarate/malate carrier protein	27	25	8	0.0043								
O95139	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 6	30	10	4	0.0042								
Q00325	Phosphate carrier protein, mitochondrial	20	28	7	0.0042	10	11	4	0.0004	2.0	2.5	1.8	9.8
Q9P2X0	Dolichol-phosphate mannosyltransferase subunit 3	24	7	2	0.0041								
Q9Y6C9	Mitochondrial carrier homolog 2	23	18	4	0.0032								
O75844	CAAX prenyl protease 1 homolog	16	27	8	0.0031	10	11	5	0.0003	1.6	2.5	1.6	9.4
Q16891	MICOS complex subunit MIC60	15	41	12	0.0029	13	14	7	0.0003	1.1	2.9	1.7	11.3
Q9Y6N5	Sulfide:quinone oxidoreductase, mitochondrial	13	24	6	0.0029								
P37268	Squalene synthase	22	22	8	0.0028								
Q99640	Membrane-associated tyrosine- and threonine-specific cdc2-inhibitory kinase	19	25	8	0.0027								
P49006	MARCKS-related protein	19	9	3	0.0025								
Q9NZ01	Very-long-chain enoyl-CoA reductase	17	14	6	0.0024								
P04844	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	19	27	9	0.0023	12	11	4	0.0002	1.5	2.5	2.3	9.4
Q9NRV9	Heme-binding protein 1	19	8	3	0.0023								
O43169	Cytochrome b5 type B	17	6	3	0.0022								
Q8TBX8	Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma	24	16	8	0.0020								
Q8N118	Cytochrome P450 4X1	13	19	7	0.0020	3	6	1	0.0002	5.2	3.2	7.0	12.2
Q9Y6I9	Testis-expressed protein 264	19	11	6	0.0019								
Q9H2G2	STE20-like serine/threonine-protein kinase	15	41	15	0.0018								
Q9H3K2	Growth hormone-inducible transmembrane protein	15	11	5	0.0017								
Q96N66	Lysophospholipid acyltransferase 7	13	15	5	0.0017								
Q6IAN0	Dehydrogenase/reductase SDR family member 7B	17	10	4	0.0017								
P36021	Monocarboxylate transporter 8	9	16	3	0.0016								
Q8TCT9	Minor histocompatibility antigen H13	7	11	3	0.0016								
P53007	Tricarboxylate transport protein, mitochondrial	18	9	5	0.0016								
O94804	Serine/threonine-protein kinase 10	14	28	10	0.0016								
O75208	Ubiquinone biosynthesis protein COQ9, mitochondrial	23	9	6	0.0015								
O00116	Alkyldihydroxyacetonephosphate synthase, peroxisomal	19	18	10	0.0015								
Q96S66	Chloride channel CLIC-like protein 1	14	15	7	0.0015								
O95470	Sphingosine-1-phosphate lyase 1	18	14	9	0.0013								
P43490	Nicotinamide phosphoribosyltransferase	19	12	6	0.0013								
Q9UBX3	Mitochondrial dicarboxylate carrier	15	7	4	0.0013								

Q9H1P3	Oxysterol-binding protein-related protein 2	14	11	7	0.0012							
P51648	Fatty aldehyde dehydrogenase	13	11	6	0.0012							
Q6Y1H2	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 2	8	5	2	0.0011							
Q56VL3	OCIA domain-containing protein 2	16	3	2	0.0010							
O00178	GTP-binding protein 1	9	13	5	0.0010							
O60218	Aldo-keto reductase family 1 member B10	9	6	3	0.0010							
Q02318	Sterol 26-hydroxylase, mitochondrial	11	10	5	0.0010							
Q9Y4W6	AFG3-like protein 2	13	15	11	0.0010							
Q9BX68	Histidine triad nucleotide-binding protein 2, mitochondrial	14	3	2	0.0010							
Q9H0U3	Magnesium transporter protein 1	9	6	3	0.0010							
Q969X5	Endoplasmic reticulum-Golgi intermediate compartment protein 1	7	5	2	0.0009							
Q7Z2K6	Endoplasmic reticulum metallopeptidase 1	7	15	5	0.0009	3	6	2	0.0001	2.4	2.5	2.5	9.6
Q9BQL6	Fermitin family homolog 1	11	11	6	0.0009							
Q96GG9	DCN1-like protein 1	8	4	2	0.0008							
P41440	Folate transporter 1	9	9	4	0.0008							
Q9H0X9	Oxysterol-binding protein-related protein 5	10	13	7	0.0008							
P16930	Fumarylacetoacetase	14	6	3	0.0008							
Q9UBT2	SUMO-activating enzyme subunit 2	6	9	3	0.0008							
Q16850	Lanosterol 14-alpha demethylase	7	7	4	0.0007							
P02765	Alpha-2-HS-glycoprotein	5	5	3	0.0007							
Q7L5N7	Lysophosphatidylcholine acyltransferase 2	12	7	5	0.0007							
Q03518	Antigen peptide transporter 1	6	10	3	0.0007							
Q3SXM5	Inactive hydroxysteroid dehydrogenase-like protein 1	7	4	2	0.0007							
P18031	Tyrosine-protein phosphatase non-receptor type 1	11	5	4	0.0006							
E9PIE4	Mitochondrial carrier homolog 2 (Fragment)	14	3	2	0.0006							
Q9UBV2	Protein sel-1 homolog 1	7	9	3	0.0006							
O15533	Tapasin	4	5	2	0.0006							
Q9BUB5	MAP kinase-interacting serine/threonine-protein kinase 1	7	5	3	0.0006							
Q86UE4	Protein LYRIC	5	6	3	0.0006							
Q8NC56	LEM domain-containing protein 2	6	5	3	0.0005							
Q9H3H5	UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase	4	4	3	0.0005							
O95197	Reticulon-3	7	10	4	0.0005							
Q9UBU6	Protein FAM8A1	6	4	2	0.0005							
Q9BZF3	Oxysterol-binding protein-related protein 6	3	9	2	0.0005							
Q10713	Mitochondrial-processing peptidase subunit alpha	6	5	3	0.0005							
Q8NBX0	Saccharopine dehydrogenase-like oxidoreductase	12	4	3	0.0005							

Q96KA5	Cleft lip and palate transmembrane protein 1-like protein	11	5	4	0.0005
O43353	Receptor-interacting serine/threonine-protein kinase 2	7	5	3	0.0005
Q86V85	Integral membrane protein GPR180	6	4	2	0.0005
Q06136	3-ketodihydrosphingosine reductase	6	3	2	0.0005
Q8NBM4	Ubiquitin-associated domain-containing protein 2	10	3	2	0.0005
O94901	SUN domain-containing protein 1	4	7	3	0.0005
Q9BU23	Lipase maturation factor 2	6	6	3	0.0005
Q9NQC3	Reticulon-4	3	10	3	0.0005
Q9H6U8	Alpha-1,2-mannosyltransferase ALG9	5	5	2	0.0004
P23634	Plasma membrane calcium-transporting ATPase 4	8	10	4	0.0004
Q14527	Helicase-like transcription factor	7	8	7	0.0004
O76024	Wolframin	4	7	3	0.0004
Q6P1M0	Long-chain fatty acid transport protein 4	5	5	2	0.0004
Q6ZWT7	Lysophospholipid acyltransferase 2	5	4	2	0.0004
Q8N2K0	Monoacylglycerol lipase ABHD12	6	3	2	0.0004
Q7L576	Cytoplasmic FMR1-interacting protein 1	5	9	5	0.0004
Q969V3	Nicalin	6	4	3	0.0004
O95202	Mitochondrial proton/calcium exchanger protein	8	5	4	0.0004
O43272	Proline dehydrogenase 1, mitochondrial	4	4	2	0.0004
Q16720	Plasma membrane calcium-transporting ATPase 3	5	8	3	0.0004
O96005	Cleft lip and palate transmembrane protein 1	5	4	3	0.0003
Q16851	UTP--glucose-1-phosphate uridylyltransferase	4	3	2	0.0003
Q2TAY7	WD40 repeat-containing protein SMU1	4	3	2	0.0003
Q9Y2U8	Inner nuclear membrane protein Man1	5	5	4	0.0003
P35610	Sterol O-acyltransferase 1	4	3	2	0.0003
O00519	Fatty-acid amide hydrolase 1	4	3	2	0.0003
P27694	Replication protein A 70 kDa DNA-binding subunit	4	3	2	0.0003
Q86X29	Lipolysis-stimulated lipoprotein receptor	4	3	2	0.0002
Q8N5G2	Macoilin	4	3	2	0.0002
Q13045	Protein flightless-1 homolog	3	4	3	0.0002
Q96GD3	Polycomb protein SCMH1	5	2	2	0.0002
O95239	Chromosome-associated kinesin KIF4A	2	3	2	0.0001
Q5JPE7	Nodal modulator 2	2	3	2	0.0001
Q8NEV8	Exophilin-5	3	4	4	0.0001

Table A.3.8 Targets of probe **P1**. Comparison with DMSO in solution digestion.
5 µM, MDA-MB-468 cells, cytosolic/membrane fraction.

a = % of protein sequence covered; b = PSM (spectral count); c = number of unique peptides; d = NSAF

UniProt ID	Protein name	Probe P1				DMSO				Probe/DMSO Ratio			
		a	b	c	d	a	b	c	d	a	b	c	d
P17858	ATP-dependent 6-phosphofructokinase, liver type	43	258	24	0.0267	17	22	8	0.0009	2.5	11.7	3.0	29.7
O00264	Membrane-associated progesterone receptor component 1	23	61	5	0.0252								
Q01813	ATP-dependent 6-phosphofructokinase, platelet type	34	180	17	0.0185	22	30	11	0.0012	1.5	6.0	1.5	15.2
Q9UNL2	Translocon-associated protein subunit gamma	13	34	2	0.0148								
Q9BRX8	Peroxiredoxin-like 2A	38	31	8	0.0109								
Q15738	Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	43	48	14	0.0104	16	8	4	0.0007	2.7	6.0	3.5	15.2
Q9BW60	Elongation of very long chain fatty acids protein 1	14	33	3	0.0095								
Q9NRV9	Heme-binding protein 1	26	17	4	0.0073	20	5	4	0.0008	1.3	3.4	1.0	8.6
P05141	ADP/ATP translocase 2	27	24	9	0.0065	6	4	2	0.0004	4.4	6.0	4.5	15.2
Q8TBX8	Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma	29	29	11	0.0056								
Q9H2G2	STE20-like serine/threonine-protein kinase	21	69	24	0.0045	2	5	2	0.0001	9.8	13.8	12.0	35.0
O94804	Serine/threonine-protein kinase 10	24	50	20	0.0042	7	3	2	0.0002	3.3	16.7	10.0	19.3
P36021	Monocarboxylate transporter 8	10	24	4	0.0036								
P69905	Hemoglobin subunit alpha	17	6	2	0.0034								
Q9NZ01	Very-long-chain enoyl-CoA reductase	15	13	6	0.0034								
O75844	CAAX prenyl protease 1 homolog	20	20	10	0.0034	2	2	1	0.0001	8.5	10.0	10.0	25.3
Q15067	Peroxisomal acyl-coenzyme A oxidase 1	18	27	9	0.0033	3	2	1	0.0001	5.9	13.5	9.0	34.2
P55084	Trifunctional enzyme subunit beta, mitochondrial	27	19	11	0.0032								
Q9BV40	Vesicle-associated membrane protein 8	29	4	3	0.0032								
O43353	Receptor-interacting serine/threonine-protein kinase 2	24	21	9	0.0031								
P40939	Trifunctional enzyme subunit alpha, mitochondrial	26	29	17	0.0031	13	9	5	0.0004	2.0	3.2	3.4	8.2
Q9UM00	Calcium load-activated calcium channel	20	7	4	0.0030								
P37268	Squalene synthase	20	15	8	0.0029								
P51571	Translocon-associated protein subunit delta	18	6	2	0.0028								
O00178	GTP-binding protein 1	18	23	11	0.0028	9	3	2	0.0005	2.0	7.7	5.5	5.7
Q9H1P3	Oxysterol-binding protein-related protein 2	13	16	6	0.0027								
P16930	Fumarylacetoacetase	12	13	4	0.0025	7	3	2	0.0002	1.8	4.3	2.0	11.0
Q9Y6I9	Testis-expressed protein 264	8	7	2	0.0018								
Q8N118	Cytochrome P450 4X1	13	11	6	0.0017	3	2	1	0.0001	4.1	5.5	6.0	12.6
Q8TCT9	Minor histocompatibility antigen H13	5	8	2	0.0017								
O95433	Activator of 90 kDa heat shock protein ATPase homolog 1	13	7	5	0.0017								

Q96LJ7	Dehydrogenase/reductase SDR family member 1	11	6	2	0.0015								
P41440	Folate transporter 1	6	11	3	0.0015								
O95470	Sphingosine-1-phosphate lyase 1	11	10	6	0.0014								
Q15392	Delta(24)-sterol reductase	7	9	4	0.0014	2	2	1	0.0001	4.5	4.5	4.0	11.4
Q9BUB5	MAP kinase-interacting serine/threonine-protein kinase 1	8	8	3	0.0014								
P02765	Alpha-2-HS-glycoprotein	7	6	4	0.0013								
Q12933	TNF receptor-associated factor 2	7	8	4	0.0013	2	2	1	0.0001	3.6	4.0	4.0	10.1
Q02978	Mitochondrial 2-oxoglutarate/malate carrier protein	14	5	4	0.0013								
O43826	Glucose-6-phosphate exchanger SLC37A4	8	6	3	0.0011								
Q9NTJ5	Phosphatidylinositide phosphatase SAC1	10	8	5	0.0011								
Q9Y6N5	Sulfide:quinone oxidoreductase, mitochondrial	7	6	3	0.0011								
Q96S66	Chloride channel CLIC-like protein 1	8	7	4	0.0010								
Q9UJC3	Protein Hook homolog 1	9	8	4	0.0009	5	9	2	0.0004	1.8	0.9	2.0	2.3
Q9NQC3	Reticulon-4	4	13	5	0.0009								
Q96N66	Lysophospholipid acyltransferase 7	6	5	2	0.0009								
Q9H0X9	Oxysterol-binding protein-related protein 5	6	9	5	0.0008								
Q16850	Lanosterol 14-alpha demethylase	4	5	2	0.0008								
Q9NVX7	Kelch repeat and BTB domain-containing protein 4	6	5	2	0.0008								
Q6P1M0	Long-chain fatty acid transport protein 4	7	6	3	0.0008								
O00116	Alkyldihydroxyacetonephosphate synthase, peroxisomal	10	6	5	0.0007								
Q86V85	Integral membrane protein GPR180	6	4	2	0.0007								
Q9H0U3	Magnesium transporter protein 1	5	3	2	0.0007								
O43143	Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	5	6	4	0.0006								
Q96KA5	Cleft lip and palate transmembrane protein 1-like protein	9	4	3	0.0006								
P35610	Sterol O-acyltransferase 1	6	4	3	0.0006								
Q9BT22	Chitobiosyldiphosphodolichol beta-mannosyltransferase	5	3	2	0.0005								
P49748	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	4	4	2	0.0005								
Q7L5N7	Lysophosphatidylcholine acyltransferase 2	4	3	2	0.0004								
P08842	Steryl-sulfatase	5	3	2	0.0004								
Q9NZB2	Constitutive coactivator of PPAR-gamma-like protein 1	4	5	4	0.0004	2	2	1	0.0001	2.8	2.5	4.0	6.3
Q13233	Mitogen-activated protein kinase kinase kinase 1	1	3	2	0.0002								