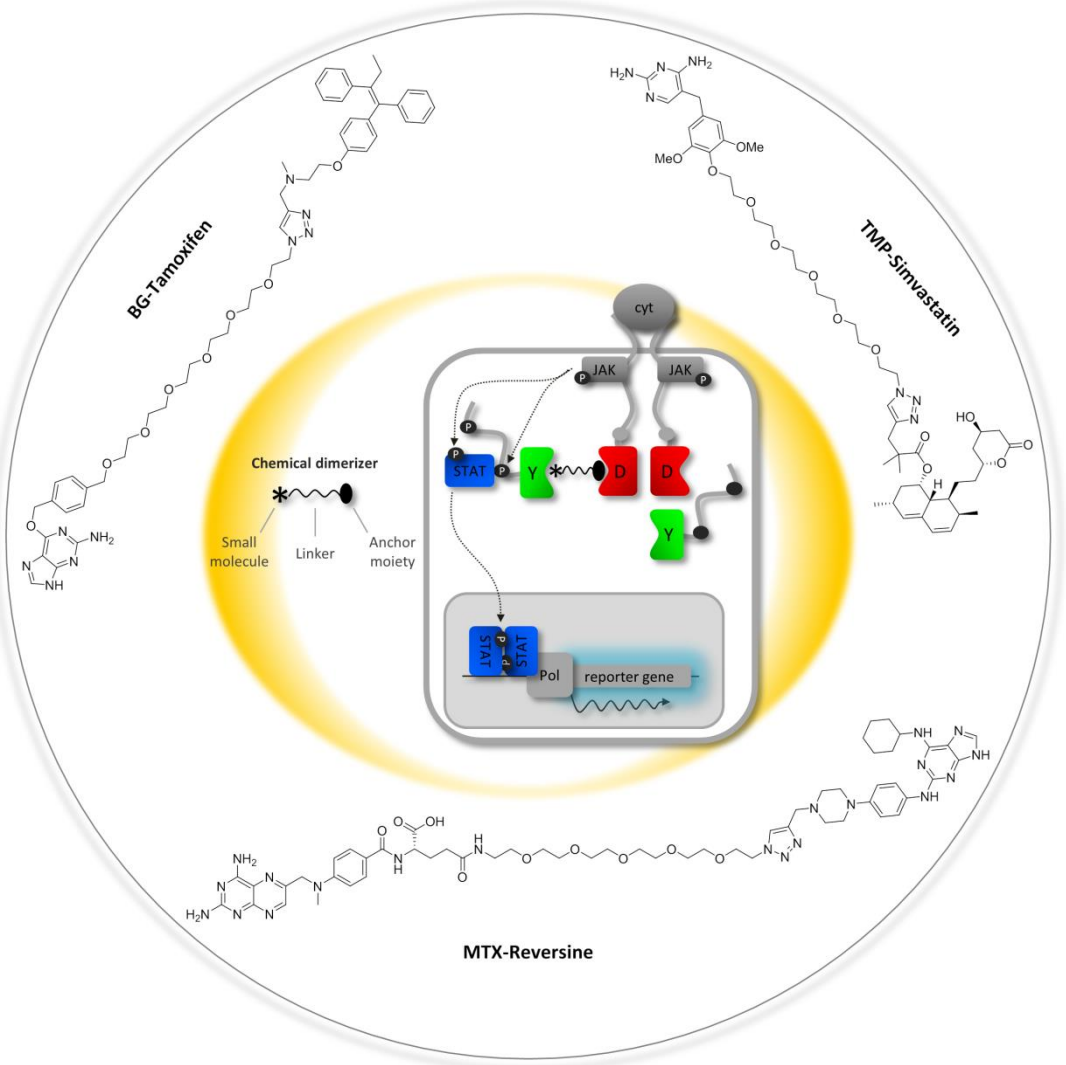
UNIVERSITEIT GENT

FACULTY OF **PHARMACEUTICAL SCIENCES**

Synthesis of Chemical Dimerizers for the Optimization of MASPIT



Dries De Clercq

Promoters:

Prof. Dr. S. Van Calenbergh Prof. Dr. J. Tavernier Dr. S. Lievens

Academic year 2015-2016

Thesis submitted in fulfillment of the requirements for the degree of Doctor in Pharmaceutical Sciences
Proefschrift voorgelegd tot het bekomen van de graad van Doctor in de Farmaceutische Wetenschappen

Synthesis of Chemical Dimerizers for the Optimization of MASPIT

Dries De Clercq

Ghent University
Faculty of Pharmaceutical Sciences
Laboratory for Medicinal Chemistry

Promoters:
Prof. Dr. S. Van Calenbergh
Prof. Dr. J. Tavernier
Dr. S. Lievens

Academic year 2015-2016

Thesis submitted in fulfillment of the requirements for the degree of Doctor in Pharmaceutical Sciences
Proefschrift voorgelegd tot het bekomen van de graad van Doctor in de Farmaceutische Wetenschappen

Composition of the Jury

Chairman:
Prof. Dr. Filip De Vos
Laboratory of Radiopharmacy, Faculty of Pharmaceutical Sciences, Ghent University

Promoters:
Prof. Dr. Serge Van Calenbergh
Laboratory for Medicinal Chemistry, Faculty of Pharmaceutical Sciences, Ghent University

Prof. Dr. Jan Tavernier
Cytokine Receptor Lab, VIB & Ghent University

Dr. Sam Lievens
Cytokine Receptor Lab, VIB & Ghent University

Members of the Examination Committee:
Dr. Jurgen Joossens
Department of Pharmaceutical Sciences, University of Antwerp

Prof. Dr. Katrien Remaut
Laboratory for General Biochemistry and Physical Pharmacy, Faculty of Pharmaceutical Sciences, Ghent University

Prof. Dr. Steven Verhelst
Laboratory of Chemical Biology, Faculty of Medicine, KU Leuven

Dr. Richard Wombacher
Chemical Biology Laboratory, Institute of Pharmacy and Molecular Biotechnology, Heidelberg University

Dries De Clercq was supported by a PhD Grant from the Special Research Fund of Ghent University (BOF).

Table of Contents

Summary

The identification of the molecular targets and mechanisms underpinning the beneficial or detrimental effects of small-molecule leads and drugs constitutes a crucial aspect of current drug discovery. Over the last two decades, three-hybrid (3H) systems have progressively taken an important position in the armamentarium of target protein profiling technologies for small molecules. MASPIT (mammalian small molecule-protein interaction trap) is a mammalian three-hybrid (M3H) system that enables the identification of cytosolic target proteins of small molecules of interest in intact human cells. MASPIT is the 3H component of the MAPPIT (mammalian protein-protein interaction trap) technology platform, an approach that uses the JAK-STAT pathway that mediates cytokine receptor (CR) signaling. MASPIT employs *E. coli* dihydrofolate reductase (eDHFR), fused to a mutated, functionally inactivated, CR to present a methotrexate (MTX) fusion compound (MFC) to the intracellular environment. This allows screening of a small-molecule bait against a collection of chimeric prey proteins. As a result of the interaction between the small-molecule bait and a target prey protein, the JAK-STAT pathway is activated, resulting in the expression of a reporter gene.

A prerequisite for successful MASPIT analysis is the availability of appropriate bifunctional small-molecule probes, capable of artificially bringing two (chimeric) proteins together to form a stable ternary complex. A comprehensive overview of these so-called chemical inducers of dimerization (CIDs) specifically applied in both yeast three-hybrid (Y3H) and M3H systems for small molecule-target protein profiling is provided in Chapter 1. Furthermore, examples for typical components of chemical dimerizers for 3H systems are discussed. Additionally, a number of variations on the classical 3H themes as well as 3H-based screening platforms are illustrated.

The aim of this thesis was to optimize the performance of MASPIT from a medicinal chemical perspective. Chapter 3 describes a versatile, clickable MTX reagent that allows swift γ-selective conjugation to yield MFCs appropriate for MASPIT. The general applicability of the reagent was demonstrated using three structurally diverse pharmacologically active compounds, i.e. tamoxifen, reversine, and FK506. In analytical mode, MASPIT produced concentration-dependent reporter signals for the established target proteins of these model baits. Furthermore, in a proof-of-concept experiment, the FK506 MFC was explored in a cell array screen for targets of FK506. Out of a pilot collection of nearly 2000 full-length human ORF (open reading frame) preys, FKBP12 was selectively identified as an interaction partner of FK506, thereby further validating the MASPIT system and showing its potential towards uncovering new intracellular targets of small molecules of interest.

Two alternative chemical dimerizer approaches aimed at increasing the sensitivity of MASPIT are evaluated in Chapter 4. To circumvent any potential limitations related to the tight binding of MTX to endogenous human DHFR, trimethoprim (TMP) was explored as an alternative prokaryote-specific eDHFR ligand. MASPIT evaluation of TMP fusion compounds (TFCs) with tamoxifen, reversine, and simvastatin as model baits, resulted in dose-response curves being shifted towards lower EC_{50} values than those of their MTX congeners. Additionally, a scalable TMP-azido reagent was synthesized that displayed a similar improvement in sensitivity, possibly owing to increased membrane permeability relative to the MTX anchor.

Furthermore, to stabilize the ternary complex (CR-CID-Prey), the concept of covalent bonding was introduced into MASPIT, which hitherto relied on reversible interactions of either component of the dimerizers. As a starting point, a fusion compound was explored that allows selective and covalent immobilization to the CR by using a SNAP-tag-based system. Unexpectedly, this approach did not yield the hypothesized increase in sensitivity, as the O^6-benzylguanine-tamoxifen heterodimerizer produced an inferior read-out compared to its MTX or TMP congeners.

In a next step, the implementation of covalent bonding on the bait-end of the dimerizers was explored, as described in Chapter 5. For this purpose, the optimized TMP-azido reagent was equipped with a double ligation handle that enabled sequential introduction of both the tamoxifen pilot bait and photoactivatable crosslinkers, thereby affording heterotrimeric photoaffinity labeling (PAL) probes. However, evaluation of these PAL probes in MASPIT turned out to be less straightforward than anticipated and no clear-cut proof of effective photocrosslinking was obtained thus far.
Therefore, taking a short cut, a test system based on a panel of irreversible and reversible EGFR kinase inhibitor TFCs was explored to assess the contribution of covalent bait-prey interactions to the MASPIT read-out. Unfortunately, preliminary biological experiments yielded contradictory outcomes, which warrant further investigation.

Finally, Chapter 7 is dedicated to the discussion and critical evaluation of MASPIT, and 3H systems in general, within the broader context of the most frequently applied strategies to small-molecule target deconvolution in chemical biology and drug discovery research. Furthermore, the relevance and potential of 3H approaches is discussed. To conclude, future perspectives regarding anchor moieties and conjugation methods applicable to the synthesis of chemical dimerizers for 3H applications are outlined.

Samenvatting – Summary in Dutch

De identificatie van de moleculaire doelwitten en mechanismen, die aan de basis liggen van de gunstige en schadelijke effecten van kleine therapeutische *lead* verbindingen en geneesmiddelen, is cruciaal in geneesmiddelenontdekking. In de afgelopen twee decennia hebben drie-hybride (3H) systemen een belangrijke positie ingenomen in het arsenaal van technologieën die het interactieprofiel van kleine moleculen met hun doelwiteiwitten onderzoeken. MASPIT (*mammalian small molecule-protein interaction trap*) is een zoogdier-gebaseerd 3H systeem dat de identificatie van cytosolische doelwiteiwitten van interessante kleine moleculen mogelijk maakt in intacte humane cellen. MASPIT is een 3H component van het MAPPIT (*mammalian protein-protein interaction trap*) technologieplatform, een systeem dat gebruik maakt van de JAK-STAT cascade, die cytokine receptor (CR) signalisatie medieert. MASPIT gebruikt *E. coli* dihydrofolaatreductase (eDHFR) dat gefuseerd is aan een gemuteerde, functioneel geïnactiveerde CR om een methotrexaat (MTX) fusiecomponent (MFC) aan het intracellulaire milieu aan te bieden. Dit laat toe om een kleine lokaasmolecule te screenen tegen een collectie van chimere prooi-eiwitten. De interactie tussen de kleine lokaasmolecule en een doelwit-prooiproteïne resulteert in activatie van de JAK-STAT cascade en daaropvolgende expressie van een reportergen.

Voor succesvolle MASPIT analyse dient men te beschikken over geschikte bifunctionele synthetische probes, die in staat zijn om twee (chimere) eiwitten artificieel samen te brengen ter vorming van een stabiel ternair complex. Een uitvoerig overzicht van deze zogenaamde chemische induceerders van dimerisatie (CIDs) die tot op heden specifiek aangewend werden in zowel gist- als zoogdier-gebaseerde 3H systemen voor het in kaart brengen van kleine molecuul-doelwitproteïne interacties is weergeven in Hoofdstuk 1. Verder worden voorbeelden van de typische componenten van chemische dimerizeerders voor 3H systemen besproken. Daarnaast worden enkele variaties op de klassieke 3H systemen, alsook 3H-gebaseerde screeningsplatformen toegelicht.

De doelstelling van dit proefschrift was het optimaliseren van de performantie van MASPIT vanuit een medicinaal chemische invalshoek. Hoofdstuk 3 beschrijft de ontwikkeling van een veelzijdig MTX reagens dat via *click*-chemie snelle, γ-selectieve conjugatie toelaat ter vorming van MFCs die geschikt zijn voor toepassing in MASPIT. De algemene toepasselijkheid van het reagens werd aangetoond voor drie structureel uiteenlopende farmacologisch actieve verbindingen, namelijk tamoxifen, reversine, en FK506. In analytische modus produceerde MASPIT concentratieafhankelijke reportersignalen voor de gekende doelwiteiwitten van deze model lokaasmoleculen. Voorts werd, in een *proof-of-concept* experiment, de FK506 MFC onderworpen aan een cel *array* screen, op zoek naar doelwitten van FK506. Uit een initiële collectie van nagenoeg 2000 *full-length* humane ORF (*open reading frame*) prooi-eiwitten werd FKBP12 selectief geïdentificeerd als interactiepartner van FK506. Zo werd het MASPIT systeem verder gevalideerd, alsook zijn vermogen om nieuwe intracellulaire doelwitten van kleine therapeutische verbindingen te onthullen.

Hoofdstuk 4 beschrijft de synthese en evaluatie van twee alternatieve CIDs, die als doel hebben de gevoeligheid van MASPIT te verhogen. Om de potentiële beperkingen gerelateerd aan de sterke binding van MTX aan endogeen humaan DHFR te omzeilen, werd trimethoprim (TMP) als alternatief prokaryoot-specifiek eDHFR ligand onderzocht. MASPIT evaluatie van TMP fusiecomponenten (TFCs) met tamoxifen, reversine, en simvastatine als model lokaasmoleculen, resulteerde in verschuiving

van de dosis-respons curves naar lagere EC_{50} waarden in vergelijking met de corresponderende MTX analogen. Bijkomend werd een schaalbaar TMP-azido reagens gesynthetiseerd dat een gelijkaardige verbetering in gevoeligheid vertoonde, mogelijks toe te schrijven aan zijn verhoogde membraanpermeabiliteit ten opzichte van het MTX anker.
Met het oog op het stabiliseren van het ternaire complex (CR-CID-prooiproteïne), werd verder het concept van covalentie onderzocht in MASPIT, dat tot dusver gestoeld was op reversibele interacties van elk van de componenten van de dimerizeerders. Als vertrekpunt werd getracht de fusiecomponent selectief en covalent te immobiliseren op de CR door middel van een SNAP-tag-gebaseerd systeem. Deze strategie leverde onverwachts niet de veronderstelde toename in gevoeligheid op, gezien de O^6-benzylguanine-tamoxifen heterodimerizeerder een lagere readout produceerde vergeleken met zijn MTX of TMP analogen.

In een volgende fase werd de implementatie van covalentie aan de kant van de lokaasmolecule van de dimerizeerders onderzocht, zoals uiteengezet in Hoofdstuk 5. Hiertoe werd het geoptimaliseerde TMP-azido reagens uitgerust met een dubbel ligatie-handvat dat sequentiële invoering van zowel het tamoxifen modellokaas als fotoactiveerbare *crosslinkers* mogelijk maakte. Deze synthesestrategie leverde heterotrimere fotoaffiniteitslabeling (PAL) probes op. Echter, de evaluatie van deze PAL probes in MASPIT bleek minder voor de hand liggend dan vooropgesteld en tot dusver werd geen uitgesproken bewijs van effectieve fotocrosslinking verkregen.
Daarom werd besloten een shortcut te nemen en een testsysteem, gebaseerd op een reeks irreversibele en reversibele EGFR kinase inhibitor TFCs, te ontwikkelen om de bijdrage van covalente lokaas-prooi interacties tot de MASPIT readout te bepalen. Helaas leverden preliminaire biologische experimenten tegenstrijdige resultaten op, die verder onderzoek vereisen.

Tot slot worden in Hoofdstuk 7 MASPIT, en 3H systemen in het algemeen, kritisch besproken en geëvalueerd binnen de bredere context van de meest frequent toegepaste strategieën voor *target* identificatie van kleine therapeutische verbindingen in het onderzoeksdomein van de chemische biologie en geneesmiddelenontdekking. Voorts worden de relevantie en het potentieel van 3H systemen uiteengezet. Ten slotte worden enkele toekomstperspectieven geschetst met betrekking tot immobiliserende groepen en conjugatiemethoden, toepasbaar voor de synthese van chemische dimerizeerders voor drie-hybride toepassingen.

List of Abbreviations

3H	Three-hybrid
ABPP	Activity-based protein profiling
ACE-Cl	α-Chloroethyl chloroformate
ADME-Tox	Absorption, distribution, metabolism, elimination and toxicity
aq.	Aqueous
ATP	Adenosine triphosphate
BG	O^6-Benzylguanine
BGFC	O^6-Benzylguanine fusion compound
Boc	*tert*-Butoxycarbonyl
BODIPY	Boron dipyrromethene
br	Broad
calcd.	Calculated
cat.	Catalytic amount
CCCP	Compound-centric chemical proteomics
COSY	Correlation spectroscopy
C_q	Quaternary carbon atom
CR	Cytokine receptor
CuAAC	Copper-catalyzed azide-alkyne cycloaddition
d	Doublet
Da	Dalton
DAD	Diode array detection
DBAD	Di-*tert*-butyl azodicarboxylate
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCE	1,2-Dichloroethane
dd	Double doublet
ddd	Double double doublet
DHFR	Dihydrofolate reductase
DIPEA	*N*,*N*-Diisopropylethylamine
DMAP	4-(Dimethylamino)-pyridine
DMEM	Dulbecco's modified Eagle's medium
DMF	*N*,*N*-Dimethylformamide
DMSO	Dimethyl sulfoxide
dt	Double triplet
E. coli	*Escherichia coli*
EDC	*N*-(3-Dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride
EGFR	Epidermal growth factor receptor
eq.	Molar equivalent(s)
ER1	Estrogen receptor α
Et	Ethyl
Et_2O	Diethyl ether
Et_3N	Triethylamine
EtOAc	Ethyl acetate
EtOH	Ethanol
g	Gram(s)
h	Hour(s)
hAGT	Human O^6-alkylguanine-DNA alkyltransferase
HATU	O-(7-Azabenzotriazol-1-yl)-*N*,*N*,*N*',*N*'-tetramethyluronium hexafluorophosphate

HEK	Human embryonic kidney cell line
HMBC	Heteronuclear multiple bond correlation
HMGCR	3-Hydroxy-3-methylglutaryl-CoA reductase
HPLC	High pressure/performance liquid chromatography
HRMS	High resolution mass spectrometry
HSQC	Heteronuclear single-quantum coherence spectroscopy
Hz	Hertz
hν	Irradiation with light
IC_{50}	Inhibitor concentration resulting in 50% inhibition of enzyme activity
J	Coupling constant
JAK	Janus kinase
K_D	Dissociation constant
K_i	Inhibition constant
KO*t*Bu	Potassium *tert*-butoxide
LDA	Lithium diisopropylamide
m	Multiplet
M	Molar
M3H	Mammalian three-hybrid
MAPPIT	Mammalian protein-protein interaction trap
MASPIT	Mammalian small molecule-protein interaction trap
Me	Methyl
MeOH	Methanol
MFC	Methotrexate fusion compound
mg	Milligram(s)
MHz	Megahertz
min	Minute(s)
mL	Milliliter(s)
mmol	Millimole(s)
MOA	Mechanism of action
MS	Mass spectrometry
Ms	Methanesulfonyl
mTOR	Mammalian target of rapamycin
MTX	Methotrexate
NaH	Sodium hydride
NBS	*N*-Bromosuccinimide
NHS	*N*-Hydroxysuccinimide
NMR	Nuclear magnetic resonance spectroscopy
NOESY	Nuclear Overhauser effect spectroscopy
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PAL	Photoaffinity labeling
PCR	Polymerase chain reaction
Pd/C	Palladium on activated charcoal
PEG	Poly(ethylene glycol)
Ph	Phenyl
PI3K	Phosphoinositide 3-kinase
PKC	Protein kinase C
PLCγ	Phospholipase C-gamma
ppm	Parts per million

PTM	Post-translational modification
PyBOP	(Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
q	Quartet
quant.	Quantitative
RP	Reversed-phase
RT	Room temperature
s	Singlet
SAR	Structure-activity relationship
sat.	Saturated
SDS	Sodium dodecyl sulphate
STAT	Signal transducer and activator of transcription
t	Triplet
TAM	Tamoxifen
TBAF	Tetra-*n*-butylammonium fluoride
TBDMS	*tert*-Butyldimethylsilyl
TBTA	Tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine
*t*Bu	*tert*-Butyl
*t*BuOH	*tert*-Butanol
Tf, Triflyl	Trifluoromethanesulfonyl
TFA	Trifluoroacetic acid
TFC	Trimethoprim fusion compound
THF	Tetrahydrofuran
TKI	Tyrosine kinase inhibitor
TLC	Thin-layer chromatography
TMP	Trimethoprim
TMS	Tetramethylsilane
TOCSY	Total correlation spectroscopy
TPTU	2-(2-Pyridon-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate
t_R	Retention time
Ts, Tosyl	*p*-Toluenesulfonyl
UV	Ultraviolet
v	Volume
Y2H	Yeast two-hybrid
Y3H	Yeast three-hybrid
δ	Chemical shift
Δ	Heating (reflux)

Note: the one and three letter abbreviations for the amino acids follow the recommendations of IUPAC. *J. Biol. Chem.* **1968**, *243*, 3557-3559 and *J. Biol. Chem.* **1972**, *247*, 977-983.

CHAPTER I

GENERAL INTRODUCTION

Chemical Dimerizers in Three-Hybrid Systems for Small Molecule-Target Protein Profiling

The content of this chapter was derived from:
D. J. H. De Clercq, J. Tavernier, S. Lievens, S. Van Calenbergh, *ACS Chemical Biology* **2015**, *under review*.

I. Chemical Dimerizers in Three-Hybrid Systems for Small Molecule-Target Protein Profiling

I.1. Introduction

Drug discovery research has mainly relied on two strategies to identify potential drug candidates, molecular and empirical approaches.[1] The latter relate to phenotypic screening, which implies the measurement of phenotypic responses to small bioactive modulators in complex biological systems, such as intact cells, isolated organs or model organisms. Traditional, unbiased phenotype-based drug discovery has been applied for over a century and successfully led to the development of numerous important drugs.[2] However, with molecular biology and biochemistry moving to the center stage since the genomics era in the 1990s,[3] drug discovery predominantly concentrated on molecular approaches. These include hypothesis-driven target-based approaches in which compounds that interact with a validated (protein) target are identified via high-throughput screening. Subsequently, these hit compounds are optimized with respect to potency, selectivity and ADME-Tox properties in further stages of the drug discovery process (Figure I.1).

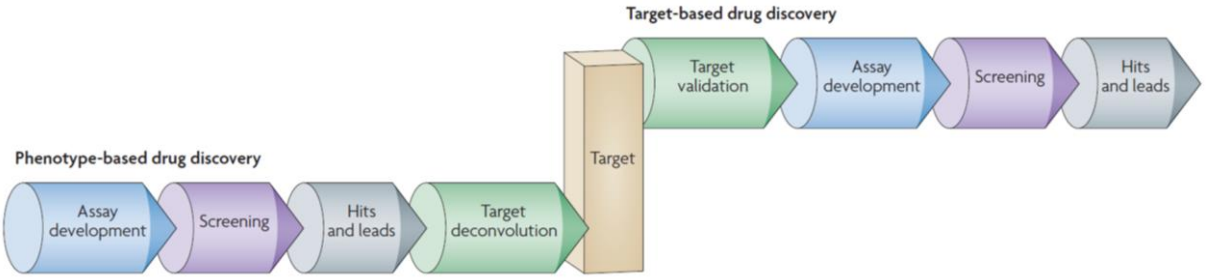


Figure I.1: Schematic diagram of phenotype-based versus target-based drug discovery. The phenotype-based approach is initiated by hit and lead discovery, followed by identification of the molecular targets that underlie their observed phenotypic effects. The target-based strategy starts from validated targets to identify hits and leads via extensive assays and screenings. Figure from: *Nature Reviews Drug Discovery* 2007, 6 (11), 891-903.

Yet, many clinically approved drugs have been found to be more promiscuous than originally thought, with every drug interacting on average with six targets.[4] Therefore, the 'one gene, one drug, one disease' core assumption that frames target-based drug discovery is largely outdated.[5] Furthermore, phenotypic screening initially was suggested to be the more successful strategy for the discovery of small-molecule, first-in-class drugs in an era in which the major focus was on targets (i.e. 1999-2008).[6] Indeed, target-based approaches for first-in-class drugs were postulated to be the root cause for high attrition rates and low productivity, which pharmaceutical R&D is currently facing. This was attributed to the fact that target-based screening focuses narrowly on a particular target, thereby inherently failing to address the molecular complexities underlying drug action.[7],[1]

However, a recent analysis of the origins of all first-in-class drugs approved by the US Food and Drug Administration (FDA) from 1999 to 2013 significantly deviates from this previous report,[1],[6] as it shows that the majority (70%) were discovered through target-based approaches.[8] This discrepancy originates from several factors. Firstly, the important five year extension over the former analysis (i.e. 2013 vs. 2008) illustrates that owing to the long time frames of drug development, the relatively new technologies underpinning target-based approaches (such as the sequencing of the human genome) may only be starting to substantially affect drug approvals and some even suggest that the

best is yet to come.[8] Secondly, different terminologies and definitions have been used to describe the drug discovery approaches, phenotypic screening in particular, which influences interpretation of the data. Nevertheless, elucidating the exact origin of new drugs is less important to drive drug development than establishing the causal links between target engagement, molecular effects on cell biology and the desired phenotypic response.[9] Toward this end, current drug discovery programs should not rely on one single screening approach, but rather combine them sensibly and productively in order to deliver drugs with optimal efficacy/safety as recently exemplified by the introduction of the 'mechanism-informed phenotypic drug discovery' (MIPDD) paradigm in the oncology field, which implies a mechanistically defined, clinically relevant phenotypic screening and ensuing optimization approach.[9] This way phenotypic screening might contribute by expanding the discovery scope from a single molecular target to multiple processes, signaling pathways and complex molecular functions related to untreated or understudied diseases.[10] The subsequent identification of the target spectrum of active hits discovered in such phenotypic screens, i.e. target deconvolution, constitutes an important aspect of current drug discovery.[11],[12]

Over the past decades, numerous case studies within the target profiling field that apply, for example, chemical proteomics-based methods have proved successful. However, despite these success stories, to date target identification often remains an important bottleneck in drug discovery, as a generally applicable methodology is still lacking.[13],[14] In case the target protein of interest is low abundant or unstable, expression-cloning-based methods, such as yeast[15] and mammalian[16] three-hybrid systems, can be a valuable alternative to classical pulldown approaches.[12] However, a *conditio sine qua non* for successful three-hybrid analysis is the availability of appropriate bifunctional small-molecule probes, capable of artificially bringing two proteins together to form a stable ternary complex - termed chemical inducers of dimerization (CIDs),[17] or chemical dimerizers. Thorough reviews providing a broader overview on CID applications[18],[19],[20] as well as those summarizing yeast 'N'-hybrid systems[21],[22] have been recently published. However, reports elaborating on a combination of these themes, more specifically the use of CIDs in three-hybrid systems, are scarce and go back more than a decade.[23] Therefore, in this chapter we aim to fill this gap, focusing on the design of CIDs and their specific application in both yeast and mammalian three-hybrid systems for small molecule-target protein profiling within the broader scope of target deconvolution and drug discovery.

I.2. Chemically induced dimerization

The ability to bind two proteins or protein domains simultaneously represents the main characteristic of chemical inducers of dimerization. Homodimerizers are symmetrical bifunctional ligands that act to bring two identical proteins together. If two different proteins are recognized, the molecules are referred to as heterodimerizers.[18] Historically, mechanism of action studies on the immunosuppressive drugs FK506, rapamycin and cyclosporin A (CsA) initiated the development of the concept of chemically induced dimerization.[24] These studies revealed that the latter compounds act as naturally occurring heterodimerizers, since they first form a binary complex with their respective immunophilin, which subsequently competitively binds to and concomitantly inhibits a second protein enzyme. More specifically, FK506 and rapamycin both bind to FK506 binding protein 12 (FKBP12), after which the complex recruits calcineurin[25] or the FKBP12-rapamycin-binding (FRB)

domain of FRAP (FKBP12-rapamycin-associated protein),[26] respectively. In the case of CsA, complex formation with cyclophilin is followed by interaction with calcineurin.[25]

The first application of the CID technology was introduced in 1993 by the Schreiber and Crabtree groups, reporting on the use of cell permeable, synthetic ligands to control an endogenous signal transduction pathway.[17] In particular, an artificial dimer of FK506, FK1012, was constructed to homodimerize chimeric FKBP12 proteins fused to the cytoplasmic ζ subunit of the T-cell receptor, allowing ligand-regulated signal transmission and specific target gene activation. Since this seminal paper, the chemical toolbox to induce protein-protein interactions has been greatly expanded,[27] with prominent roles reserved for both FK506, rapamycin and their synthetic analogues.[28],[29] A comprehensive general overview of the CID field is beyond the scope of this chapter. Rather, in the following sections we focus on the spectrum of dimerizer systems that has been applied in yeast as well as mammalian three-hybrid assays over the last two decades.

I.3. Yeast three-hybrid systems

The yeast three-hybrid (Y3H) system[15] is an extension of the original yeast two-hybrid (Y2H) assay,[30] which was introduced in 1989 by Fields and Song for the identification of protein-protein interactions. In this genetic system the interaction between two proteins of interest (X and Y) is detected via the reconstitution of an active transcription factor, which has been split into a separate DNA binding domain (DBD) and a transcriptional activation domain (AD), and the subsequent activation of a reporter gene under the control of this transcription factor (Figure I.2A).[31]

As the name suggests, the three-hybrid version includes a third hybrid component, i.e. a CID consisting of an anchor moiety which is covalently coupled to a small molecule of interest via an appropriate linker. Accordingly, this adaptation of the Y2H assay shifts the application scope from protein-protein interactions to the detection of small molecule-target protein interactions. In the Y3H assay, the small molecule of interest is displayed as 'bait' through a high affinity interaction of the anchor moiety of the CID with the ligand binding domain (LBD) of a LBD-DBD fusion protein. The latter is screened against a collection of chimeric candidate target proteins (referred to as 'preys'; Y), encoded by a cDNA library fused to the AD. Positive interactions reconstitute a functional transcription-activating trimeric complex, which upon recruitment to specific DNA binding sites in the promoter region of a reporter construct results in the activation of the downstream reporter gene, enabling for example yeast growth on a selective medium (Figure I.2B). Finally, sequencing of the prey fusion gene present in these positive clones completes the target identification process.

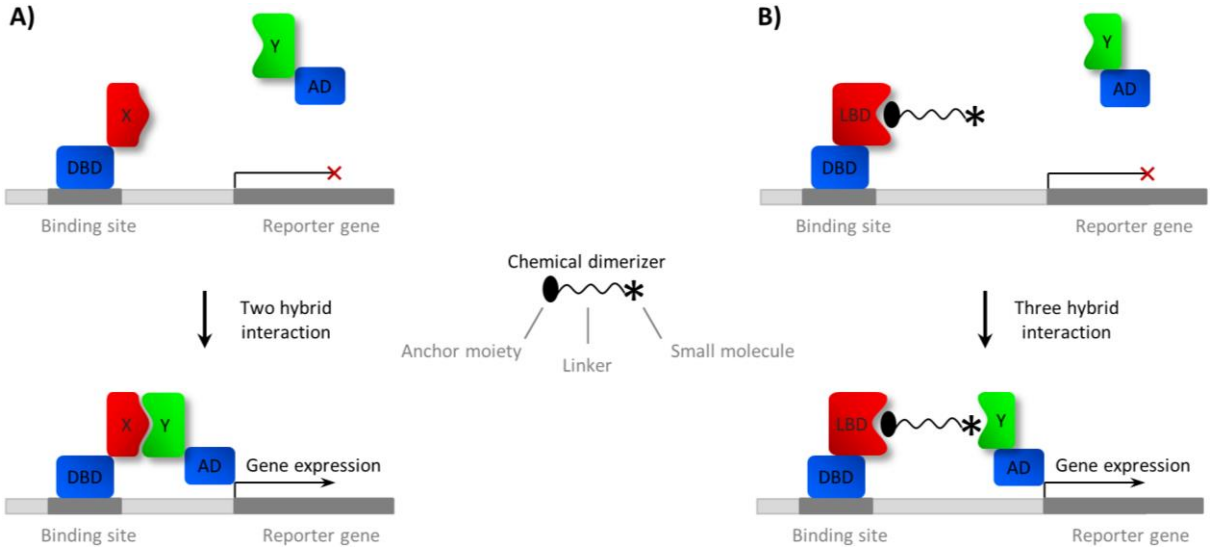


Figure I.2: Schematic overview of the yeast two-hybrid (Y2H) and three-hybrid (Y3H) system. A) Y2H: A chimeric protein (DBD-X) consisting of a DNA binding domain (DBD) fused to a protein X is coexpressed with a second protein chimera (AD-Y) comprising a transcriptional activation domain (AD) fused to a protein Y. If protein X and Y bind to one another, a functional transcription factor is reconstituted, leading to transcription of a downstream reporter gene. **B) Y3H:** Successful reconstitution of a functional trimeric complex through interaction of three hybrid components implies small molecule-target protein binding and results in the expression of a reporter gene (see text for more details). These hybrids include a hybrid protein that consists of a DBD fused to a ligand binding domain (LBD, e.g. dihydrofolate reductase), a CID comprising an anchor moiety (e.g. methotrexate) linked to a small molecule of interest and a second protein chimera composed of an AD fused to a library of candidate target proteins (Y).

The initial report describing the Y3H system was published by Licitra and Liu nearly two decades ago.[15] In this pioneering work, they introduced the first chemical heterodimerizer, comprising a dexamethasone-FK506 hybrid ligand (**1**, Figure I.3), for effectively bridging two fusion proteins: the hormone-binding domain of the rat glucocorticoid receptor (rGR) fused to the LexA DNA-binding domain and FKBP12 fused to the B42 transcriptional activation domain. Moreover, in a proof-of-principle experiment using this heterodimerizer they demonstrated the potential of the Y3H system to uncover the target proteins of small organic compounds by identifying FKBP12 as the interaction partner of FK506 from a Jurkat cDNA library fused to the B42 activation domain. The application of Y3H as a screening tool for novel compound-binding proteins was later confirmed by identifying dihydrofolate reductase (DHFR) clones from a mouse cDNA library screen using the dexamethasone-methotrexate (Dex-MTX) heterodimer as a test compound in a GAL4-AD-based system.[32] Hence, a major focus in the Y3H field has been the development of new ligand-receptor CID pairs, as systematically outlined in Table I.1.

Table I.1: Overview of chemical heterodimerizers used in yeast three-hybrid (Y3H) systems.

Dimerizer	Description	DBD-fusion	AD-fusion	Comments	Ref.
Dex-FK506	Dexamethasone and FK506	rGR	FKBP12	Used in Jurkat cDNA library screen	[15]
Dex-MTX	Dexamethasone and methotrexate	rGR	mDHFR	Used in mouse cDNA library screen	[32]
FK506 Rapamycin	Natural product Natural product	FKBP12 FKBP12	CNA FRB	Using synthetic ligands as a transcriptional on-off switch	[33]
C40-PhOPh-FK506	C40-*p*-phenoxyphenyl bumped FK506 analogue	Mutant mCAB library	FKBP12	Screening for restored binding of compensatory mutant mCABs to a calcineurin-resistant FK506 derivative	[34]
Rapamycin	Natural product	FKBP12	FRB	Towards large-scale screening of protein-ligand interactions in nanodroplets	[35]
Estradiol-biotin	7-α-substituted β-estradiol and biotin	ER-α/β	Tetrameric SA (Y43A mutant)	Enhanced potency and efficacy compared to Dex-biotin	[36]
Estrone oxime-biotin	Estrone-17-(*O*-carboxymethyl)oxime and biotin	ER-α/β	Tetrameric SA (Y43A mutant)	More readily synthesized than estradiol congeners	[37]
MTX-Dex	Methotrexate and dexamethasone	eDHFR/ mDHFR	rGR	Most commonly used anchor moiety	[38],[39],[40], [41],[42]
MTX-SLF	Methotrexate and synthetic ligand for FKBP12	eDHFR	FKBP12 mutants	K_D of ligand-receptor interaction correlates with transcription read-out	[43]
MTX-PurB[1]	Methotrexate and purvalanol B	eDHFR	CDK	First successful identification of novel (CDK-like/unrelated) kinase targets	[44],[45]
MTX-PurB[2]	Methotrexate and purvalanol B	C_{ub}-eDHFR	N_{ub}-PCTK3	Split-ubiquitin-based Y3H system for small-molecule-protein interaction analysis	[46]
MTX-RGB-286147	Methotrexate and a pyrazolopyrimidinone kinase inhibitor	eDHFR	CDK/CRK	Application to MOA studies	[47],[48]
MTX-AA	Methotrexate and anecortave acetate	eDHFR	PDE6D	Identification of new target of drug with unknown MOA	[49]
MTX-Cmpd2	Methotrexate and an imidazopyridine PKG inhibitor	eDHFR	TgBRADIN/ GRA24	Application to MOA studies of antiparasitic drug	[50],[51]
TMP-Dex	Trimethoprim and dexamethasone	eDHFR	rGR	No cross-reactivity with endogenous DHFR	[52],[53]
TMP-SLF	Trimethoprim and synthetic ligand for FKBP12	eDHFR	FKBP12	No endogenous (mammalian) targets	[54]
BG-MTX	O^6-Benzylguanine and methotrexate	hAGT	eDHFR	Selective covalently immobilizing anchor moiety (SNAP-tag)	[55],[56]
BG-Sulfa[3]	O^6-Benzylguanine and sulfasalazine	hAGT	SPR	Identification of known and novel targets of clinically approved drugs	[57],[58]
BG-VC490004[4]	O^6-Benzylguanine and an anti-TB drug lead	hAGT	CoxF	Target deconvolution and validation of anti-TB drugs	[59]

Abbreviations used: AD, transcriptional activation domain; CDK, cyclin-dependent kinase; CID, chemical inducer of dimerization; CNA, calcineurin A subunit; CoxF, carbon monoxide dehydrogenase F protein; CRK, CDK-related kinase; C_{ub}/N_{ub}, C/N-terminal half of ubiquitin; DBD, DNA binding domain; eDHFR/mDHFR, *Escherichia coli*/mammalian dihydrofolate reductase; ER, estrogen receptor; FKBP12, FK506 binding protein with a mass of 12 kDa; FRB, FKBP12-rapamycin-binding domain of FRAP (FKBP12-rapamycin-associated protein); hAGT, human O^6-alkylguanine-DNA alkyltransferase; mCAB, modular calcineurin A/B fusion protein domain; MOA, mechanism of action; PCTK3, PCTAIRE protein kinase 3; PDE6D, phosphodiesterase 6-delta; PKG, cGMP-dependent protein kinase; rGR, rat glucocorticoid receptor; SA, streptavidin; SPR, sepiapterin reductase; TB, tuberculosis; TgBRADIN, *Toxoplasma gondii* bradyzoite differentiation inhibitor.
[1] As well as (R)-roscovitine and indenopyrazole-1.
[2] As well as dexamethasone.
[3] As well as methotrexate, dasatinib, purvalanol B, erlotinib, atorvastatin, furosemide and indomethacin.
[4] As well as methotrexate, fluphenazine, mefloquine, ofloxacin, clofazimine and rifampicin.

Initially, the Schreiber and Crabtree laboratories exploited the use of the natural heterodimerizers FK506 and rapamycin to control transcription of integrated genes in yeast.[33] In an effort to circumvent the limitations of these dimerizers related to interference with cellular functions of their endogenous binding partners, novel, orthogonal dimerizer systems were engineered following a 'bump-hole' strategy.[60] In this approach, an artificial cavity is introduced into the ligand-binding site of the protein to accommodate an additional, bulky substituent installed on the ligand and which inhibits binding to the wild-type protein. In the case of FK506, a Y3H screen was applied as a chemical genetic selection tool to identify compensatory mutant modular calcineurin A/B fusion protein domains (mCABs) that restored binding to a calcineurin-resistant bumped analogue.[34] Over the last two decades, these natural dimerizers and their synthetic derivatives found widespread use as powerful research tools in an investigational[29],[28] and therapeutic[61],[62],[18] context, but have only been scarcely applied in small molecule-target protein profiling research. An interesting yeast-based example is the concept of detecting protein-ligand interactions in stochastic nanodroplets on a large scale.[35] Here, rapamycin is coupled to TentaGel resin beads via a photocleavable linker and its time-dependent photochemical release results in the dimerization of FKBP12-DBD and FRB-AD fusion proteins, which triggers transcription of a *HIS3* reporter gene, allowing selective yeast growth in nanodroplets lacking histidine. In an extension of this initial design, the authors postulate the potential of the system to screen an anchored combinatorial library, consisting of large numbers of combinatorial small molecules covalently coupled to a constant anchor moiety, against a cDNA library of candidate target proteins in a single experiment. However, since this first proof of concept, to our knowledge no follow-up studies on this topic have been reported.

Relying on the rather weak binding of dexamethasone to yeast-expressed GR proteins[63] and its relatively low activity in recombinant yeast systems,[64] Peterson and co-workers investigated steroidal estrogens as alternative immobilizing anchor moieties for profiling natural products in Y3H.[36] They synthesized both a hybrid 7-α-substituted β-estradiol derivative fused to biotin, providing a simple model of more complex natural products, and its dexamethasone congener. To directly compare dimerizer-mediated activation of gene expression, yeast were engineered to co-express tetrameric streptavidin (SA)-AD and estrogen receptor (ER) α/β- or GR-DBD fusion proteins, respectively. Markedly, this new β-estradiol-based CID showed enhanced potency and efficacy compared to the previous Dex-GR CID pair. Though, the routine use of these 7-α-substituted β-estradiol derivatives as CIDs is hampered by their elaborate 14-step synthesis.[65],[36] As a dramatically streamlined alternative, a facile two-pot synthesis starting from estrone afforded in good overall yield a novel, cell-permeable CID comprising estrone-17-(*O*-carboxymethyl)oxime linked to biotin,

which displayed a dose-dependent activation of reporter gene expression in the ER-SA Y3H assay comparable to its estradiol congener.[37]

A significant evolution in the Y3H field was the introduction of the well-established methotrexate (MTX)-dihydrofolate reductase (DHFR) CID pair by the Cornish group,[38] which represents the most commonly used anchor system to date. In an effort to characterize three-hybrid systems at the biochemical level, they examined the influence of the ligand-receptor pair on the transcriptional read-out.[39] The latter was found to be much more sensitive to the structure of the fusion proteins than to variations in the MTX-Dex dimerizer design. Most surprisingly, using the bacterial *E. coli* DHFR (eDHFR), consistently higher transcription activation levels were obtained compared to its mammalian homologue (mDHFR), despite the fact that both exhibit similar binding affinities for MTX (K_i = 1.0-3.4 pM).[66] Subsequently, this initial plasmid-based system was optimized to stabilize the transcriptional read-out in a manner that allows the straightforward variation of the AD chimera.[40] Therefore, the genes encoding the DHFR anchor protein and the complementary *LEU2* and *lacZ* reporters were integrated into the yeast chromosome. Given its remarkably reduced percentage of false negatives (from 20% to 3%) and a highly consistent reporter read-out, Cornish and colleagues assumed that this MTX-based integrated system would be suitable for the *de novo* identification of small molecule-target proteins in AD-fused prey cDNA library screens.

In fact, in a follow-up paper they showed that the MTX-DHFR Y3H system has the requisite sensitivity for drug discovery.[43] For this, a systematic study was undertaken to quantitatively characterize the correlation between ligand-receptor affinity and the transcription read-out using a heterodimer of MTX and the synthetic ligand for FKBP12 (SLF), which was previously described in the context of a bacterial RNA polymerase small molecule three-hybrid system.[67] As a starting point, de Felipe *et al.* designed a series of FKBP12 point mutants spanning several orders of magnitude in their K_D for SLF, as measured in vitro using a fluorescence polarization assay. Subsequently, the levels of transcriptional activation for these variants were determined in the Y3H assay. Interestingly, they observed that the strength of the SLF-FKBP12 interaction does correlate with the transcriptional read-out, albeit over a small range of K_D's (1 order of magnitude on log scale). Furthermore, the authors concluded that in this MTX-based Y3H assay it is not possible to detect interactions whose K_D's are over ca. 50 nM, a cutoff considered high enough to realistically use the Y3H system for drug discovery. A question that may be asked, however, is whether such a system would indeed be sufficiently sensitive for detecting medium/low-affinity interactions of different types of small molecules of interest with off-target proteins that contribute to unwanted side-effects or toxicity.

In 2004 this major outstanding question was addressed by Kley and co-workers in a key publication reporting the first successful application of the Y3H assay to discover novel drug targets.[44],[45] More specifically, the target spectrum of cyclin-dependent kinase (CDK) inhibitors, including purvalanol B, was determined in a MTX-eDHFR-based system using both cDNA library and focused screening of yeast cell arrays displaying selected polypeptide open reading frames (ORFs). The target profile of the latter purine analogue proved to be significantly broader than originally anticipated, since in addition to a range of known targets (CDK1, CDK4, and CK1, among others) several novel candidate binders, such as CDC/CDK-like kinases, other types of serine/threonine kinases and even receptor and non-receptor tyrosine kinases were identified. In vitro, these enzymes are inhibited by purvalanol B with IC_{50} values in the higher nanomolar to low micromolar range, defining a sensitivity

limit of this Y3H system comparable to that of yeast two-hybrid[68] and significantly broadening the range of detectable interactions in the context of target deconvolution and drug discovery. As such, this work represents a significant step forward by enlarging the scope of Y3H towards profiling of small-molecule kinase inhibitors, one of the main components of drug development pipelines to date.

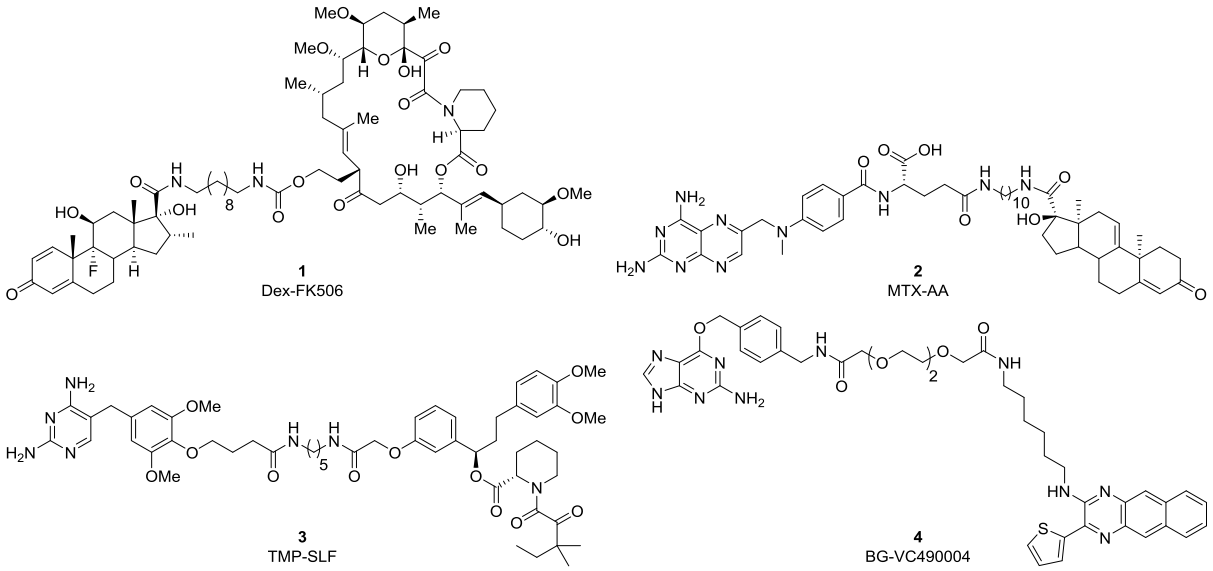


Figure I.3: Chemical structures of selected dimerizers used in yeast three-hybrid systems. Dexamethasone-FK506 (**1**); methotrexate-anecortave (**2**); trimethoprim-synthetic ligand for FKBP12 (**3**); and O^6-benzylguanine-VC490004 (**4**).

Another interesting application of Y3H are mechanism of action studies, which are essential for linking phenotypic responses with molecular targets of small molecules. This was first successfully demonstrated in 2005 by Caligiuri *et al.*, exploring the potential applications of compounds derived from a trisubstituted pyrazolopyrimidinone kinase inhibitor scaffold (such as the putative CDK inhibitor RGB-286147).[47],[48] Recently, Shepard *et al.* elucidated the MOA of anecortave acetate (AA), an intraocular pressure-lowering cortisene for treating glaucoma.[49] By displaying the latter in a MTX-anchored Y3H system (**2**, Figure I.3), they identified phosphodiesterase 6-delta (PDE6D) as a new molecular target and validated it all the way from quantitative *in vitro* measurements to an animal model of the disease.

So as to overcome the partial limitations of MTX-based CIDs related to their cross-reactivity with endogenous yeast DHFR and the ensuing potentially impaired transcriptional activation in Y3H, Cornish and colleagues introduced trimethoprim (TMP) as an alternative prokaryote-specific[69] DHFR anchor moiety.[52] This new CID pair (TMP-eDHFR) was originally established as an *in vivo* chemical protein labeling method, i.e. TMP-tag,[70] which recently has been rendered both covalent and fluorogenic for high-resolution intracellular live cell imaging applications[71] (for thorough reviews on chemical tags, see [72],[73],[74]). In the context of protein dimerization studies, TMP-based CIDs proved particularly useful in mammalian systems, as exemplified by the biocompatible TMP-SLF heterodimerizer (**3**, Figure I.3)[54] and further discussed in chapter 4. The latter CID **3** was successfully employed to activate transcription in an eDHFR-FKBP12 Y3H assay and to modulate a Golgi-resident glycosyltransferase in Chinese hamster ovary (CHO) cells.

While the spectrum of CIDs discussed so far exclusively relies on reversible interactions, some dimerizers have been developed to present the small molecule of interest in an irreversible, covalent fashion based on yet another protein labeling approach, i.e. SNAP-tag.[75] This strategy is centered around the human DNA-repair protein O^6-alkylguanine-DNA alkyltransferase (hAGT). The Johnsson laboratory exploited the low substrate specificity of this enzyme to covalently label SNAP-tag-fused proteins *in vivo* with a ligand of interest by conjugating the latter to the *para* position of O^6-benzylguanine (BG). Extrapolating this method to the CID field, they introduced the first chemical inducer of hemicovalent dimerization comprising a BG-MTX bifunctional molecule.[55] In a proof-of-principle hAGT-eDHFR Y3H system, this dimerizer induced transcriptional activation at a level comparable to that achieved with noncovalent CIDs.[55] By optimizing this initial concept, the Johnsson group established a powerful yeast-based platform that enabled the identification of both known and novel interaction partners of clinically approved drugs covering a wide range of therapeutic areas.[57] The methodology couples the identification of a candidate target to its subsequent unambiguous validation by combining a new, more sensitive SNAP-tag-based Y3H system with an effective pulldown or activity assay that utilizes the same BG drug derivative as in the Y3H screens. In this respect, they uncovered, for example, the first non-kinase target of erlotinib (ORP7) and characterized several off-targets of atorvastatin (such as PDE6D and NQO2). Most intriguingly, this milestone study revealed that the anti-inflammatory drug sulfasalazine and its metabolites inhibit sepiapterin reductase (SPR) and consequently decrease tetrahydrobiopterin (BH4) biosynthesis, thereby suggesting new and improved therapeutic applications for this long-standing drug. Finally, this SNAP-tag-based Y3H platform was recently adapted for target deconvolution of anti-tuberculosis drugs by screening cDNA libraries encoding mycobacterial proteins with BG-derivatized small molecules, such as the synthetic anti-TB drug lead VC490004 (**4**, Figure I.3).[59] A comparable approach was followed by Odell *et al.* to identify *T. gondii* parasite proteins that interact with a MTX-functionalized imidazopyridine PKG inhibitor ('Compound 2').[50] Of note, both TMP- and SNAP-tag-based Y3H assays are being applied as compound profiling tools as part of a commercial drug discovery service.[76]

I.4. Mammalian three-hybrid systems

Analogous to the yeast three-hybrid systems described above, several classes of chemical dimerizers have been applied in a mammalian (human) context, as listed in Table I.2. The interest in developing these mammalian systems was instigated by a number of limitations of Y3H, including its restricted membrane permeability, primitive cellular context, and incapacity to monitor interactions outside of the nucleus. Starting in 1996, FK1012 homodimerizers were shown to activate gene transcription in Jurkat cells[33],[77] and human skin keratinocytes and fibroblasts[78]. In a next stage, wholly synthetic dimeric SLFs (such as AP1510) were developed by Ariad Gene Therapeutics, which combine a reduced complexity and greater adaptability with higher potency compared to the first generation natural product-derived CIDs.[79],[80],[81] Finally, in order to potentially reduce nonproductive interactions and further improve performance, a series of novel, homodimeric bumped AP1510 analogues (e.g. AP1889) was designed to bind specifically to the F36V point mutant and only minimally to endogenous FKBP12.[82],[83]

Table I.2: Overview of chemical dimerizers used in mammalian three-hybrid (M3H) systems.

Dimerizer	Description	DBD/Anchor protein[1]-fusion	AD/Prey[1]-fusion	Comments	Ref.
Homodimerizers					
FK1012	Dimeric FK506	FKBP12 (3 copies)	FKBP12 (3 copies)	Using synthetic ligands as a transcriptional on-off switch	[33],[77], [78]
AP1510 and congeners	Dimeric synthetic FKBP12 ligands	FKBP12 (3 copies)	FKBP12 (3 copies)	Reduced complexity, greater adaptability and more potent than natural product CIDs	[79],[80], [81],[84]
AP1889 and congeners	Bumped analogues of AP1510 series	FKBP12 F36V mutant (3 copies)	FKBP12 F36V mutant (3 copies)	No affinity for endogenous FKBP12	[82],[83]
Heterodimerizers					
FKCsA	FK506 and cyclosporin A	FKBP12 (3 copies)	CyP (2 copies)	Inducible transcriptional activation in Jurkat cells	[85]
AP1867-THOXs	Synthetic FKBP12 ligand and substituted tetrahydrooxazepines	FKBP12 F36V mutant	?	Diversified library of cell-permeable CIDs to identify novel ligand-receptor pairs	[86]
C40-PhOPh-FK506[2]	C40-*p*-phenoxyphenyl bumped FK506 analogue	FKBP12 (3 copies)	mCAB VKMGC mutant	No affinity for endogenous calcineurin	[34]
Rapamycin	Natural product	FKBP12 (3 copies)	FRB	Pharmacological control of gene expression in human cells	[87],[88], [89]
C16-(R)-methallyl[3] and other rapalogues	C16-(R)-methallyl and other bumped rapamycin analogues	FKBP12 (3 copies)	FRB LFP mutant	No affinity for endogenous FRB	[90],[62]
Dex-Fum	Dexamethasone and fumagillin	rGR	PfMetAP2	SAR evaluation of fumagillin analogues via competition assay	[91],[92]
MTX-PD173955[4]	Methotrexate and a potent ABL kinase inhibitor	eDHFR	ABL	MASPIT cDNA library screening and competition assays for small molecule-target protein profiling in intact human cells	[16]

Abbreviations used: ABL, Abelson tyrosine kinase; AD, transcriptional activation domain; CID, chemical inducer of dimerization; CyP, cyclophilin; DBD, DNA binding domain; eDHFR, *Escherichia coli* dihydrofolate reductase; FKBP12, FK506 binding protein with a mass of 12 kDa; FRB, FKBP12-rapamycin-binding domain of FRAP (FKBP12-rapamycin-associated protein); MASPIT, mammalian small molecule-protein interaction trap; mCAB, modular calcineurin A/B fusion protein domain; PfMetAP2, *Plasmodium falciparum* methionine aminopeptidase type 2; rGR, rat glucocorticoid receptor.

[1] In MASPIT, anchor protein (X) and prey (Y) denote the cytokine receptor (CR)-X and Y-gp130 fusion proteins, respectively (see Figure I.5 for details).

[2] As well as C40-phenyl-FK506.

[3] As well as C16-(R)-isopropoxy rapamycin.

[4] As well as AP1867, E7070, RGB-285978, RGB-285961 and RGB-286147.

The first synthetic heterodimerizer implemented in a mammalian three-hybrid setting was described by Crabtree and Schreiber in 1996, and consists of a fusion of the natural products FK506 and cyclosporin A (FKCsA **5**, Figure I.4).[85] By dimerizing FKBP12-DBD and cyclophilin (CyP)-AD chimeras, the hybrid ligand strongly stimulated the expression of a GAL4-responsive secreted alkaline phosphatase (SEAP) reporter gene in Jurkat cells. As for the Y3H system, initial attention was also

paid to the well-established natural heterodimerizers FK506, rapamycin and derivatives thereof. Having identified a set of selected mutant mCAB hits that restored binding to C40-*p*-phenoxyphenyl FK506 using a Y3H system (see Section I.3), the Schreiber group applied a reciprocally configured M3H assay in a subsequent validation step.[34] Toward that end, these hits were fused to an AD domain and evaluated for their ability to restore binding to the bumped FK506 analogue, displayed by FKBP12-DBD chimeras. As a result, the VKMGC mutant was shown to most optimally respond to the calcineurin-resistant derivative. Previously, the group had followed a similar M3H-based screening approach for the discovery of novel ligand-receptor CID pairs based on rapamycin.[90] As mentioned in Section I.3, next to their application as tool compounds for fundamental biology studies, rapamycin[87] and its bumped analogues[62] paved the way for the development of suitable systems for pharmacological control of therapeutic gene expression.

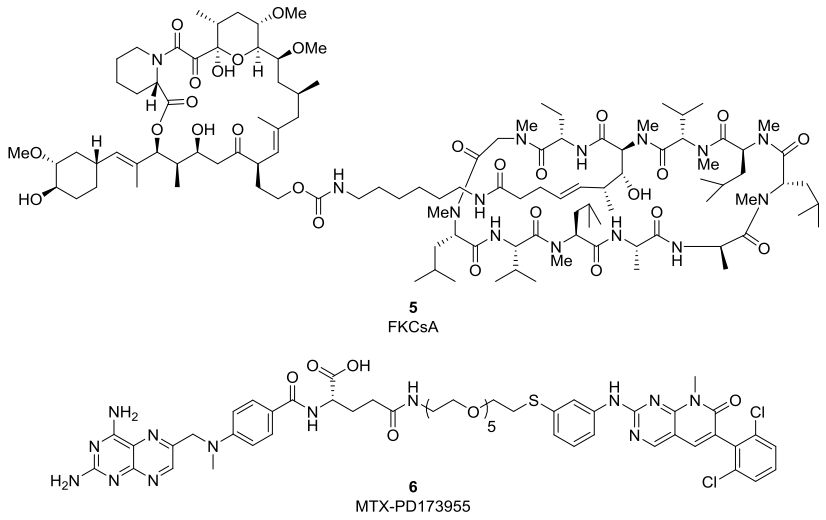


Figure I.4: Chemical structures of selected dimerizers used in mammalian three-hybrid systems. FK506-cyclosporin A (**5**); and methotrexate-PD173955 (**6**).

In another study aiming at the identification of new ligand-receptor CID pairs, Koide *et al.* constructed a diversified library of heterodimerizers comprising an invariant synthetic FKBP12 ligand (i.e. AP1867) attached to 320 substituted tetrahydrooxazepines (THOXs).[86] The latter system was selected as 'privileged scaffold' for its structural novelty in the cyclic form and convertibility to a linear form through reduction of the N-O bond. Furthermore, the presence of the olefin and three chiral centers creates a further functionalization site and the possibility to exploit stereochemistry as a diversity element, respectively. Subsequently, the cell permeability of these candidate heterodimerizers was assessed by analyzing a representative panel of 25 library members spanning a broad range of hydrophobicities using an AP1889-based M3H competition assay. In this experiment, a cell-permeable heterodimerizer competes with the AP1889 homodimerizer for binding to FKBP12 F36V chimeras, thereby disrupting the complex that activates SEAP expression. All of the tested compounds as well as an additional sample of six of the most polar library members were found to be freely permeable to human fibrosarcoma cells, as each gave strong inhibition of SEAP expression levels. Unfortunately, the authors did not report on the identification of any functional, novel CID pairs and only briefly referred to several initial promising 'hits'. However, to our knowledge, no follow-up studies have been published ever since. In addition to the evaluation of cell permeability properties, M3H competition experiments with native unfused small molecules also allow for their

prioritization with respect to *in cell* target binding affinity. Accordingly, structure-activity relationships (SAR) can be mapped out based on the side-by-side comparison of analogue series. An interesting example of this strategy is provided by Chen *et al.*, who performed dexamethasone-fumagillin-mediated M3H assays in the absence and presence of a series of potential competitors of this antimalarial lead.[92],[91]

The possibility to perform such competition assays in a host system that is readily permeable to a wide variety of small molecules, offers an important advantage of mammalian over yeast three-hybrid systems. Namely, wild-type yeast are naturally resistant to many drugs, whose intracellular accumulation is hampered by constitutively active multidrug-resistance (MDR) mechanisms.[93] In fact, a complex network of genes, termed the pleiotropic drug resistance (PDR) network, contributes to this MDR phenotype. For example, three such genes, i.e. *PDR5*, *SNQ2*, and *YOR1*, encode plasma membrane ATP-binding cassette (ABC) transporters that actively mediate drug efflux, thereby lowering intracellular drug levels.[94],[95] On the other hand, the *ERG6* gene, encoding the C-24 sterol methyltransferase enzyme acting in the late steps of the ergosterol biosynthetic pathway, limits the rate of passive drug diffusion across the fungal membrane.[96] Hence, in an effort to tackle this uptake issue, reporter yeast strains have been genetically modified by specific deletion of the latter genes, resulting in a significantly enhanced permeability to small molecules, as exemplified in both two-hybrid[97] and three-hybrid[57],[98] studies. An alternative and complementary approach to increase the sensitivity of yeast to drugs consists of inducing the hexose transporters *HXT11* and *HXT9*, which have been postulated to influence ABC pump function via negative feedback regulation or to allow the translocation of other classes of compounds across the membrane.[99]

I.4.1. MASPIT and KISS

An inherent limitation to both the mammalian three-hybrid systems discussed so far and Y3H in general is their restriction to the detection of interactions with proteins or protein domains that can be functionally expressed and effectively translocated into the nucleus. MASPIT (mammalian small molecule-protein interaction trap) circumvents this disadvantage by shifting the analysis of binding events between small organic compounds and their target proteins to the cytosol of living human cells (Figure I.5A).[16] Additionally, the intact human cellular context (HEK293T) ensures compound profiling in a physiological relevant environment and conserves normal protein conformation as well as post-translational modifications. Furthermore, this might reveal potential effects of the target's association with additional proteins or other intracellular molecules on small molecule binding.

MASPIT is the three-hybrid component of the MAPPIT (mammalian protein-protein interaction trap) technology platform,[100],[101] an approach that uses the JAK-STAT pathway which mediates cytokine receptor (CR) signaling. MASPIT employs *E. coli* dihydrofolate reductase (eDHFR), fused to a mutated, functionally inactivated, CR to present a methotrexate (MTX) fusion compound (MFC) to the intracellular environment. This allows screening of a small-molecule bait against a collection of chimeric prey proteins. As a result of the interaction between the small-molecule bait and a target prey protein, the JAK-STAT pathway is activated, resulting in the expression of a reporter gene (Figure I.5A).

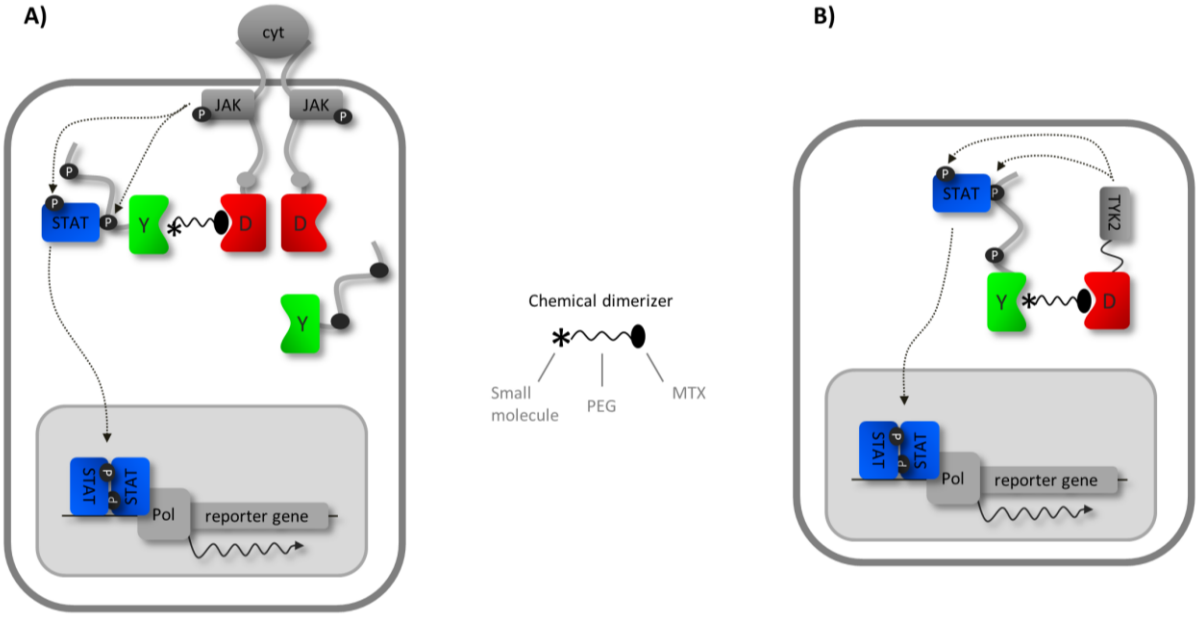


Figure I.5: Outline of the MASPIT and KISS mammalian three-hybrid (M3H) systems. A) MASPIT: *E. coli* dihydrofolate reductase (D) is fused to a cytokine receptor (CR) that is rendered signaling-deficient by mutating STAT3 recruitment sites in its cytoplasmic tail (grey dots). A prey protein (Y) is tethered to a gp130 CR fragment containing functional STAT3 transcription factor docking sites (black dots). When a fusion compound consisting of a small molecule of interest (asterisk) coupled to methotrexate (MTX) is added to the cells, MTX binds to eDHFR, resulting in the compound of interest being displayed as bait. Upon administration of the appropriate cytokine ligand (cyt), the CR-eDHFR chimeric receptor undergoes a conformational change, activating the associated JAK2 kinases through crossphosphorylation (P). Interaction between the small-molecule bait and the prey-gp130 fusion protein brings the latter into proximity of the activated JAK kinases, reconstituting a functional JAK-STAT signaling pathway. Sequential phosphorylation of STAT3 docking sites on the gp130 chain (P), STAT3 recruitment, and STAT3 phosphorylation (P) ultimately leads to activated STAT3 dimers that induce the expression of a reporter gene coupled to a STAT3-dependent promoter. **B) KISS:** eDHFR (D) is tethered to the C-terminal (kinase-domain-containing) portion of tyrosine kinase 2 (TYK2) and co-expressed with a prey protein (Y) fused to a gp130 CR fragment. Upon physical interaction between the small molecule and the prey protein, TYK2 phosphorylates STAT3 docking sites on the prey chimera (P), which eventually triggers reporter gene activation as in the case of the MASPIT assay.

The original MASPIT paper was published by Kley and coworkers in 2006, focusing on MTX-based CIDs incorporating diverse small-molecule chemotypes, such as the potent pyridopyrimidine ABL tyrosine kinase inhibitor PD173955 (**6**, Figure I.4).[16] The latter MFC was applied in a MASPIT cDNA library screen aiming at the proteome-wide *de novo* identification of its target proteins. Only kinases were detected in this screen; besides different known kinases (e.g. LYN, FYN and FGFR1), the authors also identified several novel kinase targets of this compound, such as various ephrin receptors and cyclin G-associated kinase. In contrast to classical target-based profiling, this M3H system can thus provide information regarding unanticipated small molecule-target protein interactions. Furthermore, this study also describes for the first time the use of competition assays to compare and rank unmodified small molecules (e.g. PD173955 and imatinib) on the basis of their relative ability to displace a CID (MTX-PD173955) from its target protein (ABL) in intact mammalian cells (analogous to SAR evaluation in [92] as discussed above). Interestingly and unlike *in vitro* K_D or IC_{50} values, such quantitative cellular targeting potencies integrate many variables associated with compound cell-bioavailability and intracellular targeting efficiency. However, an important limitation of the system is its incompatibility with (full size) transmembrane proteins. Indeed, since the small-

molecule bait is recruited to a plasma membrane anchored cytokine receptor chimera, only interactions with soluble target proteins can be detected.

The latter shortcoming was addressed by the introduction of KISS (kinase substrate sensor), a novel binary protein-protein interaction mapping approach that enables *in situ* analysis in living mammalian cells of protein interactions and additionally small molecule-target protein profiling in a three-hybrid setup.[102] In this assay, small molecules are displayed as baits through recruitment of their corresponding MFCs to an *E. coli* DHFR (eDHFR)-tyrosine kinase 2 (TYK2) fusion protein, and interactions with target protein prey fusions are detected through reporter induction as in the case of the MASPIT assay (Figure I.5B). Contrary to MASPIT, the KISS 3H approach allows screening for integral membrane target proteins as this interaction sensor is freely diffusible.

I.5. Design aspects of chemical dimerizers

Ideally a CID for three-hybrid applications should be accessible via a modular, versatile and practical synthesis in a minimal number of steps and acceptable overall yield. Furthermore, the CID should be readily soluble and cell permeable. In the following three subsections examples of anchor moieties, linkers and conjugation methodologies are discussed (Figure I.6).

I.5.1. Anchor moieties

For a comprehensive outline of the ligand-receptor pairs used in three-hybrid systems to date, we refer to Table I.1 and Table I.2. Representative examples of prominent anchor moieties based upon reversible or irreversible immobilization are included below (Figure I.6).

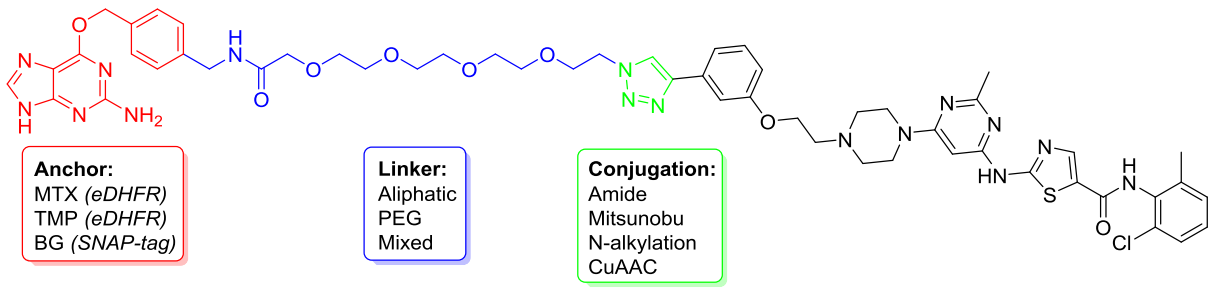


Figure I.6: Typical components of chemical dimerizers for three-hybrid systems. Schematic non-comprehensive overview of anchor moieties, linkers and conjugation methods applicable to the synthesis of chemical dimerizers as exemplified by the O^6-benzylguanine-$(PEG)_4$-dasatinib[57] conjugate. Anchor proteins are in brackets.

I.5.2. Linkers

To discourage steric hindrance of fusion partners, which could cause the CID to bind suboptimally to its anchor protein or allow one to overlook targets that might interact with the unconjugated small-molecule bait, a linker is introduced to allow optimal interaction with prey chimeras. Toward this end, the attachment position of the spacer, both on the anchor moiety and bait, must be meticulously chosen so as to minimize loss in binding affinity through perturbation of

pharmacophores or induction of conformational changes. In this respect, 'linkerology' studies are often guided by available SAR, structural, and molecular modeling data, which facilitate selection of the attachment site and incorporation of appropriate chemical handles. Indeed, in the case of DHFR ligands attachment positions are well established. For example, SAR studies of MTX derivatives have emphasized the importance of selective conjugation to the γ-carboxylic acid of the glutamate moiety to ensure high affinity binding to eDHFR through interaction of this enzyme with the free α-carboxylic acid.[103],[104] However, appropriate derivatization of small-molecule baits is often less obvious and is generally initiated from SAR information of their primary targets (as far as known). Of note, for comprehensive target deconvolution studies one should ideally rely on phenotypical SAR data, acquired from testing a panel of derivatives modified at different positions for retained activity in appropriate functional assays.[57] Two main factors that govern the design of a particular linker are its length and chemical composition.

Though crucial, no hard and fast rules are available for identifying the optimal linker length, which is typically determined empirically. For instance, one can envision that sufficient length and flexibility would be necessary to effectively bridge both binding sites and successfully heterodimerize the two fusion proteins, while on the other hand the linker should be short enough to ensure the desired proximity-related dimerization event to occur. Hence, this supports the development of modular synthesis strategies, enabling linker length and conjugation method to be swiftly mixed and matched as required (Figure I.6).[51] Previously, the effect of CID linker length modifications on the transcription read-out in both yeast and mammalian three-hybrid systems has been investigated, yet yielding contradictory outcomes. Piloting work by Cornish and colleagues suggested an increase in Y3H sensitivity by increasing the linker length of MTX-Dex CIDs.[39] However, only subtle differences among five-, eight-, and ten-methylene linkers were revealed. Yet, the latter trend was later confirmed in several other yeast-based studies,[36],[44],[46] but contrasts sharply with the reports by Amara[79] and Schreiber[77], observing a clear inverse correlation between linker length and activity. These discrepancies might be explained by the relative strength of the interactions, as well as geometrical and steric effects governed by the nature of the prey chimeras and hence frequently underscore the need for application-specific CID optimizations.

Three main classes of linker types have been applied for the construction of CIDs in the context of 3H systems, i.e. straight-chain aliphatic (methylene),[39] poly(ethylene glycol) (PEG),[44],[46],[51] and mixed-type spacers (e.g., *N*,*N*'-(ethane-1,2-diyl)bis(2-hydroxyacetamide)-derived).[81],[79] The chemical composition of the linker might significantly influence the stability, solubility, cell permeability and hence overall efficiency of CIDs. Clearly, one should keep away from reactive functional groups, such as esters, which could be cleaved *in vivo*. Johnsson and co-workers have shown that a BG-based CID comprising an aliphatic linker proved to be more effective than its PEG congener, most probably due to the fact that the former spacer relatively increases the overall hydrophobicity of the CID, favoring its cellular uptake.[55] However, this frequently observed enhanced permeability of methylene spacer-based CIDs is often at the expense of their solubility and too hydrophobic linkers could promote undesired aggregation in aqueous media.[20] PEG linkers on the other hand were reported to yield dimerizers with generally good water solubility properties and, remarkably, ditto yeast cell permeability.[44],[23]

I.5.3. Conjugation methodologies

Various (bio)conjugation methods are available for fusing both the anchor moiety and small-molecule bait to either end of the linker. Most synthetic strategies apply general anchor-spacer reagents (e.g. MTX-$(PEG)_5$-NH_2 or BG-$(PEG)_4$-NH_2)[44],[57] for subsequent ligation to (appropriately functionalized) small molecules using standard peptide coupling, Mitsunobu or N-alkylation reactions, thereby yielding amide, ether or amine linkages, respectively.[23] Remarkably, only by 2011, the highly powerful, regioselective and reliable copper(I)-catalyzed azide alkyne cycloaddition 'click' reaction (CuAAC)[105] found acceptance in the Y3H field.[57] Two years later, Tran *et al.* reported on newly developed synthetic strategies, implementing 'clickable' MTX reagents for the straightforward generation of CIDs for Y3H applications.[51]

I.6. Alternative three-hybrid systems

I.6.1. Alternative yeast three-hybrid systems

A number of variations on the classical Y3H theme, as depicted in Figure I.2, have appeared in the course of the past decade. One of the most notable adaptations is the chemical complementation assay, introduced by the Cornish group (Figure I.7).[106] The approach entails a reaction-independent, genetic high-throughput assay, which uses a MTX-Dex Y3H system to link enzyme catalysis of bond formation or cleavage reactions to transcription of an essential or repression of a toxic downstream reporter gene, respectively. In the bond-making direction, enzyme-mediated covalent coupling of a MTX-disaccharide-fluoride donor and a Dex-disaccharide acceptor reconstitutes a functional heterodimerizer capable of increasing transcription of a *LEU2* reporter gene, which allows for a growth selection in the absence of leucine (Figure I.7A).[107] Hence, in this forward direction, the system has been mainly applied as a platform for the directed evolution of glycosynthases.[108],[109]

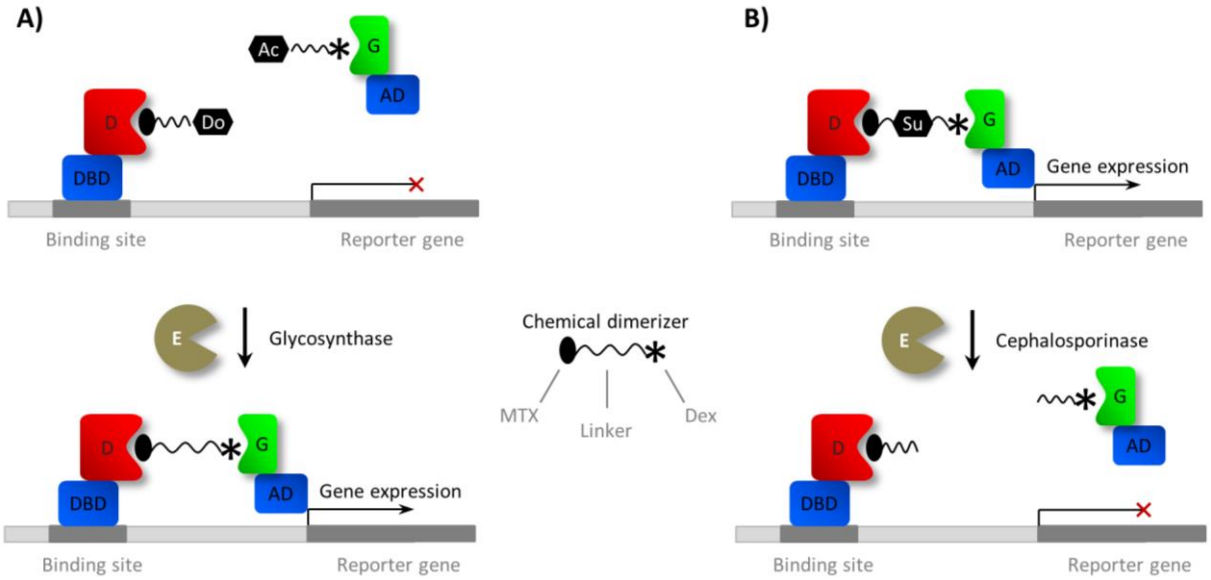


Figure I.7: Schematic overview of the chemical complementation assay. A) Bond formation: Glycosynthase-catalyzed glycosidic bond formation between a MTX glycosyl donor (Do) and a dexamethasone glycosyl acceptor (Ac) is detected by reconstitution of the MTX-Dex heterodimerizer and activation of reporter gene expression. **B) Bond cleavage:** Addition of cephalosporinase to the Y3H system results in cleavage of the MTX-Cephem-Dex substrate and disruption of reporter gene transcription. D, eDHFR; E, enzyme; G, glucocorticoid receptor; Su, substrate.

On the other hand, to monitor enzyme-catalyzed bond cleavage reactions, specialized CIDs comprising the substrate for the reaction under study as alternative linker between the anchor moiety and small-molecule bait (i.e., MTX-substrate-Dex) are constructed (Figure I.7B). Addition of an active enzyme as a fourth component to the Y3H system, results in cleavage of the dimerizer and subsequent disruption of reporter gene transcription.[106],[110] In combination with combinatorial mutagenesis, this assay variant was used to investigate the mechanism of resistance to the third-generation cephalosporin cefotaxime by the class C β-lactamase P99.[111] Alternatively, by switching to the toxic *URA3* reporter gene, the detection of enzymatic activity was coupled to yeast cell survival, enabling the first high-throughput selection for cellulase catalysts.[112] Taken together, the generality of chemical complementation represents a fundamental asset of this approach and should allow its extrapolation to detect new enzymatic activities for important chemical transformations.[112]

Two other modified versions of the traditional Y3H system were described by Peterson and co-workers, which are essentially built around the phosphorylation of a universal protein tyrosine kinase (PTK) substrate[113] and biotinylation of a BirA biotin protein ligase substrate,[114] respectively. The former system was initially designed for the discovery and characterization of novel PTK enzymes.[113] In a successful validation experiment, v-Abl or v-Src kinase mediated phosphorylation of a tetrameric universal tyrosine-containing substrate linked to green fluorescent protein (GFP) and fused to the LexA DBD enabled recognition by the coexpressed phosphotyrosine-binding Grb2 SH2 domain-B42 AD chimera, ultimately activating the expression of the reporter gene. In a follow-up report, the authors described a permeability-enhanced variant of this system, allowing analysis of PTK inhibitors and suggesting its application to the discovery of drug leads in high-throughput screening assays.[98] The second alternative and related Y3H approach involves highly specific BirA-facilitated covalent acylation of the key lysine-10 residue of the Avitag-GFP-B42 AD substrate by exogenous biotin.[114]

The resulting biotinylated Avitag fusion protein subsequently tightly binds the SA-LexA DBD, leading to reporter gene expression. Afterwards, Weber *et al.* exploited this specific use of biotin as a covalent heterodimerizer to control transgene expression in mammalian cells.[115],[116]

A last alternative three-hybrid technology is based on the split-ubiquitin system, an *in situ* protein interaction sensor.[117] In the three-hybrid setup, a CID mediates the proximity of the split-ubiquitin sensor via interaction with its two receptors, each of which fused to the N- (N_{ub}) or C-terminal half of ubiquitin (C_{ub}).[46] Subsequent reconstitution of quasi-native ubiquitin results in proteolytic cleavage by ubiquitin-specific proteases (UBPs) and ensuing degradation of the destabilized R-Ura3p reporter (Figure I.8). Successful proof-of-concept studies were conducted with model MTX-Dex dimerizers, comprising different PEG linker lengths (see Section I.5.2) and a comparison with their effectiveness in the classical Y3H assay was performed. The application of this alternative Y3H system was further extended to a pharmaceutically relevant context upon monitoring the interaction of purvalanol B with PCTAIRE protein kinase 3 (PCTK3).

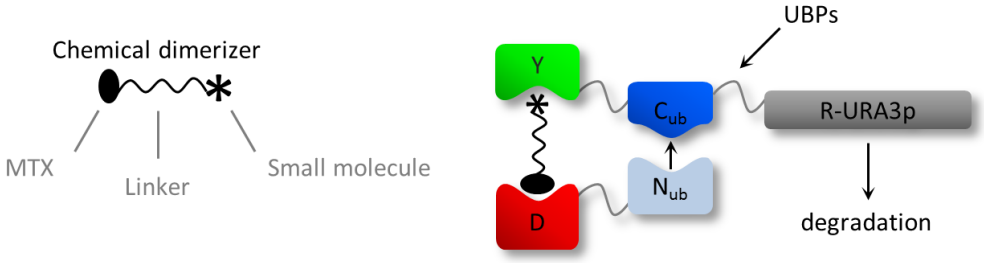


Figure I.8: Schematic representation of the split-ubiquitin-based yeast three-hybrid system. A successful small molecule-target protein interaction mediates reconstruction of ubiquitin, resulting in cleavage by ubiquitin-specific proteases (UBPs) and degradation of the R-Ura3p reporter. D, eDHFR; N_{ub}/C_{ub}, N/C-terminal half of ubiquitin; Y, target protein.

I.6.2. Alternative mammalian three-hybrid systems

Next to the conventional M3H system, MASPIT and KISS, several strategies have been developed to study small-molecule induced protein-protein interactions in a mammalian setting based on protein-fragment complementation assays (PCAs). A first example of such an assay is the split-DHFR approach, which was described by Remy and Michnick and preceded the previously outlined split-ubiquitin sensor (see Section I.6.1).[118] The strategy is based on the reassembly of two designed complementary fragments of the split-DHFR sensor, each fused to either of two proteins or protein fragments that dimerize upon small-molecule addition; in this case the well-characterized rapamycin-induced association of FKBP to FRB is used. Correct folding of DHFR reconstitutes enzyme activity, which can be detected *in vivo* by DHFR-negative cell survival in medium depleted of nucleotides or by stoichiometric binding and retention of fluorescein-conjugated MTX followed by fluorescence microscopy/spectroscopy, or fluorescence-activated cell sorting (FACS). Since the successful introduction of these first test DHFR assays, there has been a veritable explosion of applications utilizing split-protein systems in numerous fields such as cell biology, drug discovery and chemical biology.[119],[120],[121] In fact, many split-protein reporters were initially validated using the FKBP-rapamycin-FRB CID system including split β-lactamase,[122],[123] split firefly luciferase,[124] split *Gaussia* luciferase,[125] and split TEV (tobacco etch virus protease).[126],[127] However, to our knowledge, these

various split-protein systems have not been applied in the context of small molecule-target protein profiling so far.

A final intriguing methodology introduced by Jester *et al.* is the coiled-coil enabled split-luciferase three-hybrid system for profiling kinase inhibitors (Figure I.9).[128] Contrary to the latter split-protein approaches, this strategy utilizes a cell free translation system, thereby circumventing potential issues related to cell permeability, large-scale protein purification and off-target effects. In the initial design, a protein kinase of interest is coupled to the C-terminal fragment of split firefly luciferase whereas its N-terminal part is fused to the coiled-coil Fos peptide, which is specific for the coiled-coil Jun. The establishment of a ternary complex, mediated by addition of a heterodimerizer consisting of Jun tethered to the pan-kinase inhibitor staurosporine, results in the reassembly of active luciferase, allowing a straightforward luminescent read-out. Kinase inhibitors were then rapidly identified by competitively displacing the staurosporine-based CID with reversal of luciferase reassembly and concomitant loss in luminescence.

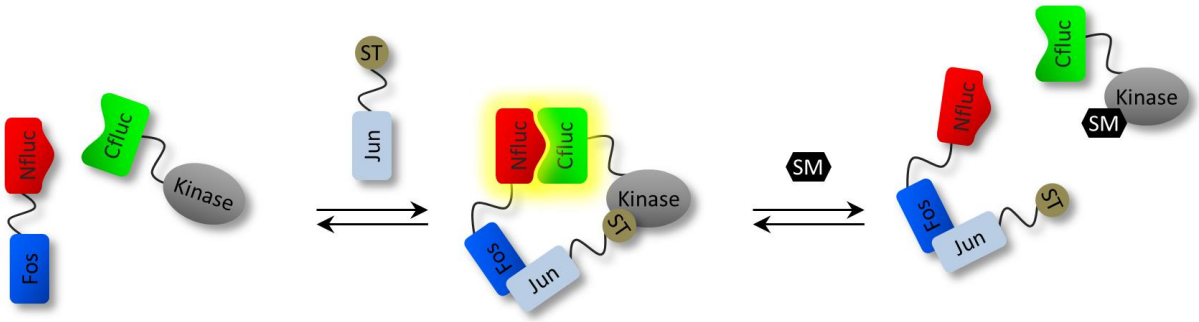


Figure I.9: Design of the coiled-coil enabled split-luciferase three-hybrid system. The two halves of split-luciferase are tethered to a kinase and coiled-coil Fos, respectively. Addition of the Jun-staurosporine CID results in ternary complex formation and reconstitution of active luciferase. Small-molecule kinase inhibitors competitively displace the CID, thereby decreasing luciferase signal. Nfluc/Cfluc, N/C-terminal half of firefly luciferase; SM, small molecule; ST, staurosporine.

At first, the potential generality of this new three-hybrid biosensor was demonstrated by successfully interrogating ten representative protein kinases from six groups in competitive binding assays. However, in a recent follow-up paper it was found that the Fos and either of six tyrosine kinase (TK) fusion proteins (PTK6, EPHB3, RIPK2, ABL, SRC and EPHA2) did not show significant luminescence in the presence of the staurosporine warhead based CID, as anticipated from the reported low affinity interactions of the latter inhibitor with these TKs.[129] Hence, in an effort to expand the coiled-coil enabled method to the entire kinome, particularly the TK group, dasatinib based CIDs were designed and implemented as active site directed ligands. Most notably, the authors showed that a CID based on a lower affinity derivative of dasatinib was necessary for being amenable to displacement by free inhibitors in the three-hybrid assay. Finally, as a proof-of-principle experiment, this modified 3H split-luciferase approach was employed to screen a broad panel of known kinase inhibitors against ABL, providing potential starting points for new drug design campaigns.

I.7. Three-hybrid screening platforms

Adoption of 3H technologies in a small molecule profiling platform, aimed at *de novo* identification of (on- or off-) target proteins, requires access to a set of potential target proteins (preferably proteome-wide) and an approach that enables efficient screening of this collection. Essentially two general methodologies can be pursued in high-throughput 3H applications: cDNA library and array screening.

Traditionally, mainly the former has been applied because this requires relatively little resources with respect to both building and screening extensive libraries. Moreover, since the screening format is essentially the same as in the case of (yeast or mammalian) two-hybrid technologies for the identification of protein-protein interactions, tools and libraries, which are broadly distributed within the research community and can be acquired commercially, can be applied as such in a 3H setup. In brief, a mRNA sample derived from (a) relevant cell type(s) is cloned as cDNA in a suitable prey vector and the resulting cDNA library is co-expressed together with the anchor protein (e.g. the LBD-DBD fusion protein in the case of Y3H) in the appropriate cell type. The resulting pool of cells, ideally each expressing a different prey protein, is treated with a CID containing the bait small molecule and cells harboring an interacting small molecule bait-target prey pair can be recovered through selection for reporter gene activity. In yeast systems, growth selection through an auxotrophic marker potentially combined with a β-lactamase-based colorimetric read-out is typically used.[15],[32],[44],[49] In mammalian cells, a wider variety of growth selection (mostly based on antibiotic resistance markers), enzymatic (e.g. luciferase) or fluorescent read-outs are being applied.[16],[130] Although seemingly straightforward, cDNA library screening is quite labor- and time-consuming: a screen for a single small-molecule bait typically requires several weeks, if not months, from the actual library screen up to the selection of validated clones. In addition, once these clones have been selected, target protein identity still needs to be retrieved through prey cDNA amplification and sequencing.

An alternative format that accommodates higher throughputs is array screening. This entails parallel testing in an arrayed fashion of individual combinations of the small-molecule bait of interest with a (large) set of target preys, typically in microtiterplates or using (micro-)arrays. Conceptually, an important benefit is that, because preys are assayed individually rather than in the bulk of a cDNA library-expressing cell pool, there is no competitive selection among different target proteins, which results in a higher level of sensitivity of an array format. Compared to cDNA library screening, setting up an array screening platform requires much more effort in terms of generation of the library of cDNA clones (which need to be cloned individually) and infrastructure for printing and scanning the arrays. Once operational however, array replicas can be produced batch-wise (potentially in a largely automated way) and stored as a flexible screening resource, and an efficient arrayed 3H platform enables screening large collections of target proteins within days, whereby target identity is revealed by its position in the array.

A focused yeast cell array covering a limited subset of protein kinases has been used to complement a cDNA library screening approach aimed at identifying novel targets of purvalanol B, as mentioned in Section I.3. Here, array replicas of yeast cell clones stably expressing different kinase prey fusion proteins were generated and screened with different concentrations of a number of small-molecule CIDs.[44] Over the past couple of years, the Tavernier group has invested heavily in adopting the

MASPIT and KISS M3H assays to a cell array format. In the current setup, functionalized mixtures containing individual prey-encoding plasmids are being printed at high density and screened through reverse transfection. In this approach, a HEK293T cell pool, bulk transfected with the CR-eDHFR anchor protein and treated with the small molecule-derived CID (see Figure I.5), is grown as a monolayer on the array, resulting in local uptake and expression of the target prey. In cells growing on top of a plasmid encoding a target protein of the small molecule of interest, the reporter gene is activated, flagging the interaction.

Although such an approach clearly exhibits vast potential as to scale and speed, besides the fact that it requires significant upfront investments to have such a platform in place, there are also a number of other points to consider. A potential disadvantage compared to cDNA library-based approaches is that, in particular when only full size ORFs are being used, interactions with specific subdomains that are not properly exposed (e.g. in cases where the small molecule binds to a subdomain of the target protein that is only accessible in specific protein conformers induced under particular physiologic conditions) might be missed. Additionally, a cDNA library might cover target protein isoforms not present in a full size ORF collection. These issues can be easily addressed however by growing the arrayed library beyond full size ORFs, by adding in additional protein isoforms and subdomains in order to move towards the complexity of classical cDNA libraries.

References

[1] D. C. Swinney, J. Anthony, *Nat. Rev. Drug Discov.* **2011**, *10*, 507-519.
[2] J. Drews, *Science* **2000**, *287*, 1960-1964.
[3] C. Human Genome Sequencing, *Nature* **2004**, *431*, 931-945.
[4] J. Mestres, E. Gregori-Puigjane, S. Valverde, R. V. Sole, *Mol. Biosyst.* **2009**, *5*, 1051-1057.
[5] A. L. Hopkins, *Nat. Chem. Biol.* **2008**, *4*, 682-690.
[6] D. C. Swinney, *Clin. Pharmacol. Ther.* **2013**, *93*, 299-301.
[7] F. Sams-Dodd, *Drug Discovery Today* **2005**, *10*, 139-147.
[8] J. Eder, R. Sedrani, C. Wiesmann, *Nat. Rev. Drug Discov.* **2014**, *13*, 577-587.
[9] J. G. Moffat, J. Rudolph, D. Bailey, *Nat. Rev. Drug Discov.* **2014**, *13*, 588-602.
[10] J. A. Lee, M. T. Uhlik, C. M. Moxham, D. Tomandl, D. J. Sall, *J. Med. Chem.* **2012**, *55*, 4527-4538.
[11] G. C. Terstappen, C. Schlupen, R. Raggiaschi, G. Gaviraghi, *Nat. Rev. Drug Discov.* **2007**, *6*, 891-903.
[12] J. Lee, M. Bogyo, *Curr. Opin. Chem. Biol.* **2013**, *17*, 118-126.
[13] S. Ziegler, V. Pries, C. Hedberg, H. Waldmann, *Angew. Chem.-Int. Edit.* **2013**, *52*, 2744-2792.
[14] J. N. Y. Chan, C. Nislow, A. Emili, *Trends Pharmacol. Sci.* **2010**, *31*, 82-88.
[15] E. J. Licitra, J. O. Liu, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 12817-12821.
[16] M. Caligiuri, L. Molz, Q. Liu, F. Kaplan, J. P. Xu, J. Z. Majeti, R. Ramos-Kelsey, K. Murthi, S. Lievens, J. Tavernier, N. Kley, *Chem. Biol.* **2006**, *13*, 711-722.
[17] D. M. Spencer, T. J. Wandless, S. L. Schreiber, G. R. Crabtree, *Science* **1993**, *262*, 1019-1024.
[18] A. Fegan, B. White, J. C. T. Carlson, C. R. Wagner, *Chem. Rev.* **2010**, *110*, 3315-3336.
[19] T. Clackson, in *Chemical Biology*, Wiley-VCH Verlag GmbH, **2008**, pp. 227-249.
[20] T. W. Corson, N. Aberle, C. M. Crews, *ACS Chem. Biol.* **2008**, *3*, 677-692.
[21] M. Rezwan, D. Auerbach, *Methods* **2012**, *57*, 423-429.
[22] A. Hamdi, P. Colas, *Trends Pharmacol. Sci.* **2012**, *33*, 109-118.
[23] N. Kley, *Chem. Biol.* **2004**, *11*, 599-608.
[24] G. R. Crabtree, S. L. Schreiber, *Trends Biochem.Sci.* **1996**, *21*, 418-422.
[25] J. Liu, J. D. Farmer, W. S. Lane, J. Friedman, I. Weissman, S. L. Schreiber, *Cell* **1991**, *66*, 807-815.
[26] J. W. Choi, J. Chen, S. L. Schreiber, J. Clardy, *Science* **1996**, *273*, 239-242.
[27] A. Rutkowska, C. Schultz, *Angew. Chem.-Int. Edit.* **2012**, *51*, 8166-8176.
[28] J. E. Gestwicki, P. S. Marinec, *Comb. Chem. High Throughput Screen* **2007**, *10*, 667-675.
[29] M. Putyrski, C. Schultz, *FEBS Lett.* **2012**, *586*, 2097-2105.
[30] S. Fields, O. K. Song, *Nature* **1989**, *340*, 245-246.
[31] D. Auerbach, I. Stagljar, in *Proteomics and Protein-Protein Interactions, Vol. 3* (Ed.: G. Waksman), Springer US, **2005**, pp. 19-31.
[32] D. C. Henthorn, A. A. Jaxa-Chamiec, E. Meldrum, *Biochem. Pharmacol.* **2002**, *63*, 1619-1628.
[33] S. N. Ho, S. R. Biggar, D. M. Spencer, S. L. Schreiber, G. R. Crabtree, *Nature* **1996**, *382*, 822-826.
[34] P. A. Clemons, B. G. Gladstone, A. Seth, E. D. Chao, M. A. Foley, S. L. Schreiber, *Chem. Biol.* **2002**, *9*, 49-61.
[35] A. Borchardt, S. D. Liberles, S. R. Biggar, G. R. Crabtree, S. L. Schreiber, *Chem. Biol.* **1997**, *4*, 961-968.
[36] S. L. Hussey, S. S. Muddana, B. R. Peterson, *J. Am. Chem. Soc.* **2003**, *125*, 3692-3693.
[37] S. S. Muddana, B. R. Peterson, *Org. Lett.* **2004**, *6*, 1409-1412.
[38] H. Lin, W. M. Abida, R. T. Sauer, V. W. Cornish, *J. Am. Chem. Soc.* **2000**, *122*, 4247-4248.
[39] W. M. Abida, B. T. Carter, E. A. Althoff, H. N. Lin, V. W. Cornish, *ChemBioChem* **2002**, *3*, 887-895.
[40] K. Baker, D. Sengupta, G. Salazar-Jimenez, V. W. Cornish, *Anal. Biochem.* **2003**, *315*, 134-137.
[41] L. M. Wingler, V. W. Cornish, *ChemBioChem* **2011**, *12*, 715-717.
[42] M. D. Harton, L. M. Wingler, V. W. Cornish, *Biotechnol. J.* **2013**, *8*, 1485-1491.
[43] K. S. de Felipe, B. T. Carter, E. A. Althoff, V. W. Cornish, *Biochemistry* **2004**, *43*, 10353-10363.
[44] F. Becker, K. Murthi, C. Smith, J. Come, N. Costa-Roldan, C. Kaufmann, U. Hanke, C. Degenhart, S. Baumann, W. Wallner, A. Huber, S. Dedier, S. Dill, D. Kinsman, M. Hediger, N. Bockovich, S. Meier-Ewert, A. F. Kluge, N. Kley, *Chem. Biol.* **2004**, *11*, 211-223.
[45] S. Lefurgy, V. Cornish, *Chem. Biol.* **2004**, *11*, 151-153.
[46] D. Dirnberger, G. Unsin, S. Schlenker, C. Reichel, *ChemBioChem* **2006**, *7*, 936-942.
[47] M. Caligluri, F. Becker, K. Murthi, F. Kaplan, S. Dedier, C. Kaufmann, A. Machl, G. Zybarth, J. Richard, N. Bockovich, A. Kluge, N. Kley, *Chem. Biol.* **2005**, *12*, 1103-1115.
[48] P. L. Toogood, *Chem. Biol.* **2005**, *12*, 1057-1058.

[49] A. R. Shepard, R. E. Conrow, I. H. Pang, N. Jacobson, M. Rezwan, K. Rutschmann, D. Auerbach, R. SriRamaratnam, V. W. Cornish, *ACS Chem. Biol.* **2013**, *8*, 549-558.

[50] A. V. Odell, F. Tran, J. E. Foderaro, S. Poupart, R. Pathak, N. J. Westwood, G. E. Ward, *PLoS One* **2015**, *10*.

[51] F. Tran, A. V. Odell, G. E. Ward, N. J. Westwood, *Molecules* **2013**, *18*, 11639-11657.

[52] S. S. Gallagher, L. W. Miller, V. W. Cornish, *Anal. Biochem.* **2007**, *363*, 160-162.

[53] J. E. Bronson, W. W. Mazur, V. W. Cornish, *Mol. Biosyst.* **2008**, *4*, 56-58.

[54] J. L. Czlapinski, M. W. Schelle, L. W. Miller, S. T. Laughlin, J. J. Kohler, V. W. Cornish, C. R. Bertozzi, *J. Am. Chem. Soc.* **2008**, *130*, 13186-13187.

[55] S. Gendreizig, M. Kindermann, K. Johnsson, *J. Am. Chem. Soc.* **2003**, *125*, 14970-14971.

[56] T. Gronemeyer, C. Chidley, A. Juillerat, C. Heinis, K. Johnsson, *Protein Eng. Des. Sel.* **2006**, *19*, 309-316.

[57] C. Chidley, H. Haruki, M. G. Pedersen, E. Muller, K. Johnsson, *Nat. Chem. Biol.* **2011**, *7*, 375-383.

[58] C. Chidley, H. Haruki, M. G. Pedersen, C. Fellay, S. Moser, K. Johnsson, *Chimia* **2011**, *65*, 720-724.

[59] S. Moser, K. Johnsson, *ChemBioChem* **2013**, *14*, 2239-2242.

[60] T. Clackson, *Curr. Opin. Struct. Biol.* **1998**, *8*, 451-458.

[61] R. Pollock, T. Clackson, *Curr. Opin. Biotechnol.* **2002**, *13*, 459-467.

[62] R. Pollock, M. Giel, K. Linher, T. Clackson, *Nat. Biotechnol.* **2002**, *20*, 729-733.

[63] M. J. Garabedian, K. R. Yamamoto, *Mol. Biol. Cell* **1992**, *3*, 1245-1257.

[64] R. Sitcheran, R. Emter, A. Kralli, K. R. Yamamoto, *Genetics* **2000**, *156*, 963-972.

[65] S. L. Hussey, E. He, B. R. Peterson, *Org. Lett.* **2002**, *4*, 415-418.

[66] J. R. Appleman, N. Prendergast, T. J. Delcamp, J. H. Freisheim, R. L. Blakley, *J. Biol. Chem.* **1988**, *263*, 10304-10313.

[67] E. A. Althoff, V. W. Cornish, *Angew. Chem.-Int. Edit.* **2002**, *41*, 2327-2330.

[68] J. Estojak, R. Brent, E. A. Golemis, *Mol. Cell. Biol.* **1995**, *15*, 5820-5829.

[69] D. P. Baccanari, S. Daluge, R. W. King, *Biochemistry* **1982**, *21*, 5068-5075.

[70] L. W. Miller, Y. F. Cai, M. P. Sheetz, V. W. Cornish, *Nat. Methods* **2005**, *2*, 255-257.

[71] C. R. Jing, V. W. Cornish, *ACS Chem. Biol.* **2013**, *8*, 1704-1712.

[72] M. J. Hinner, K. Johnsson, *Curr. Opin. Biotechnol.* **2010**, *21*, 766-776.

[73] R. Wombacher, V. W. Cornish, *J. Biophotonics* **2011**, *4*, 391-402.

[74] C. R. Jing, V. W. Cornish, *Accounts Chem. Res.* **2011**, *44*, 784-792.

[75] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, *Nat. Biotechnol.* **2003**, *21*, 86-89.

[76] http://www.hybrigenics-services.com/contents/our-services/discover/ultimate-ychemh

[77] S. T. Diver, S. L. Schreiber, *J. Am. Chem. Soc.* **1997**, *119*, 5106-5109.

[78] R. A. Freiberg, S. N. Ho, P. A. Khavari, *J. Clin. Invest.* **1997**, *99*, 2610-2615.

[79] J. F. Amara, T. Clackson, V. M. Rivera, T. Guo, T. Keenan, S. Natesan, R. Pollock, W. Yang, N. L. Courage, D. A. Holt, M. Gilman, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 10618-10623.

[80] V. M. Rivera, *Methods-a Companion to Methods in Enzymology* **1998**, *14*, 421-429.

[81] T. Keenan, D. R. Yaeger, N. L. Courage, C. T. Rollins, M. E. Pavone, V. M. Rivera, W. Yang, T. Guo, J. F. Amara, T. Clackson, M. Gilman, D. A. Holt, *Bioorg. Med. Chem.* **1998**, *6*, 1309-1335.

[82] W. Yang, T. P. Keenan, L. W. Rozamus, X. R. Wang, V. M. Rivera, C. T. Rollins, T. Clackson, D. A. Holt, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3181-3184.

[83] R. Pollock, V. M. Rivera, in *Methods in Enzymology, Vol. Volume 306* (Ed.: M. C. S. Joseph C. Glorioso), Academic Press, **1999**, pp. 263-281.

[84] V. Senner, A. Sotoodeh, W. Paulus, *Neurochem. Res.* **2001**, *26*, 521-524.

[85] P. J. Belshaw, S. N. Ho, G. R. Crabtree, S. L. Schreiber, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 4604-4607.

[86] K. Koide, J. M. Finkelstein, Z. Ball, G. L. Verdine, *J. Am. Chem. Soc.* **2001**, *123*, 398-408.

[87] V. M. Rivera, T. Clackson, S. Natesan, R. Pollock, J. F. Amara, T. Keenan, S. R. Magari, T. Phillips, N. L. Courage, F. Cerasoli, D. A. Holt, M. Gilman, *Nat. Med.* **1996**, *2*, 1028-1032.

[88] S. Natesan, E. Molinari, V. M. Rivera, R. J. Rickles, M. Gilman, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 13898-13903.

[89] W. Y. Go, S. N. Ho, *J. Gene. Med.* **2002**, *4*, 258-270.

[90] S. D. Liberles, S. T. Diver, D. J. Austin, S. L. Schreiber, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 7825-7830.

[91] X. C. Chen, C. R. Chong, L. R. Shi, T. Yoshimoto, D. J. Sullivan, J. O. Liu, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 14548-14553.

[92] X. C. Chen, S. Xie, S. Bhat, N. Kumar, T. A. Shapiro, J. O. Liu, *Chem. Biol.* **2009**, *16*, 193-202.

[93] E. Balzi, A. Goffeau, *J. Bioenerg. Biomembr.* **1995**, *27*, 71-76.
[94] M. Kolaczkowski, A. Kolaczkowska, J. Luczynski, S. Witek, A. Goffeau, *Microb. Drug Resist.-Mechan. Epidemiol. Dis.* **1998**, *4*, 143-158.
[95] B. Rogers, A. Decottignies, M. Kolaczkowski, E. Carvajal, E. Balzi, A. Goffeau, *J. Mol. Microbiol. Biotechnol.* **2001**, *3*, 207-214.
[96] R. Emter, A. Heese-Peck, A. Kralli, *FEBS Lett.* **2002**, *521*, 57-61.
[97] J. Kato-Stankiewicz, I. Hakimi, G. Zhi, J. Zhang, I. Serebriiskii, L. Guo, H. Edamatsu, H. Koide, S. Menon, R. Eckl, S. Sakamuri, Y. C. Lu, Q. Z. Chen, S. Agarwal, W. R. Baumbach, E. A. Golemis, F. Tamanoi, V. Khazak, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14398-14403.
[98] D. D. Clark, B. R. Peterson, *ChemBioChem* **2003**, *4*, 101-107.
[99] A. Nourani, M. WesolowskiLouvel, T. Delaveau, C. Jacq, A. Delahodde, *Mol. Cell. Biol.* **1997**, *17*, 5453-5460.
[100] S. Eyckerman, A. Verhee, J. Van der Heyden, I. Lemmens, X. Van Ostade, J. Vandekerckhove, J. Tavernier, *Nat. Cell Biol.* **2001**, *3*, 1114-1119.
[101] S. Lievens, F. Peelman, K. De Bosscher, I. Lemmens, J. Tavernier, *Cytokine Growth Factor Rev.* **2011**, *22*, 321-329.
[102] S. Lievens, S. Gerlo, I. Lemmens, D. J. H. De Clercq, M. D. P. Risseeuw, N. Vanderroost, A.-S. De Smet, E. Ruyssinck, E. Chevet, S. Van Calenbergh, J. Tavernier, *Mol. Cell. Proteomics* **2014**, *13*, 3332-3342.
[103] J. T. Bolin, D. J. Filman, D. A. Matthews, R. C. Hamlin, J. Kraut, *J. Biol. Chem.* **1982**, *257*, 13650-13662.
[104] X. E. Wells, V. J. Bender, C. L. Francis, H. M. He-Williams, M. K. Manthey, M. J. Moghaddam, W. G. Reilly, R. G. Whittaker, *Drug Dev. Res.* **1999**, *46*, 302-308.
[105] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem.-Int. Edit.* **2002**, *41*, 2596-2599.
[106] K. Baker, C. Bleczinski, H. N. Lin, G. Salazar-Jimenez, D. Sengupta, S. Krane, V. W. Cornish, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 16537-16542.
[107] H. Y. Tao, P. Peralta-Yahya, H. N. Lin, V. W. Cornish, *Bioorg. Med. Chem.* **2006**, *14*, 6940-6953.
[108] H. N. Lin, H. Y. Tao, V. W. Cornish, *J. Am. Chem. Soc.* **2004**, *126*, 15051-15059.
[109] H. Y. Tao, P. Peralta-Yahya, J. Decatur, V. W. Cornish, *ChemBioChem* **2008**, *9*, 681-684.
[110] D. Sengupta, H. N. Lin, S. D. Goldberg, J. J. Mahal, V. W. Cornish, *Biochemistry* **2004**, *43*, 3570-3581.
[111] B. T. Carter, H. N. Lin, S. D. Goldberg, E. A. Althoff, J. Raushel, V. W. Cornish, *ChemBioChem* **2005**, *6*, 2055-2067.
[112] P. Peralta-Yahya, B. T. Carter, H. Lin, H. Tao, V. W. Cornish, *J. Am. Chem. Soc.* **2008**, *130*, 17446-17452.
[113] D. D. Clark, B. R. Peterson, *J. Proteome Res.* **2002**, *1*, 207-209.
[114] S. Athavankar, B. R. Peterson, *Chem. Biol.* **2003**, *10*, 1245-1253.
[115] W. Weber, W. Bacchus, M. D. Baba, M. Fussenegger, *Nucleic Acids Res.* **2007**, *35*, 13.
[116] W. Weber, W. Bacchus, F. Gruber, M. Hamberger, M. Fussenegger, *J. Biotechnol.* **2007**, *131*, 150-158.
[117] N. Johnsson, A. Varshavsky, *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 10340-10344.
[118] I. Remy, S. W. Michnick, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 5394-5399.
[119] I. Remy, S. W. Michnick, *Biotechniques* **2007**, *42*, 137-145.
[120] S. W. Michnick, P. H. Ear, E. N. Manderson, I. Remy, E. Stefan, *Nat. Rev. Drug Discov.* **2007**, *6*, 569-582.
[121] S. S. Shekhawat, I. Ghosh, *Curr. Opin. Chem. Biol.* **2011**, *15*, 789-797.
[122] A. Galarneau, M. Primeau, L. E. Trudeau, S. W. Michnick, *Nat. Biotechnol.* **2002**, *20*, 619-622.
[123] T. Wehrman, B. Kleaveland, J. H. Her, R. F. Balint, H. M. Blau, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 3469-3474.
[124] K. E. Luker, M. C. P. Smith, G. D. Luker, S. T. Gammon, H. Piwnica-Worms, D. P. Piwnica-Worms, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12288-12293.
[125] I. Remy, S. W. Michnick, *Nat. Methods* **2006**, *3*, 977-979.
[126] M. C. Wehr, R. Laage, U. Bolz, T. M. Fischer, S. Grunewald, S. Scheek, A. Bach, K. A. Nave, M. J. Rossner, *Nat. Methods* **2006**, *3*, 985-993.
[127] D. J. Williams, H. L. Puhl, S. R. Ikeda, *PLoS One* **2009**, *4*, 14.
[128] B. W. Jester, K. J. Cox, A. Gaj, C. D. Shomin, J. R. Porter, I. Ghosh, *J. Am. Chem. Soc.* **2010**, *132*, 11727-11735.
[129] L. O. Ogunleye, B. W. Jester, A. J. Riemen, A. H. Badran, P. Wang, I. Ghosh, *MedChemComm* **2014**, *5*, 328-332.
[130] S. Lievens, J. Van der Heyden, E. Vertenten, J. Plum, J. Vandekerckhove, J. Tavernier, in *Flow Cytometry Protocols, Vol. 263* (Eds.: T. Hawley, R. Hawley), Humana Press, **2004**, pp. 293-309.

CHAPTER II

OBJECTIVES

II. Objectives

The overarching aim of this thesis is to optimize the performance of the MASPIT system from a medicinal chemical perspective (Figure II.1). As described in the general introduction, a prerequisite for successful MASPIT analysis is the availability of appropriate methotrexate fusion compounds (MFCs). Structure-activity relationship (SAR) studies of methotrexate (MTX) derivatives have emphasized the importance of selective conjugation to the γ-carboxylic acid of the glutamate moiety to ensure high affinity binding to dihydrofolate reductase (DHFR) of *E. coli*, since the α-carboxylic acid significantly contributes to the overall binding by forming an ion pair with the guanidinium group of Arg-57, a residue that is highly conserved in prokaryotic and eukaryotic dihydrofolate reductases, while the γ-carboxylate appears to interact only with the solvent. Hence, a first objective of this work is to design and synthesize a versatile MTX-based reagent and explore its use for easy and straightforward ligation to bait small molecules of interest. In this respect, we also set out to determine the optimal linker length and assess different conjugation methodologies (Figure II.1, panel A).

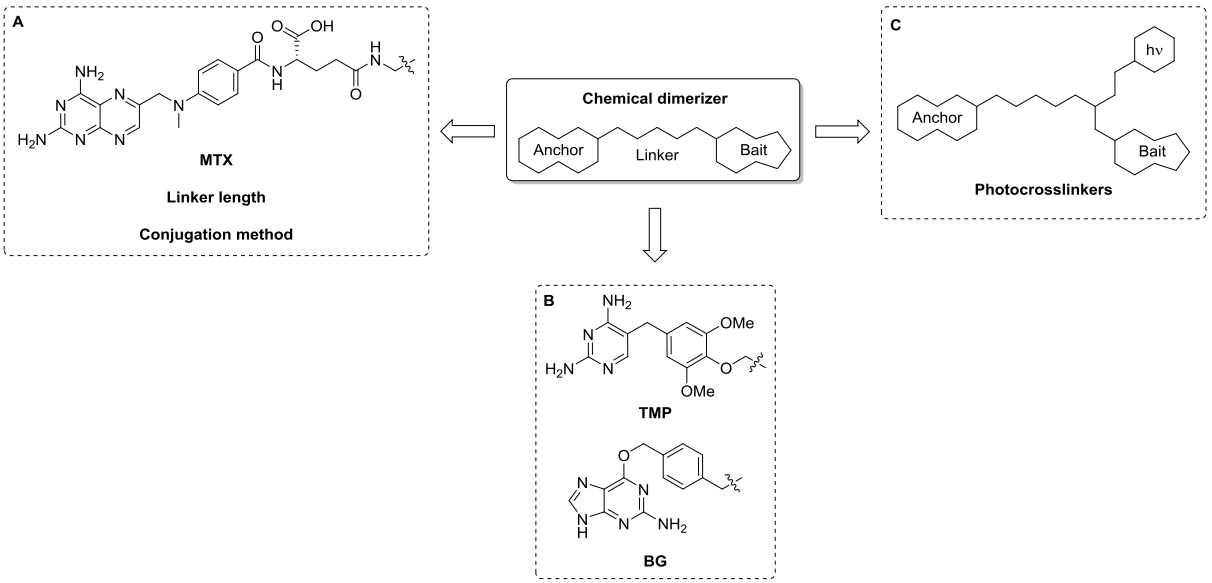


Figure II.1: Schematic overview of the chemical dimerizer optimization approaches for MASPIT. BG, O^6-benzylguanine; hν, photoactivatable group; MTX, methotrexate; TMP, trimethoprim.

In a next step, two alternative chemical dimerizer approaches, i.e. TMP- and SNAP-tag, will be investigated with the aim of increasing the sensitivity of MASPIT. First, we will try to circumvent any potential limitations related to the tight binding of MTX to endogenous human DHFR, which might titrate out a portion of the fusion compound and induce cellular toxicity through perturbation of the endogenous folate metabolism. As an *Escherichia coli* enzyme is employed in MASPIT, we will explore trimethoprim (TMP) as an alternative prokaryote-specific DHFR ligand (Figure II.1, panel B).

Furthermore, so as to stabilize the ternary complex (cytokine receptor-fusion compound-prey chimera) in order to improve the system's sensitivity, we will introduce the concept of covalent bonding into the MASPIT assay, which hitherto relies on reversible interactions of either component of the dimerizers. As a starting point, we aim to selectively and covalently immobilize the fusion

compound to the cytokine receptor (CR) by using a SNAP-tag-based system, which is centered around the human O^6-alkylguanine-DNA alkyltransferase (hAGT) enzyme. Previously the low substrate specificity of this enzyme has been exploited to covalently label SNAP-tag-fused proteins *in vivo* with a ligand of interest by conjugating the latter to the *para*-position of O^6-benzylguanine (BG) (Figure II.1, panel B). Therefore, by fusing hAGT to the cytoplasmic domain of the CR and synthesizing the appropriate BG fusion compound (BGFC), we plan to present a covalently coupled bait small molecule.

Our final optimization efforts will be aimed at introducing covalent bonding on the bait-end of the dimerizer. For this purpose, we envision that photoactivatable crosslinkers will be particularly suitable for capturing transient and low-affinity bait-prey interactions. Indeed, such photocrosslinkers convert upon photoactivation to highly reactive species, capable of introducing covalent bonds between interacting molecules. In the context of the MASPIT assay, this strategy thus might give rise to increased signal output. Hence, we will equip the anchor-spacer reagent with a double ligation handle, enabling introduction of both bait small molecules of interest and suitable photophores, to yield heterotrimeric photoaffinity labeling (PAL) probes (Figure II.1, panel C).

CHAPTER III

A 'CLICKABLE' MTX REAGENT

A 'Clickable' MTX Reagent as a Practical Tool for Profiling Small-Molecule–Intracellular Target Interactions via MASPIT

The content of this chapter was derived from:
M. D. P. Risseeuw,* D. J. H. De Clercq,* S. Lievens, U. Hillaert, D. Sinnaeve, F. Van den Broeck, J. C. Martins, J. Tavernier, S. Van Calenbergh, *ChemMedChem* **2013**, *8*, 521-526.
*Co-first author

III. A 'Clickable' MTX Reagent as a Practical Tool for Profiling Small-Molecule–Intracellular Target Interactions via MASPIT

III.1. Small-molecule baits

To optimize the CIDs for use in MASPIT, we selected four structurally diverse, pharmacologically active compounds as model baits: tamoxifen, reversine, FK506, and simvastatin. In the following subsections some background on their mechanism of action (MOA) as well as the specific motivation for our interest in these baits is provided. Furthermore, we discuss available structure-activity relationship (SAR), structural, and molecular modeling data underlying our rationale for choosing a particular attachment position, which tolerates the incorporation of appropriate chemical handles necessary for the derivatization of these small-molecule baits into MASPIT-compatible CIDs.

III.1.1. Tamoxifen

III.1.1.1. MOA and motivation

Tamoxifen (TAM) is a selective estrogen receptor (ER) modulator, which displays antagonistic effects on the ERs of breast tissue, accounting for its ability to inhibit tumor growth.[1] Hence, tamoxifen has been part of the standard therapy for ER-positive breast cancer treatment since the 1970s.[2] The drug is extensively metabolized, predominantly by the liver cytochrome P450 (CYP) enzymes 2D6 and 3A4, yielding several active metabolites such as (*Z*)-4-hydroxytamoxifen (4-OH-TAM) and (*Z*)-4-hydroxy-*N*-desmethyltamoxifen (endoxifen) (Figure III.1).[3] The latter exhibit both a much higher antiestrogenic potency in breast cancer cells and a greater affinity for ERα compared to the parent drug (e.g., K_i(4-OH-TAM) = 0.25 nM vs. K_i(TAM) = 12 nM),[4] emphasizing their contribution in mediating the overall clinical outcome.[5] Of note, the *Z*-configuration of the double bond of tamoxifen and its metabolites is required for high affinity ER binding; the *E* isomers have low binding affinity.[6]

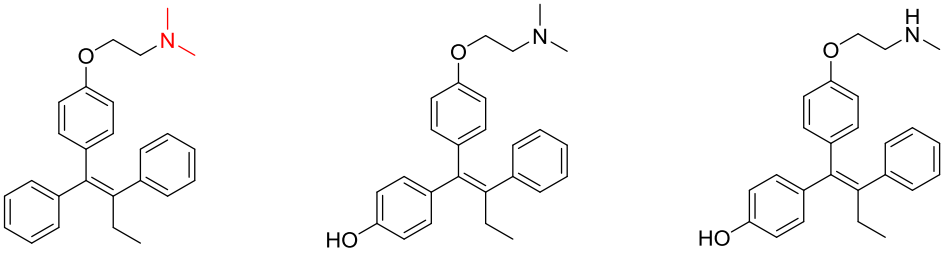


Figure III.1: Chemical structures of tamoxifen and its main active metabolites, (*Z*)-4-hydroxytamoxifen (4-OH-TAM) and (*Z*)-4-hydroxy-*N*-desmethyltamoxifen (endoxifen). The functional group selected as conjugation site is highlighted in red.

In humans, two ER isoforms are found, ERα and ERβ, each encoded by separate genes and exhibiting different affinities for certain ligands.[7],[8] The vast majority of breast carcinomas, i.e. 80%, are positive for and driven by ERα. The latter is a master transcription factor and belongs to the steroid hormone receptor superfamily.[9] ERα can be activated in a ligand-dependent as well as independent fashion, via modulation by its natural ligand 17β-estradiol or by a number of posttranslational

modifications, such as direct phosphorylation by several protein kinases, respectively.[10] In the classical ligand-dependent signaling pathway estradiol diffuses into the cell and binds ERα, located in the cytosol, triggering its dissociation from the stabilizing heat shock protein 90 (HSP90) and subsequent homodimerization. Next, activated ERα dimers translocate into the nucleus and bind the DNA at estrogen response elements (ERE), resulting in the recruitment of the transcription complex. Ultimately this cascade leads to a downregulation of anti-proliferative and pro-apoptotic genes, while upregulating responsive genes involved in cell cycle progression.[11]

Tamoxifen functions as a competitive antagonist, which upon binding to ERα alters the conformation of the ligand-binding domain (LBD), thereby preventing the formation of an active transcription complex.[10] Consequently, tamoxifen inhibits the expression of estrogen-regulated genes, including growth and angiogenic factors, resulting in a block of the cell cycle, a slowing of cell proliferation and eventually tumor regression.[1] Alternatively, tamoxifen may also directly induce apoptosis, even in ER-negative cancer cells, suggesting that it can also operate by modulating alternative targets.[12],[13] In this respect, prominent roles have been proposed for discrete signaling proteins, such as protein kinase C (PKC), calmodulin, transforming growth factor-β (TGF-β);[14] the proto-oncogene c-*MYC*;[15] oxidative stress and NF-κB[16]; mitogen-activated protein kinases (MAPK), including c-Jun N-terminal protein kinase (JNK) and p38 kinase;[17] and mitochondria and caspases.[18] However, no unique signal transduction pathway has been unambiguously attributed to TAM's pro-apoptotic effects. As yet, the exact overall MOA underlying the apparent promiscuity of this blockbuster drug remains elusive. Clearly, a better understanding of this complexity would enable the development of new and potentially more effective analogues that retain all of the beneficial effects but are devoid of toxicity and could bypass the emergence of resistance often seen with TAM.[10],[13] Taking this together, we judged this bait would be a particularly interesting test case for MASPIT.

III.1.1.2. SAR and conjugation site

Since, to our knowledge, no crystallographic data of tamoxifen in complex with ER is available, we focused on the cocrystal structure of one of its main active metabolites, 4-OH-TAM, bound to the LBD of human ERα, as described by Shiau and colleagues (Protein Data Bank (PDB) code: 3ERT).[19] The overall architecture of the ERα LBD comprises a wedge-shaped molecular scaffold consisting of three layers of 12 α-helices and a small two-stranded antiparallel β-sheet.[20] The orientation of 4-OH-TAM within the binding pocket is principally governed by the positioning of the phenolic A ring and the bulky side chain (Figure III.2A).

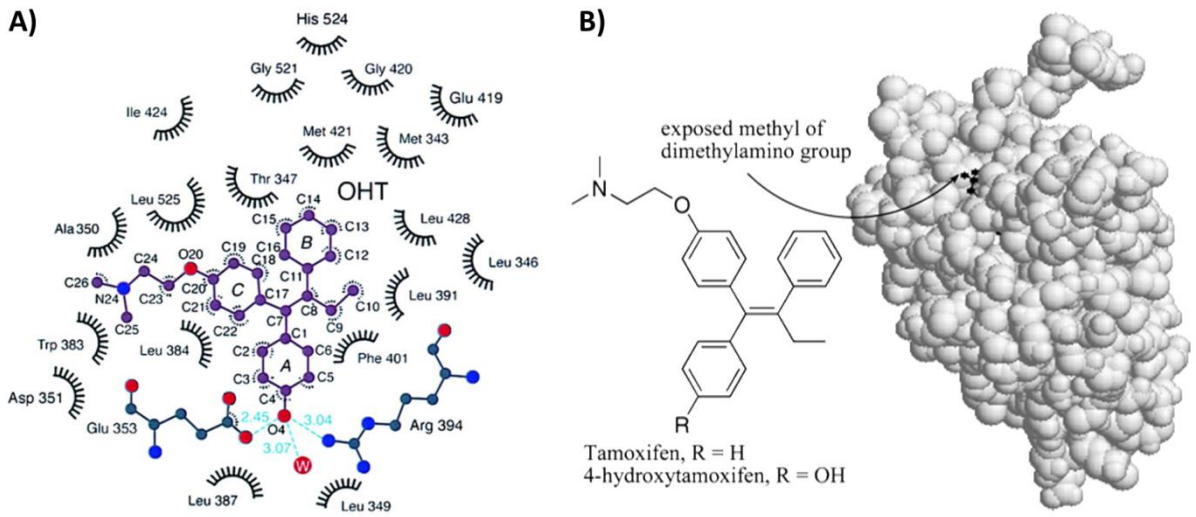


Figure III.2: A) The interactions of (*Z*)-4-hydroxytamoxifen ('OHT') with the LBD of human ERα (PDB code: 3ERT). The residues that form van der Waals contacts with the ligand are depicted as labeled arcs. Hydrogen bonding residues are shown in ball-and-stick representation, with individual hydrogen bonds represented as dashed cyan lines; the distance of each bond is indicated in Å. Figure from: *Cell* 1998, 95 (7), 927-937. **B) Space-filling model of the cocrystal structure of 4-OH-TAM bound to the LBD of ERα.** The ligand is shown in ball-and-stick representation. Figure from: *Journal of Medicinal Chemistry* 2004, 47 (5), 1193-1206.

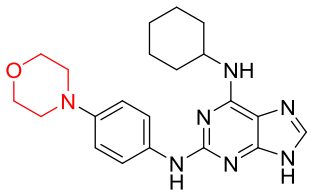
The A ring of 4-OH-TAM is bound near helices 3 and 6, and its hydroxyl function forms three hydrogen bonds with the side chains of Glu-353, Arg-394 and a structurally conserved water molecule (W), respectively. The tertiary amine group of the flexible 2-(dimethylamino)ethoxy side chain, on the other hand, is stabilized by an electrostatic interaction with the β-carboxylate of Asp-351 (3.8 Å), while the rest of the chain has van der Waals contacts with Thr-347, Ala-350, and Trp-383. Furthermore, this side chain projects out of the binding pocket between helices 3 and 11, thereby providing the structural basis of tamoxifen antagonism. Indeed, early SAR explorations already emphasized the importance of the bulky side chain of TAM analogues with respect to ER antagonism.[21],[22] More specifically, this extension of steric bulk from the rigid triphenylethylene framework promotes a conformation of the LBD, which imposes helix 12 to reposition over the coactivator recognition groove, thus preventing recruitment of coactivators and subsequent transcriptional activation.[23]

Closer examination of a space-filling model of this cocrystal structure revealed that the methyl groups of the dimethylamino function of 4-OH-TAM are solvent exposed, providing an optimal conjugation site for derivatization without disruption of binding affinity (Figure III.2B).[24] Henceforth, this particular structural information has been exploited to successfully tether tamoxifen to a variety of scaffolds, such as doxorubicin-formaldehyde conjugates,[24] gold nanoparticles,[25] histone deacetylase (HDAC) inhibitors,[26] and bile acids,[27] for targeted breast cancer therapy. Guided by these structural and experimental data, we decided to explore the dimethylamino moiety of the tamoxifen bait as starting point for our conjugation efforts.

III.1.2. Reversine

III.1.2.1. MOA and motivation

Reversine, a 2,6-disubstituted purine analogue (Figure III.3), was discovered in 2004 from a library of 50 000 heterocyclic compounds designed around a large number of kinase-directed scaffolds in a search for small molecules which induce dedifferentiation of somatic cells.[28] In particular, reversine was found to promote the reversion of myogenic lineage-committed cells to multipotent mesenchymal progenitor cells, which can proliferate and redifferentiate into osteoblasts and adipocytes under lineage-specific inducing conditions. Additionally, it was shown that reversine treatment converts primary murine and human dermal fibroblasts into myogenic-competent cells, which can be induced to differentiate into skeletal muscle at high frequency both *in vitro* and *in vivo*.[29] As such, this small molecule holds great therapeutic potential related to its ability to reprogram somatic cells to a state of increased plasticity that can be manipulated to direct differentiation in different types of functional cells for repairing damaged tissues in e.g. cardiovascular or musculoskeletal diseases.[30]



reversine

Figure III.3: Chemical structure of reversine, 2-(4-morpholinoanilino)-N^6-cyclohexyladenine. The functional group selected as conjugation site is highlighted in red.

Initial mechanism of action studies by Schultz and coworkers suggest that reversine acts as a dual inhibitor of nonmuscle myosin II (NM II) and MAPK and ERK kinase 1 (MEK1), and that inhibition of both is required for its activity.[31] The former is a cytoskeleton protein that plays a major role in processes that require cellular reshaping, such as cytokinesis.[32],[33] By blocking this protein, reversine leads to accumulation of cells in G_2/M phase, which might contribute to the increase in cellular plasticity. MEK1, on the other hand, is part of the evolutionary conserved ERK/MAPK pathway (ERK, extracellular signal-regulated kinase; MAPK, mitogen-activated protein kinase), which is involved in the control of many fundamental cellular processes including cell proliferation, survival, differentiation, apoptosis, and stem cell fate determination.[34] Moreover, this signaling pathway also regulates histone acetylation, either via direct phosphorylation of histone acetyltransferase (HAT), or via indirect modulation of its activity.[31] It was shown that reversine treatment resulted in decreased acetylation of histone H3 by inhibition of MEK-dependent signaling, thereby potentially suppressing cell-fate-determining genes. Furthermore, the authors postulate that yet other factors might contribute to the activity of reversine, including activation of the phosphatidylinositol 3-kinase (PI3K) signaling pathway, which is implicated in the regulation of functions as diverse as cell metabolism, proliferation, cell cycle and survival.[35]

In an effort to provide a molecular explanation for the function of reversine, D'Alise *et al.* tested its activity on a broad panel of protein kinases, the majority of which are involved in cell cycle regulation

and cytokinesis.[36] The small molecule was initially shown to specifically inhibit Aurora kinases, predominantly Aurora B (IC_{50} = 98.5 nM),[37] both *in vitro* and in cell-based assays. The Aurora B protein kinase is involved in multiple mitotic functions, and its inhibition produces various effects including chromosome segregation failure, spindle checkpoint override and impairment of cytokinesis.[38],[39] Collectively, these outcomes result in polyploidy and suggest the direct involvement of Aurora B on reversine-induced dedifferentiation. However, a follow-up paper from the same group showed that the inhibition of these mitotic functions *per se* is not sufficient for dedifferentiation.[40] Rather, these authors suggest a model wherein inhibition of Aurora B-mediated histone H3 phosphorylation results in chromatin remodeling and ensuing restoration of the multipotent state. Furthermore, they performed *in vitro* inhibition assays on a battery of human mitotic kinases to ascertain whether reversine is a selective Aurora B inhibitor. Interestingly, their analysis indicated that reversine is a very potent (IC_{50} = 2.8 nM) and relatively selective ATP-competitive inhibitor of human TTK (threonine and tyrosine kinase; also known as monopolar spindle 1, MPS1, kinase).[37],[41] In fact, TTK was attributed to act as a primary target kinase for reversine in mitosis, operating downstream from Aurora B.

Aurora kinases have also emerged as promising targets in cancer therapy, as they are frequently amplified and overexpressed in human tumors and as their inhibition is associated with the potential to induce cell-cycle arrest or apoptosis in a range of tumor cell types.[39],[42] In this respect, reversine was reported to display potent antitumor effects with concomitant good therapeutic selectivity in human acute myeloid leukemia cells,[36] and three human thyroid cancer cell lines.[43] However, as for its dedifferentiation capability, the exact molecular mechanism of reversine's tumor suppression effect remains to be elucidated. Hence, this spurred our interest in profiling this bait in MASPIT, not least given its potential as a useful agent in regenerative medicine and cancer chemotherapy.

III.1.2.2. SAR and conjugation site

Preliminary SAR studies on reversine indicated the necessity of both the N9 hydrogen and the NH substitution at the C2 position of the purine ring, as activity completely plummeted upon removal of either.[28] Yet, a wide variety of oxygen- or sulfur-linked substituent groups were tolerated at C6 without loss of activity, suggesting that a H-bond donor at this position is not essential. Furthermore, the scope of substituents accepted on the aniline moiety at the C2 position turned out to be rather restricted. These observations are consistent with the crystal structure of reversine in the active site of the Aurora B protein kinase (PDB code: 2VGO) (to our knowledge, no cocrystal structure of reversine bound to its primary target kinase, TTK, has been published).[36] The purine ring of reversine is positioned in the ATP-binding pocket via stacking interactions with the side chains of Ala-120, Leu-170, Ala-173, Leu-99, and Leu-223 (Figure III.4.). More important, both its N3 and N9 hydrogen bond to Ala-173 and Glu-171, respectively. Moreover, in agreement with earlier suggestions from the SAR data, the NH substituent at C2 was shown to donate an additional hydrogen bond to the carbonyl group of the latter alanine residue.

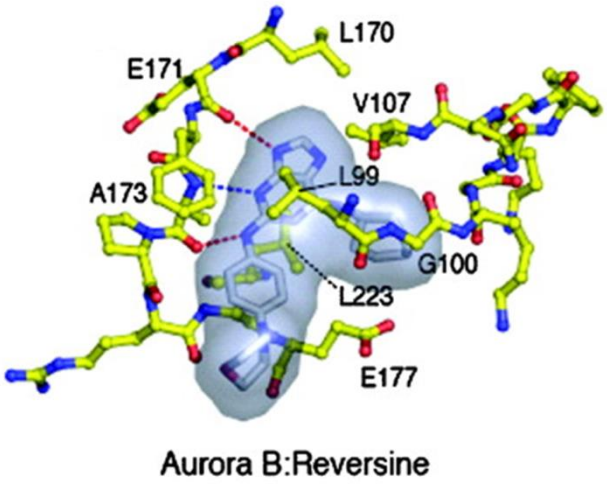


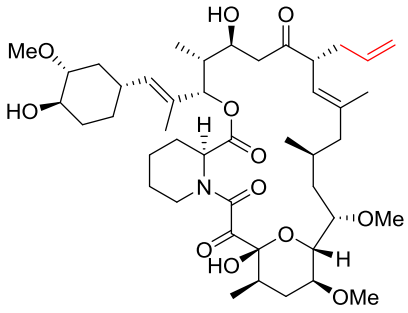
Figure III.4: Ball-and-stick representation of the interaction of reversine with selected residues in the ATP-binding pocket of *X. laevis* Aurora B (PDB code: 2VGO). A semitransparent molecular surface of reversine is shown. Hydrogen bonds are represented as dashed red/blue lines. Figure from: *Molecular Cancer Therapeutics* 2008, 7 (5), 1140-1149.

Hence, so as not to perturb this crucial H-bonding network, we selected the accessible morpholine moiety of reversine as the conjugation site, taking the liberty to exchange it for a piperazine in order to be able to attach an appropriate ligation handle. This strategy was based on earlier work by the Schultz group, successfully conjugating reversine to an agarose matrix for subsequent target identification studies by affinity chromatography.[31] Interestingly, these authors also synthesized a negative control probe comprising an inactive reversine analogue bearing an N9-methyl substituent. Indeed, later crystallographic data established the structural basis for this observed drop in activity, i.e. the presence of a substituent at N9 proved incompatible with the orientation of the purine ring within the Aurora B binding pocket since it would clash with the kinase main chain near Glu-171 (Figure III.4.).[36]

III.1.3. FK506[44]

III.1.3.1. MOA and motivation

FK506, or tacrolimus (Figure III.5.), is an immunosuppressant macrolide produced by *Streptomyces tsukubaensis*, which has found widespread use in organ transplantations as a means to lower the risk of organ rejection.[45] The cellular target of FK506 was identified as FK506 binding protein 12, FKBP12 ('12' refers to the originally discovered 12 kDa isoform), a peptidyl-prolyl *cis*-*trans* isomerase (rotamase) that catalyzes the interconversion of the *cis*- and *trans*-rotamers of the peptidyl-prolyl amide bond of peptide and protein substrates.[46] As a naturally occurring heterodimerizer (see Section I.2), FK506 is known to first form a high affinity binary complex (K_d = 0.4 nM)[47] with its principal cytoplasmic target FKBP12, which subsequently competitively binds to and concomitantly inhibits a second protein enzyme, calcineurin.[48] The latter is a heterodimeric (A/B) Ca^{2+}- and calmodulin-dependent serine/threonine phosphatase essential for T cell activation and interleukin-2 production.



FK506

Figure III.5: Chemical structure of FK506, or tacrolimus. The functional group selected as conjugation site is highlighted in red.

MASPIT profiling of FK506 was pursued to evaluate whether this system would be applicable to the analysis of natural products or biologically active compounds that are not limited in molecular size and/or complexity. Furthermore, the well-established FK506-FKBP12 ligand-receptor pair has been frequently applied in proof-of-concept studies, as exemplified by the pioneering yeast three-hybrid work by Licitra and Liu (see Section I.4).[49] Hence, analogous to this landmark report, we selected this bait to validate the MASPIT cell-array screening platform for small molecule-target protein profiling.

III.1.3.2. SAR and conjugation site

An extensive amount of structure-activity data is available regarding the binding of FK506 to its receptors, which facilitates selection of the conjugation site. FK506 has been shown to comprise two distinct protein binding surfaces, an immunophilin-binding domain (binding element) and a calcineurin-binding one (effector element) (Figure III.6A).[47]

Schreiber demonstrated that the allyl group within the calcineurin-binding domain of FK506 can be converted into a hydroxyethyl handle without significant loss in affinity for FKBP12.[50] Indeed, X-ray analysis of the FKBP12-FK506-calcineurin inhibitory complex confirmed that the terminal olefin (C39-C40) of FK506 protrudes into an interface between calcineurin A and calcineurin B,[51] making this an attractive moiety for derivatization (Figure III.6B).[52]

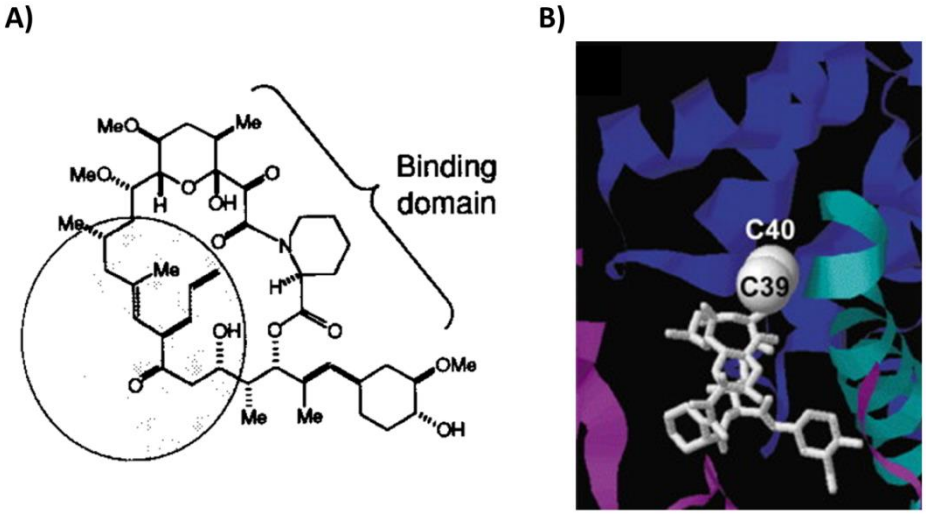


Figure III.6: A) Structure of FK506 with indication of its FKBP12-binding domain (bracket) and calcineurin-binding surface (circle). Figure from: *Science* 1991, 251 (4991), 283-287. **B) Close-up of the binding interface** of FKBP12 (magenta), FK506 (white), calcineurin A (green), and calcineurin B (blue) (PDB code: 1TCO). The terminal olefin (C39-C40) of FK506 is shown in space-filling representation. Figure from: *Chemistry & Biology* 2002, 9 (1), 49-61.

III.1.4. Simvastatin

III.1.4.1. MOA and motivation

Simvastatin is a semisynthetic derivative (bearing an additional methyl group in the acyl chain) of lovastatin, a fermentation product of the fungus *Aspergillus terreus* (Figure III.7). This hypolipidemic blockbuster drug is a member of the statin class of cholesterol-lowering agents that act through competitive inhibition of 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR) (K_i = 0.1 nM),[53] the rate-limiting enzyme in the cholesterol biosynthetic pathway. Accordingly, the levels of HMGCR's product, mevalonate, are reduced, which ultimately leads to upregulation of HMGCR, other enzymes of cholesterol biosynthesis, and most importantly induction of the hepatic low-density lipoprotein (LDL) receptor. As such, the hepatic reuptake of LDL via this receptor is increased, overall resulting in a lowering of plasma LDL cholesterol levels. Simvastatin was initially approved for marketing in Sweden in 1988 and subsequently worldwide prescribed for the primary and secondary prevention of cardiovascular diseases such as myocardial infarction, angina pectoris and other coronary events associated with elevated plasma cholesterol.[54]

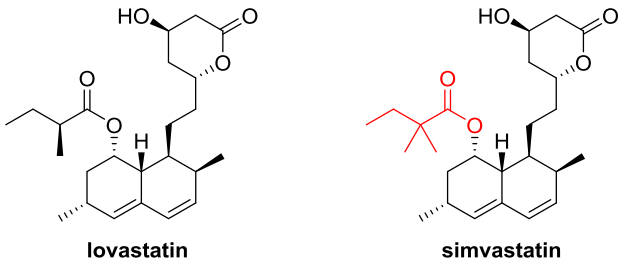


Figure III.7: Chemical structures of lovastatin and simvastatin. The functional group of simvastatin selected as conjugation site is highlighted in red.

A first motive to profile this bait in MASPIT is an attempt to map potential off-targets that may contribute to the adverse effects of the statin class of drugs. An important example of such an undesired class effect is myopathy or, in its more severe form, rhabdomyolysis.[55],[56],[57] Myopathy is characterized by a creatine kinase level greater than 10 times the upper limit of normal with accompanying symptoms including generalized muscle pain, fatigue, or weakness. Rhabdomyolysis results from a progress of this condition to severe and widespread injury to skeletal muscle and the subsequent accumulation of toxic degradation products in the blood and urine that can lead to acute renal failure. Although the typical incidence in controlled trials is less than 1%, in clinical practice statin therapy is associated with muscle problems in approximately 10-25% of treated patients.[55] For example, the risk for these adverse events increases significantly upon high dosage, long-term therapy in the elderly population or patients with renal insufficiency.[57] Moreover, the risk of statin-induced myopathy is considerably increased by drug-drug interactions resulting from concomitant administration of other myotoxic drugs (particularly the fibrate gemfibrozil)[58],[59] or agents that may increase the plasma concentration of statins by potently inhibiting or sharing key metabolic pathways (e.g. CYP 3A4, in the case of lovastatin, simvastatin and atorvastatin).[60] Most notably, in 2001, cerivastatin was voluntarily withdrawn from the global market by its manufacturer (Bayer) since the drug had been linked to a large number of fatal cases of rhabdomyolysis,[61],[62] thereby raising concerns regarding the safety of the entire statin class.[63] Remarkably, although more than 25 years have elapsed since its initial report, the molecular mechanism underlying the pathogenesis of statin-mediated muscular adverse effects remains ambiguous.[64]

In addition to their primary cholesterol lowering effect, statins have been shown to also exert cholesterol-independent or so-called pleiotropic effects, which contribute to their observed overall clinical outcome.[65],[66] By inhibiting the formation of mevalonate, statins also prevent the synthesis of important downstream isoprenoids, including farnesyl pyrophosphate (FPP) and geranylgeranyl pyrophosphate (GGPP). The latter serve as the source of the lipid moiety that is attached to proteins, such as Ras, Rho, and Rac, upon prenylation.[67] Hence, interference of statins with this post-translational modification, which controls diverse cellular events, might account for their additional effects beyond lipid lowering. For example, inhibition of geranylgeranylation of Rho proteins has been suggested as an important anticancer effect of statins,[68] and selective inhibition of GTPases has been associated with reduced amyloid-β production, thereby potentially reducing Alzheimer's disease pathogenesis.[69] On the other hand, lovastatin has been shown to function as a chemopreventive agent independently of the mevalonate pathway through inhibition of the proteasome, leading to suppression of cell proliferation and apoptosis.[70]

Apart from these biochemical effects statins have several other lipid-independent actions at the cellular and physiological level, including improvement of endothelial function, inhibition of cellular senescence, anti-inflammatory and antiatherogenic effects;[65] immunomodulatory properties;[71] bone anabolism;[72],[73] and the recently disclosed, somewhat unexpected, improvement of overall health and function of skeletal muscles in Duchenne muscular dystrophy (DMD) mice.[74] Although the involvement of certain signaling cascades, such as the Rho/Rho kinase and phosphatidylinositol-3 kinase/Akt pathway, has been suggested, the actual mechanisms by which these and other pleiotropic effects are obtained have not yet been fully clarified,[66] thereby providing the second motive for our interest in this bait.

III.1.4.2. SAR and conjugation site

Functioning as competitive inhibitors of HMGCR, all statins share a common HMG-like moiety as central pharmacophore, which mimics the natural substrate HMG-CoA.[75] In the case of the prodrug simvastatin this moiety is initially present in an inactive β-hydroxy-δ-valerolactone form, which is enzymatically hydrolyzed *in vivo* to the active ring-opened β-hydroxy-acid form (Figure III.8).[53]

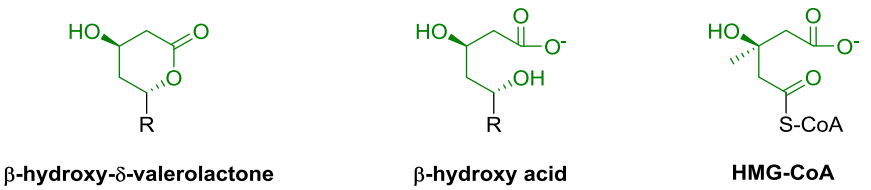


Figure III.8: Chemical structures of the inactive ('lactone') and active ('acid') form of the HMG-like moiety (green) of simvastatin as well as the natural substrate HMG-CoA. 'R' denotes the substituted decalin system of simvastatin (see Figure III.7.)

The active hydrolyzed HMG-like moiety of simvastatin occupies the HMG-binding pocket of the enzyme's active site, thereby blocking access of the substrate HMG-CoA. Accordingly, the orientation of this moiety clearly resembles that of the substrate complex and is mainly governed by an intricate network of electrostatic and hydrogen bonding interactions (Figure III.9).[75] The terminal carboxylate forms an electrostatic interaction with Lys-735 and two hydrogen bonds with Ser-684 and Lys-692. Furthermore, the β-hydroxyl group hydrogen bonds to Asp-690 and twice to Arg-590, whereas the δ-hydroxy forms hydrogen bonds with Lys-691, Asn-755, and Glu-559. Additionally, this cocrystal structure reveals that simvastatin is kinked at the δ-OH of the HMG-like moiety, with its rigid hydrophobic substituted decalin ring system being accommodated in a shallow hydrophobic groove between helices Lα1 and Lα10. In this region the hydrophobic moiety of the statin makes van der Waals contacts with the hydrophobic side chains of Leu-562, Leu-853, Ala-856, and Leu-857.

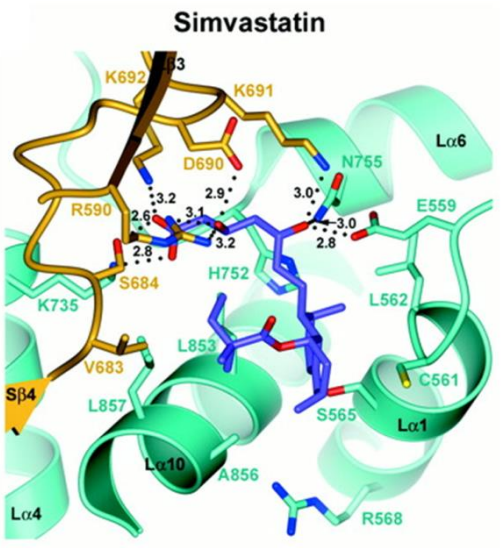
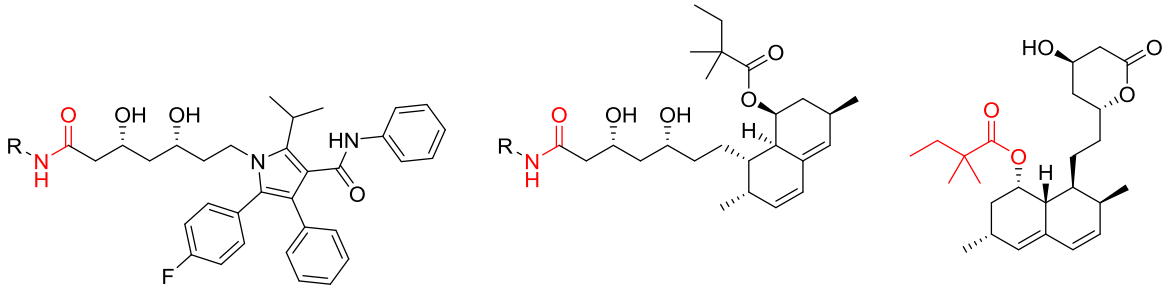


Figure III.9: Mode of binding of simvastatin (purple) to the catalytic portion of human HMGCR (PDB code: 1HW9). Selected side chains of residues that contact the statin are depicted in ball-and-stick representation. Individual electrostatic and hydrogen bonds are represented as dotted black lines; the distance of each bond is indicated in Å. Figure from: *Science* 2001, 292 (5519), 1160-1164.

Consequently, we assumed that the acyl side chain would provide a suitable site for derivatization without significant interference with the ensemble of bonding interactions outlined above. Indeed, early SAR studies on statin-like analogues by Stokker *et al.* already indicated the necessity of the terminal carboxylate in the ring-opened HMG-like moiety, as its replacement with a carboxamido group ablated activity.[76] Likewise, Chidley *et al.* recognized that derivatization of atorvastatin at its carboxyl group interferes with binding to HMGCR. Nevertheless, the authors did prepare an O^6-benzylguanine-atorvastation conjugate in such a fashion, speculating that this derivative could allow the identification of off-targets using a SNAP-tag-based Y3H assay (Figure III.10).[77] Recently, the importance of the free carboxylate was once more confirmed by Hsieh *et al.*[73] This study described an anabolic polyaspartate-simvastatin analogue, which was constructed via amidation of the latter carboxyl, thereby displaying a more than 1000-fold drop in inhibitory activity against HMGCR (IC_{50} > 30 µM) compared to the parent compound (IC_{50} = 11 nM)[75] (Figure III.10).[73]



atorvastatin conjugate
Chidley *et al.*, 2011

simvastatin conjugate
Hsieh *et al.*, 2014

alternative simvastatin conjugation site
this work

Figure III.10: Comparison of different conjugation sites for the HMGCR inhibitors atorvastatin and simvastatin.

In conclusion, since our primary aim was to uncover protein targets that may contribute to some of the obscure beneficial or detrimental effects of the statin class of drugs (rather than elucidating compound-specific effects), we decided to keep the central HMG-like pharmacophore intact and to focus our ligation efforts on the acyl side chain (Figure III.10). Remarkably, to our knowledge, no prior literature data regarding the conjugation of simvastatin at this specific site has been published.

III.2. Introduction

As we move toward systems biology and personalized medicine, it will become increasingly important to profile small molecule-target interactions and to map this information with metabolic and signaling pathways. Indeed, many clinically used drugs have been found to be more promiscuous than originally thought. However, modulation of multiple targets can also cause harmful side effects, and another considerable challenge is to uncover the mechanisms of toxicities that are not directly related to the desired pharmacological effects of drugs ('off-target pharmacology'). As classical *in vitro* target profiling requires time- and budget-consuming expression, purification, and assay setup for each individual target, it usually involves testing of a compound against a limited panel of related targets and is thus not comprehensive.

The number of 'tried-and-true' drug targets is quite small.[78],[79] The emergence of molecular biology and the completion of the human genome project have hitherto failed to produce the expected flood of compounds acting on new targets. Unbiased, phenotype-based screens represent a promising approach to uncover drugs with a novel mechanism of action. For small molecules discovered in such screens, identifying the biological targets remains largely an *ad hoc* affair. Traditional approaches using affinity pull-down reagents[80] have been successful for the identification of new targets and have, for example, been recently employed to uncover targets involved in the teratogenic effects of thalidomide.[81] However, sensitivity can be limited, particularly for compounds that exhibit low binding affinity toward their target or for targets expressed at low levels. In these cases, the target protein may be lost during the washing steps, or its binding is obscured by the presence of highly abundant (non-specifically binding) proteins.[82] A systematic, widely applicable, and robust approach is badly needed.

MASPIT (mammalian small molecule-protein interaction trap) is a three-hybrid trap variant of the original MAPPIT concept[83],[84] for the detection of small molecule-protein interactions. MASPIT makes use of a signaling-deficient cytokine receptor lacking STAT3 recruitment sites, which is fused to dihydrofolate reductase (DHFR) (Figure III.11). Fusion compounds consisting of an organic molecule of interest tethered to methotrexate (MTX) bind DHFR with very high affinity, allowing presentation of the organic molecule as 'bait'. Binding of a chimeric 'prey' protein containing functional STAT3 binding sites on the MTX fusion compound (MFC) complements the STAT3 signaling cascade. Hence, ligand binding to the receptor will lead to activation of receptor-associated JAK2 kinases, followed by tyrosine phosphorylation of the STAT3 recruitment motifs of the prey chimeras. Subsequent binding and activation of STAT3 is then easily measured using a STAT3-responsive reporter gene.

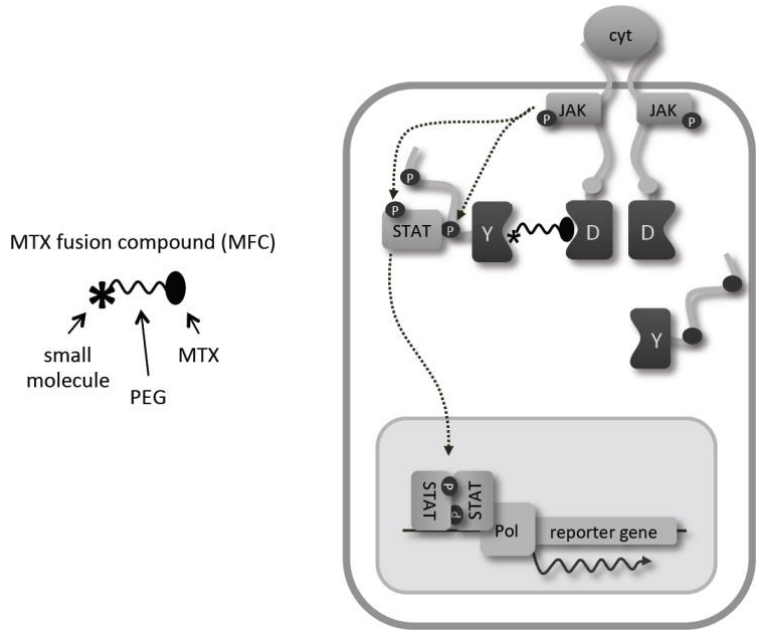


Figure III.11: Outline of the MASPIT system. Mammalian cells express a signaling-deficient cytokine receptor containing mutated STAT3 recruitment sites (grey dots), which is fused to DHFR (D). Upon addition of an MFC, which is readily taken up by the cells, the MTX moiety binds to DHFR with high affinity, resulting in the small organic molecule being displayed as bait. A second hybrid polypeptide expressed in the cells consists of a prey protein (Y) coupled to a gp130 cytokine receptor fragment that contains functional STAT3 docking sites (black dots). Physical interaction between the bait small molecule and the prey protein brings the cytokine receptor fragments into close proximity, reconstituting a functional cytokine receptor system. When these cells are stimulated with the appropriate cytokine ligand (cyt), constitutively associated JAK2 kinases are activated, leading to phosphorylation of tyrosine molecules in the gp130 moiety (P) and resulting in the recruitment of STAT3 transcription factors. Subsequently, these STATs are activated through phosphorylation (P) by the activated JAKs. Finally, activated STAT3 complexes migrate to the nucleus as dimers, where they induce expression of a STAT3-dependent reporter gene.

MASPIT can be used both analytically, to study designated small molecule-protein interactions, and in searches for interaction partners. Since 2006, our research group has been involved in a large-scale human interactome mapping program.[85],[86] As a consequence, a large portion of the human ORFeome is being transferred into MASPIT prey vectors, currently encompassing more than 12 000 ORFs.[87] To optimize screening, a cellular array screening platform was developed.[88] In brief, each prey plasmid from the collection, together with a luciferase reporter construct, was mixed with a transfection reagent to generate prey arrays in 384-well plates. After reverse transfection with a cell pool expressing the receptor-DHFR chimera and addition of the bait MFC, followed by ligand-induced activation of the system, positive interactors were detected simply by measuring the activity of a STAT3-dependent luciferase reporter gene.

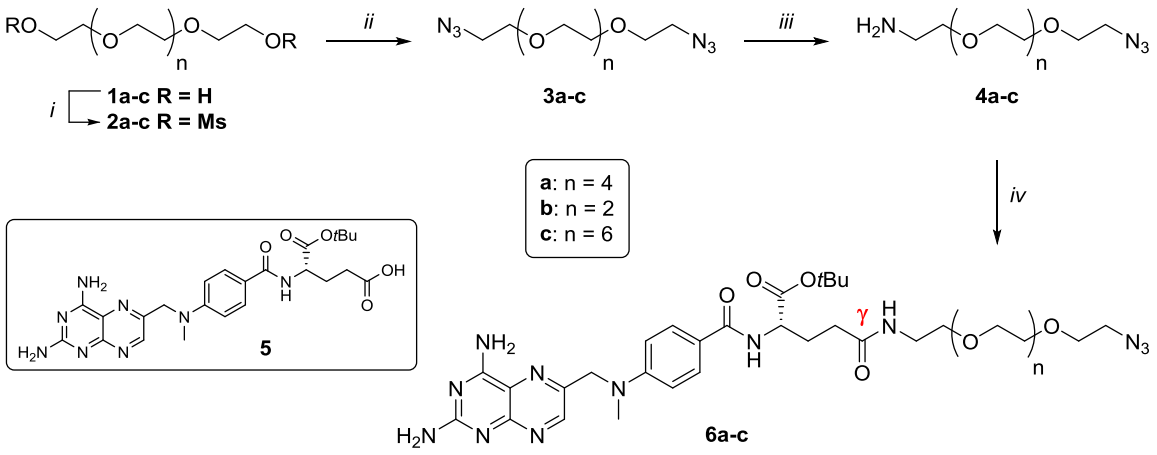
In contrast to classical target-based profiling, this mammalian three-hybrid system can provide information regarding unanticipated small molecule-target protein interactions. Another important advantage of MASPIT is the fact that the interactions between small molecules and target proteins occur in living mammalian cells rather than *in vitro*. Consequently, this might reveal potential effects of post-translational modifications of the target or of the target's association with additional proteins or other intracellular molecules on small molecule binding.

A necessary component of successful MASPIT applications, however, is the synthesis of appropriate MFCs. Structure-activity relationship (SAR) studies of MTX derivatives have emphasized the importance of selective conjugation to the γ-carboxylic acid of the glutamate moiety to ensure high affinity binding to DHFR through interaction of this enzyme with the free α-carboxylic acid.[89] Hence, an objective of this study was to synthesize a versatile MTX-based building block and explore its use for easy and straightforward ligation to bait small molecules of interest.

III.3. Results and Discussion

To swiftly access a wide variety of MFCs with minimal effort, we envisaged the synthesis of a general MTX conjugate appropriately equipped with a ligation handle. It was estimated that a copper-catalyzed 3+2 azide alkyne cycloaddition (CuAAC)[90] would be a suitable method for the attachment of the small molecule baits to the MTX-linker conjugates, given the high chemoselectivity, mild reaction conditions, and high tolerance for a wide diversity of reaction solvents.

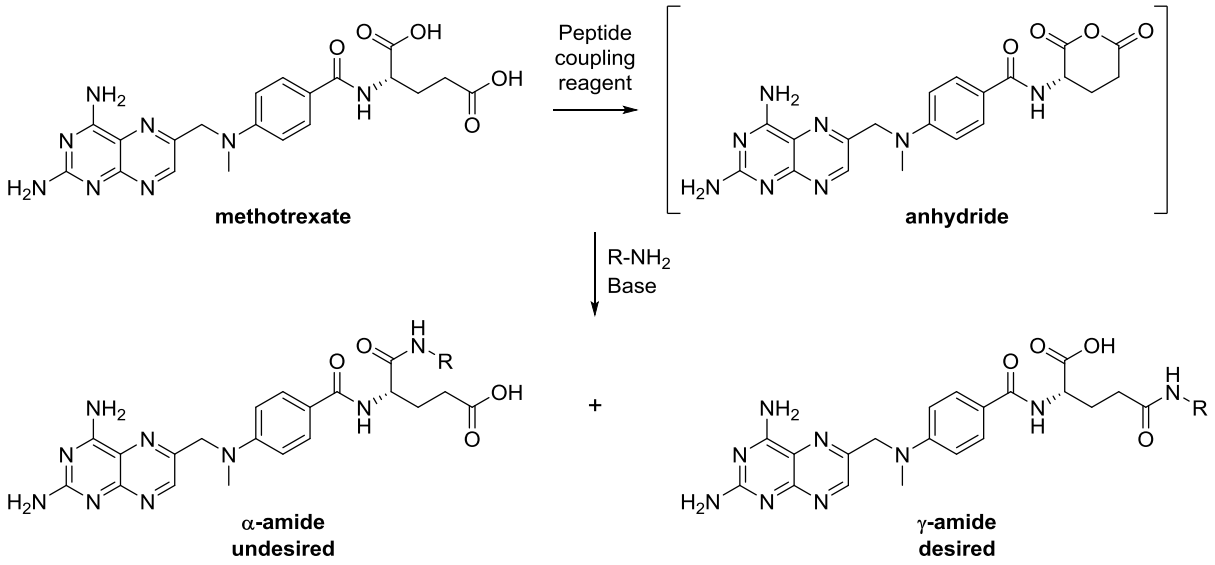
A terminal azido group was selected as a ligation handle. To discourage steric hindrance of fusion partners, which could cause the MFC to bind suboptimally to DHFR or allow one to overlook targets that might interact with the unconjugated bait, a poly(ethylene glycol) (PEG) linker was introduced between the γ-carboxylic acid and the ligation handle to allow optimal interaction with prey chimeras (Scheme III.1). The synthesis was commenced with the generation of an amino/azido bifunctionalized PEG linker. To this end poly(ethyleneglycols) were treated with methanesulfonyl chloride. Treatment of the formed bismesylates **2** with sodium azide yielded diazides **3**. Desymmetrization of these compounds towards the corresponding azidoamines **4** was carried out in a biphasic system using triphenylphosphine and dilute hydrochloric acid.[91]



Scheme III.1: Regioselective synthetic approach to a general MTX reagent with a terminal azido ligation handle. [i] MsCl, Et_3N, CH_2Cl_2, 0°C; [ii] NaN_3, DMF, 60°C, (**a** 92%; **b** 98%; **c** 93% two steps); [iii] first PPh_3, 2.0M HCl, toluene; then NaOH, (**a** 81%; **b** 88%; **c** 89%); [iv] **5**, TPTU, Et_3N, DMF, (**a** 80%; **b** 49%; **c** 42%).

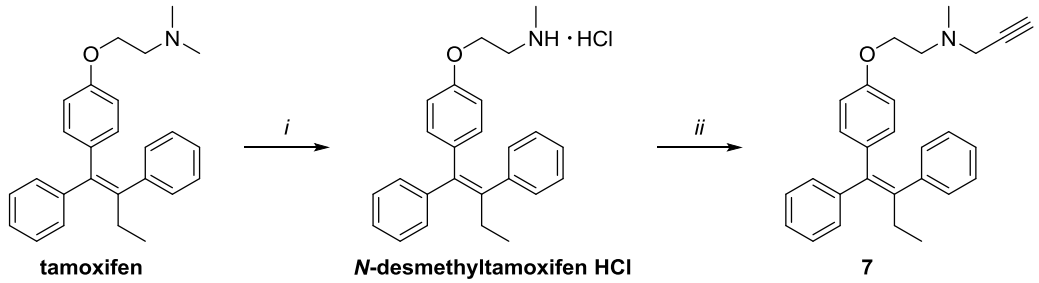
With the azido-PEG spacers in hand, attention was focused to the γ-selective condensation of MTX to **4**. Initial condensations using EDC or other standard peptide coupling reagents in DMF consistently provided an intractable mixture of the α- and the γ-amide. Variation of temperature, coupling reagent and solvent had negligible effect on the product distribution. This can be explained by the

initial formation of an acid anhydride intermediate in the glutamoyl moiety of MTX, which is subsequently opened by the amine of **4** (Scheme III.2). To circumvent the formation of this regioisomeric mixture, we condensed α-*tert*-butylmethotrexate **5**[92] with the PEG azidoamines using TPTU reagent to obtain satisfactory yields (Scheme III.1). Three different MTX-azido reagents (**6a-c**) were thus synthesized which differed with regard to the number of PEG units.

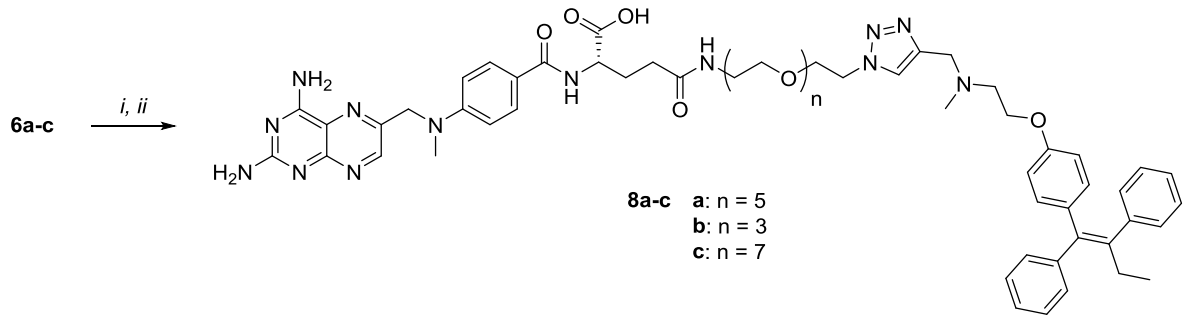


Scheme III.2: Formation of a regioisomeric mixture of methotrexylamides via an anhydride intermediate.

The utility of these MTX-based ligation reagents to form MFCs was demonstrated for three structurally diverse baits of interest, beginning with tamoxifen. The desired tamoxifen-MTX fusion compounds **8a-c** were prepared via click reaction with *N*-desmethyl-*N*-propargyltamoxifen (**7**), which was obtained by propargylation of *N*-desmethyltamoxifen hydrochloride[25] under alkaline conditions (Scheme III.3, III.4).

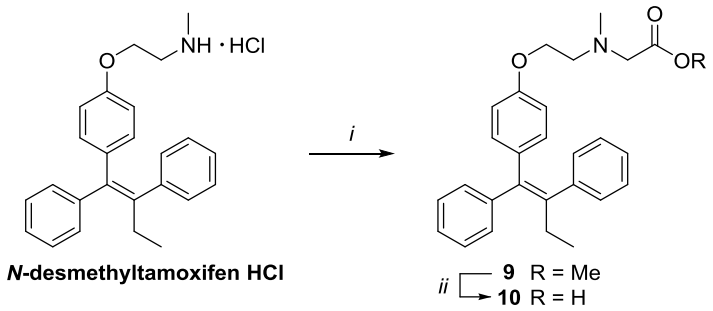


Scheme III.3: Synthesis of alkyne-functionalized tamoxifen. [i] first ACE-Cl, DCE, Δ; then MeOH, Δ, 94%; [ii] propargyl bromide, K_2CO_3, MeCN, 60°C, 57%.



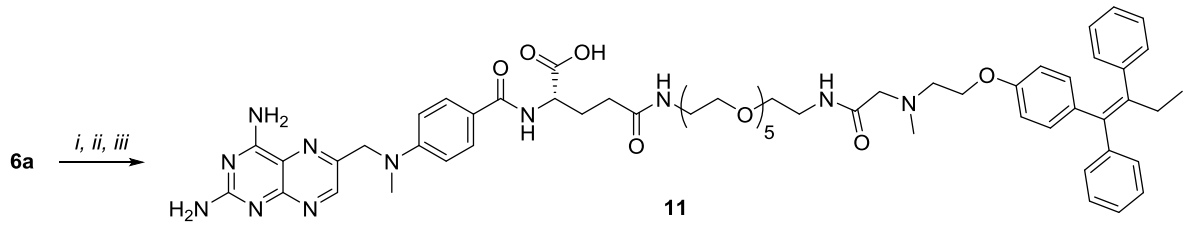
Scheme III.4: Synthesis of click-coupled tamoxifen-MFCs. [i] TFA, CH_2Cl_2; [ii] **7**, $CuSO_4$, Na ascorbate, TBTA, H_2O/*t*BuOH, 80°C, (**a** 50%; **b** 52%; **c** 63% two steps).

To investigate the possible influence of the triazole ring on the MASPIT signal, we also prepared an amide-coupled MFC. Toward this end, the required carboxylate functionalized analogue of tamoxifen was generated in two steps from *N*-desmethyltamoxifen hydrochloride. Alkylation with methyl bromoacetate followed by saponification of the ester with concomitant neutralization provided the acid (**10**) (Scheme III.5).



Scheme III.5: Synthesis of carboxylate-functionalized tamoxifen. [i] Methyl bromoacetate, DIPEA, CH_2Cl_2, 0°C, 91%; [ii] first NaOH, MeOH, 50°C; then HCl, ± quant.

With the acid in hand, the MFC was constructed. Methotrexate derivative **6a** was deprotected with trifluoroacetic acid and subsequently reduced under Staudinger conditions. Purification of the crude amine followed by PyBOP-mediated condensation to acid **10** yielded MFC **11** in 33% after HPLC purification (Scheme III.6).



Scheme III.6: Synthesis of amide-coupled tamoxifen-MFC. [i] TFA, CH_2Cl_2; [ii] Me_3P, THF, H_2O; [iii] **10**, PyBOP, DIPEA, DMF, 33%.

To evaluate tamoxifen-MTX conjugates **8a-c** and **11** in MASPIT, estrogen receptor alpha (ER1), the established primary target of tamoxifen, was selected as a prey protein. Hence, HEK293T cells were transiently transfected with receptor-DHFR chimera, ER1 prey constructs, and a STAT3-dependent luciferase reporter gene. Reporter activity was shown to be dependent on MFC concentration, with a similar pattern for all evaluated tamoxifen MFCs and a maximal signal within the 0.1-1 µM range (Figure III.12). Although the optimal spacer length may be determined by the nature of the prey, we decided to use a PEG_6 linker for the synthesis of two additional MFCs.

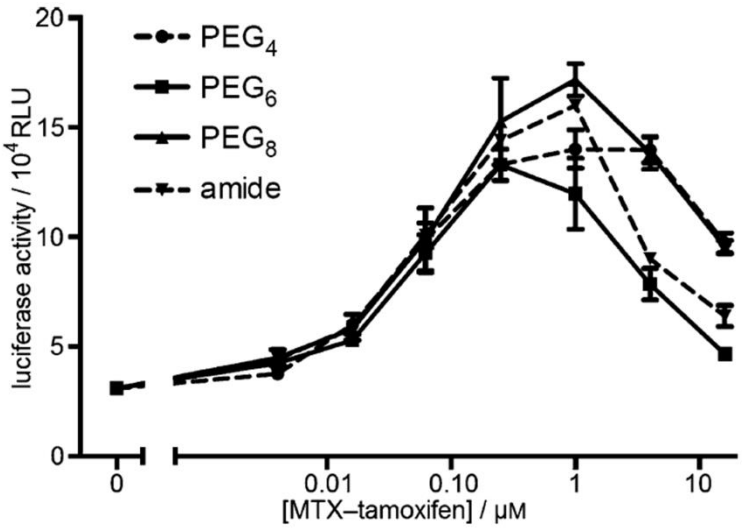
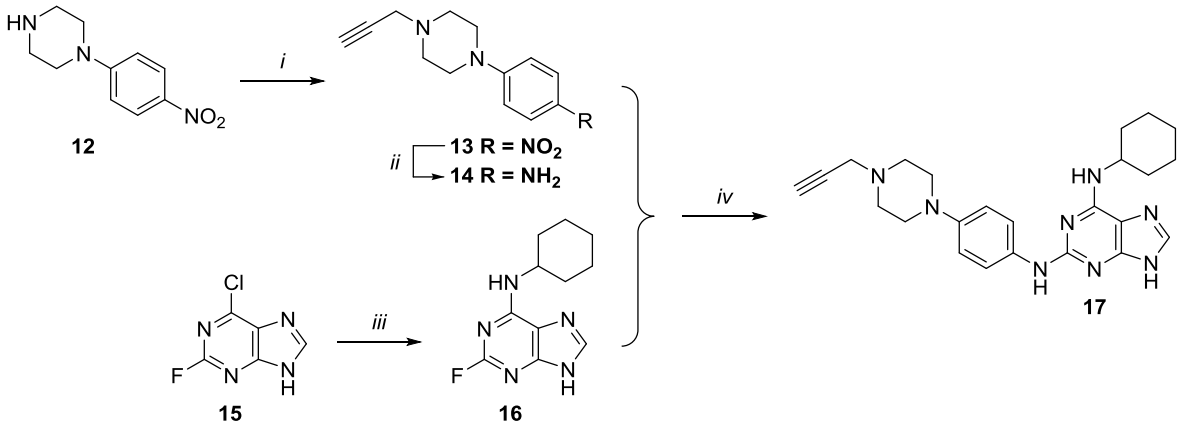


Figure III.12: Comparison of different tamoxifen MFCs. Cells were transiently transfected with a pCLG-eDHFR receptor-DHFR plasmid, an ER1 prey construct, and a luciferase reporter plasmid. They were then treated with combinations of leptin and the indicated concentration of either of the different MTX-tamoxifen fusion compounds: click-coupled through a tetra- (PEG_4), hexa- (PEG_6), or octa(ethylene glycol) linker (PEG_8), or amide-coupled through a hexa(ethylene glycol) linker (amide). The graph shows average luciferase activity of triplicate samples. Error bars represent the standard deviation.

Reversine, a small molecule found to promote dedifferentiation of committed cells into multipotent progenitor-type cells,[28] was selected as a second bait of interest. *In vitro* inhibition assays on a battery of human mitotic kinases recently indicated that TTK (also known as MPS1) acts as a primary target kinase for reversine (IC_{50} = 2.8 nM).[37] A click-coupled MFC with reversine (**18**) was obtained by CuAAC between the hydrolyzed **6a** and the alkynylated reversine derivative **17** (Scheme III.7). To obtain the latter analogue, first *N*-(4-nitrophenyl)piperazine was propargylated under alkaline conditions. Tin(II)-mediated reduction[93] of the nitro moiety gave access to aniline **14**, which could be reacted with N^6-cyclohexyl-2-fluoroadenine[28] at elevated temperature under microwave irradiation to obtain alkyne-functionalized reversine **17** as a solid.



Scheme III.7: Synthesis of alkyne-functionalized reversine. [i] propargyl bromide, K_2CO_3, MeCN, 60°C, 95%; [ii] $SnCl_2$, EtOH, Δ, 87%; [iii] cyclohexylamine, DIPEA, *n*BuOH, 80°C, 64%; [iv] EtOH, 130°C (μW), 36%.

Using TTK as a prey plasmid, stimulation with a combination of leptin and MFC **18** gave maximal luciferase activity at a MFC concentration of ~5 μM (Figure III.13).

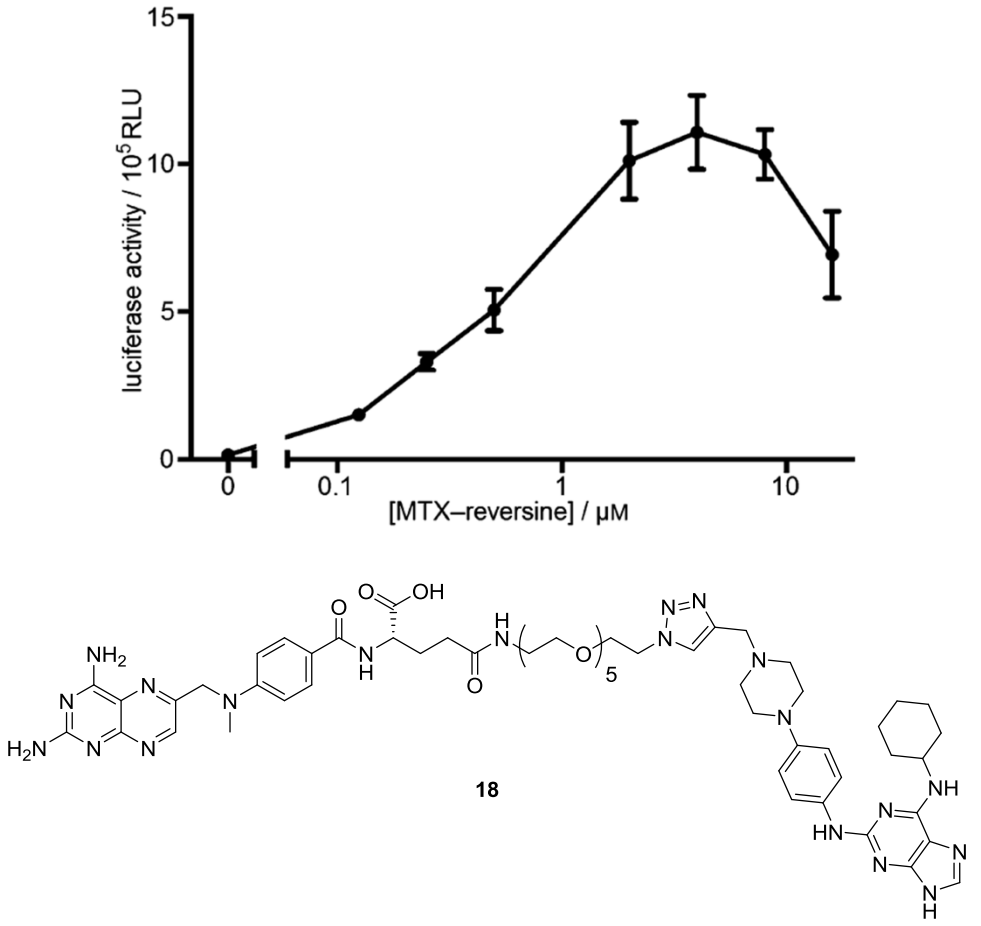
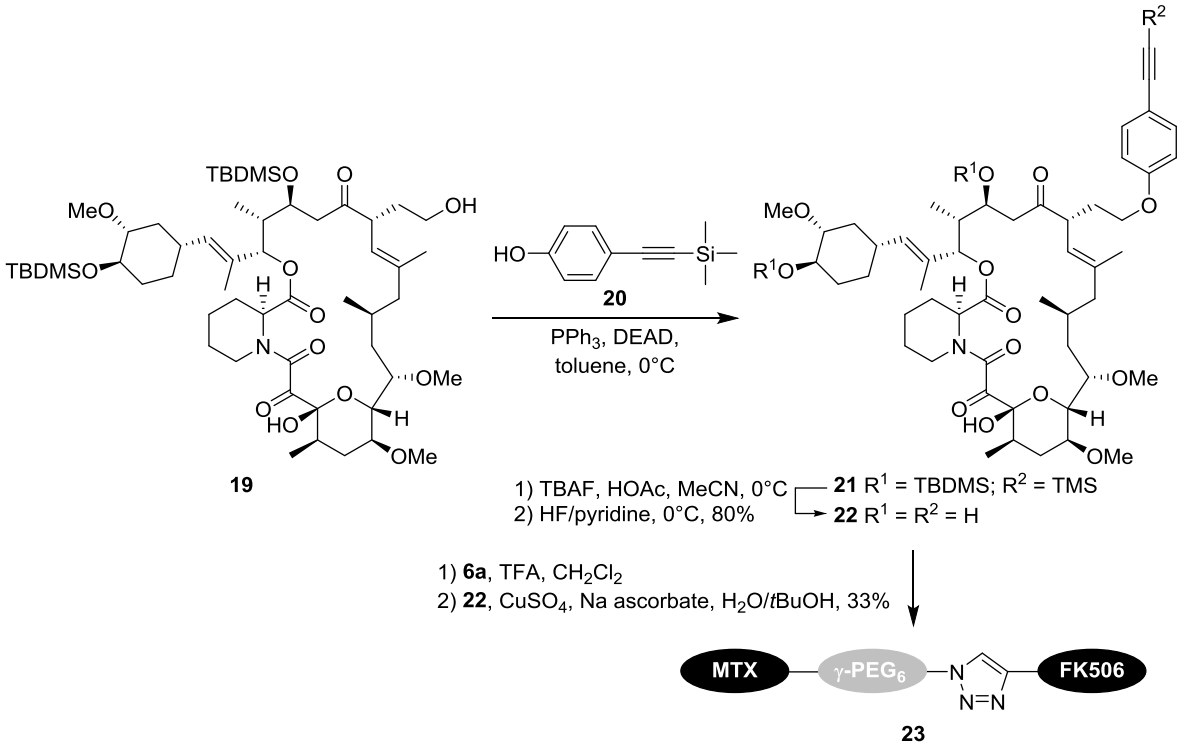


Figure III.13: Evaluation of reversine MFC. Cells transfected with a pCLL-eDHFR receptor-DHFR construct, a luciferase reporter plasmid, and a TTK prey plasmid were stimulated with a combination of leptin and the indicated concentration of MFC. The graph shows average luciferase activity of triplicate samples. Error bars represent the standard deviation.

Tacrolimus (FK506), an immunosuppressant macrolide produced by *Streptomyces tsukubaensis*, was selected as a third model bait. FK506 has found widespread use in organ transplantations as a means to lower the risk of organ rejection. The cellular target of FK506 was identified as peptidyl-prolyl *cis-trans* isomerase FKBP12. Binding of FK506 to FKBP12 inhibits calcineurin, a protein phosphatase essential for T cell activation and interleukin expression.[48] FK506 is known to bind to its principal intracellular target, FKBP12, with high affinity,[47] and an extensive amount of structure-activity data is available regarding the binding of FK506 to its receptors, which facilitates selection of the attachment site. FK506 has been modified to create affinity reagents for the isolation and identification of its receptors.[46]

Successful MASPIT profiling of FK506, a highly complex natural product, was pursued to provide clear-cut proof that this system is not confined to evaluation of biologically active compounds that are limited in molecular size and/or complexity. We additionally used this bait to show that selective conjugation of MTX to FK506 via the γ-carboxylic acid offers advantages in readout sensitivity in comparison with the non-regiomeric conjugate mixture previously used. Our final goal was to demonstrate the feasibility of FK506 MFC in identifying protein targets of small molecules using MASPIT in the cellular array assay by screening for proteins that bind to FK506.

In order to attach FK506 to the azido-functionalized MTX to form the desired MFC, a terminal acetylene had to be grafted onto the macrolide (Scheme III.7). The attachment position of this acetylene was meticulously chosen so as to minimize the loss in affinity for FKBP12. Schreiber demonstrated that the allyl moiety of FK506 can be converted into a hydroxyethyl handle (as with compound **19**) without significant loss of activity.[50] Simple alkylation of the primary hydroxy with propargyl bromide under alkaline conditions proceeded with poor regioselectivity, presumably due to the lower pKa of the hydroxyl group of the ketal functionality. Significantly better regioselectivity was achieved upon reaction of compound **19** with TMS-ethynylphenol[94] under Mitsunobu conditions.[95] Subsequent treatment with TBAF/HOAc and hydrofluoric acid[96] cleanly removed all silyl protecting groups to give acetylene **22**. After removal of the remaining *tert*-butylester of MTX-azide **6a**, CuAAC with **22** afforded the desired MFC (**23**).



Scheme III.7: Synthesis of alkyne-functionalized FK506 and its corresponding MFC.

Evaluation of MTX-FK506 conjugate **23** in MASPIT showed that reporter activity was induced only in cells that were treated with both the bait MFC and the cytokine ligand that activates the assay (Figure III.14A). No luciferase activity was measured in cells transfected with a combination of the receptor-DHFR chimera with an empty control prey construct. Co-transfection of the receptor-DHFR chimera with a positive control prey that binds to the receptor chimera itself (EFHA1) resulted in bait MFC-independent reporter gene induction.

Next, performance of the γ-substituted MTX-FK506 conjugate was compared with that of a non-regiomeric conjugate mixture.[97] HEK293T cells expressing receptor-DHFR chimera and FKBP12 prey were treated with the cytokine ligand and a concentration gradient of either of the two bait MFCs (Figure III.14B). Clearly, stronger signals were obtained with the regioselective γ-substituted MTX-FK506 fusion compound. This observation was anticipated, as the α-substituted fusion compound, which constituted roughly half of the non-regiomeric conjugate mixture, inhibits formation of the three-hybrid complex necessary for restoration of the functional MASPIT receptor complex, due to its inability to bind to DHFR.

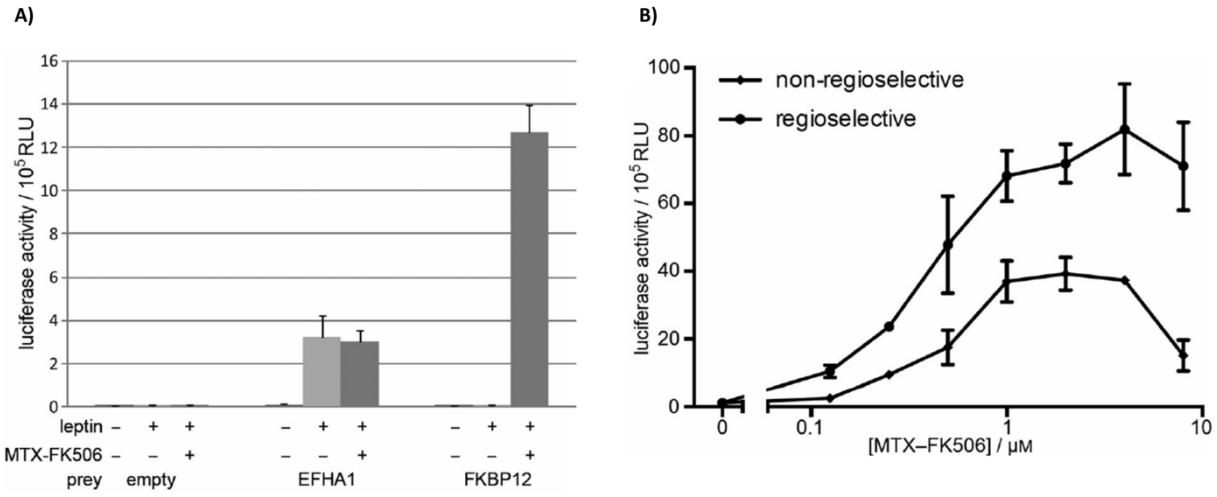


Figure III.14: A) Evaluation of the MTX-FK506 conjugate 23 in MASPIT. Cells transfected with a pCLL-eDHFR receptor-DHFR construct, a luciferase reporter plasmid, and either an empty, EFHA1, or FKBP12 prey construct were treated with leptin and/or 1 µM MTX-FK506. Luciferase activity was expressed as relative light units (RLU) and calculated as the average signal of triplicate samples. Error bars represent the standard deviation. **B) Comparison of the regioselective and a non-regiomeric MTX-FK506 conjugate mixture.** Cells transfected with a receptor-DHFR construct (pCLL-eDHFR), FKBP12 prey plasmid, and a luciferase reporter plasmid were treated with leptin and the indicated concentration of either MTX-FK506 conjugate. The graph shows the average luciferase activity of triplicate samples. Error bars represent the standard deviation.

Having confirmed the functionality of the MTX-FK506 fusion compounds in MASPIT, we next evaluated whether the regioselective MFC **23** could be applied in a cellular array screen for targets of FK506. A pilot collection of nearly 2000 full-length human ORF preys, spotted as transfection mixtures in 384-well microtiter plates,[88] was reverse-transfected with a pool of HEK293T cells transiently transfected with the receptor-DHFR plasmid. Duplicate wells were treated with either MTX-FK506 alone or in combination with the cytokine ligand. The results are shown as a dot plot of normalized luciferase readings for both treatments (Figure III.15). Applying a cutoff of tenfold induced luciferase activity for MTX-FK506/ligand-treated over MTX-FK506-treated, the only prey that scored positive corresponds to FKBP12. Importantly, no other previously reported FK506 target proteins were present in the screened collection.

This arrayed screening approach nicely complements the MASPIT cDNA library screening protocol[98] that has previously been used to search for targets of the kinase inhibitor PD173955. The complementarity of the latter assay lies mainly in the fact that, in contrast to the full-length ORF collection screened by the cellular array assay, a cDNA library also contains partial ORFs encoding protein fragments or domains. Interacting sub-modules in proteins, when isolated from regulatory domains, can allow an interaction to be identified that does not occur in the presence of the regulatory domains. In addition, a cDNA library generally covers a larger portion of the proteome, including multiple protein isoforms for many genes. However, the increased complexity of cDNA libraries, along with the fact that such collections are pooled and not arrayed, makes the screening process much more complicated and time-consuming.

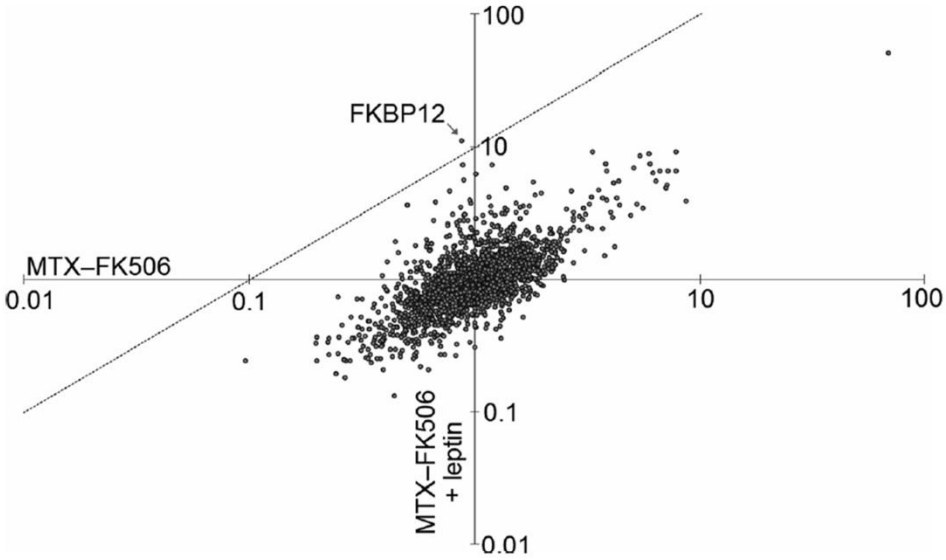


Figure III.15: MASPIT cellular array screen with the MTX-FK506 conjugate 23. An array containing 1879 distinct preys was reverse transfected with cells transfected with a receptor-DHFR plasmid (pCLL-eDHFR), and duplicate wells were treated with either 1 μM MTX-FK506 alone or with 1 μM MTX-FK506 combined with leptin. The dot plot shows normalized average luciferase values for each prey. The dashed line indicates the threshold of tenfold induced luciferase activity for MTX-FK506- and leptin-treated over MTX-FK506-treated values. The position of the FKBP12 prey in the plot is indicated.

III.4. Conclusions

In conclusion, we have presented a scalable synthesis of a versatile MTX reagent that allows for the rapid synthesis of MFCs compatible with MASPIT from any acetylene-functionalized compound using 'click chemistry'. The conjugation methodology, however, is not limited to click chemistry but is also applicable for Staudinger-type ligations or standard peptide coupling conditions. This allows easy and fast access to various MFCs, thereby minimizing the number of chemical manipulations for each construct. The results presented here clearly demonstrate the versatility of the new MTX reagent to generate an MFC of interest for use in MASPIT. Furthermore, we demonstrated the clear benefit of γ-selective functionalization of methotrexate with respect to the signal output. In a cellular array screen, FKBP12 was selectively identified as an interaction partner of FK506, thereby validating the MASPIT system and showing its potential for the identification of the molecular targets responsible for the beneficial or detrimental effects of small molecule drugs.

III.5. Experimental Section

III.5.1. Synthesis

*A thorough NMR analysis of compounds **6a**, **8a**, **11**, **18** and **22** has been carried out by the NMR and Structure Analysis Unit, Department of Organic Chemistry, Faculty of Sciences, Ghent University (Prof. José Martins, Dr. Davy Sinnaeve, Dr. Freya Van den Broeck).*

General: All reactions were performed under nitrogen and at ambient temperature, unless stated otherwise. FK506 (Tacrolimus) was purchased from LC laboratories. All other reagents and solvents were purchased from Sigma-Aldrich, Acros Organics, or TCI Europe, and used as received. Reactions were monitored by thin-layer chromatography on TLC aluminum sheets (Macherey-Nagel, Alugram Sil G/UV_{254}) with detection by spraying with a solution of $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$ (25 g/L) and $(NH_4)_4Ce(SO_4)_4{\cdot}2H_2O$ (10 g/L) in H_2SO_4 (10%) followed by charring or Ninhydrin (2% in ethanol). Column chromatography was performed with 60 Å silica gel (Biosolve, 32-63 µm). LC-MS analyses were carried out on a Waters Alliance 2695 XE separation Module by using a Phenomenex Luna reversed-phase C18 column (100×2.00 mm, 3 µm) and a gradient system of HCOOH in H_2O (0.1 %, *v/v*)/HCOOH in CH_3CN (0.1 %, *v/v*) at a flow rate of 0.4 mL/min. High-resolution spectra were recorded on a Waters LCT Premier XE Mass spectrometer. ^1H and ^{13}C NMR spectra were measured on a Varian Mercury-300BB (300/75 MHz) spectrometer, an Avance III Bruker (500/125 MHz) spectrometer equipped with a BBI probe, or an Avance II Bruker (700/176 MHz) spectrometer equipped with a $^1H/^{13}C/^{15}N$ TXI-Z probe. Chemical shifts (δ) are given in ppm relative to tetramethylsilane (^1H NMR) or $CDCl_3$, CD_3OD or $SO(CD_3)_2$ (^{13}C NMR) as internal standards. Coupling constants are given in Hz. Preparative HPLC purifications were carried out by using a Laprep preparative RP-HPLC system equipped with a Xbridge Prep C18 column (19×250 mm, 5 µm) or a Phenomenex Luna C18 column (21.20×250 mm, 5 µm) with a gradient system of HCOOH in H_2O (0.2 %, *v/v*)/CH_3CN at a flow rate of 17.5 mL/min. Microwave experiments were performed using a Milestone Microsynth under fiberoptic internal temperature control.

α,ω-Diazido,α,ω-dideoxy-hexa(ethylene glycol) (3a): To an ice cooled solution of hexa(ethylene glycol) (5.0 g, 17.7 mmol) in anhydrous CH_2Cl_2 (30 mL), triethylamine (7.7 mL, 55 mmol) and MsCl (3.1 mL, 40 mmol) were added. The resulting solution was stirred overnight letting the temperature rise to RT. The reaction mixture was subsequently poured in 0.5N HCl (100 mL) and after separation of both phases, the aqueous layer was washed with CH_2Cl_2 (3 x 50 mL). The combined organic phases were next washed with sat. $NaHCO_3$ (50 mL) and dried over Na_2SO_4. After filtration and removal of all volatiles under reduced pressure, the crude bismesylate was taken up in DMF (200 mL). Sodium azide (4.0 g, 61.5 mmol) was added and the reaction mixture was stirred overnight at 60°C. The reaction mixture was concentrated *in vacuo.* To the residue were added water (100 mL) and EtOAc (100 mL) and the biphasic mixture was agitated until the residue was completely dissolved and transferred to a separation funnel. The organic layer was separated and the aqueous fraction was washed repeatedly with EtOAc (3 x 100 mL). All EtOAc fractions were pooled, dried on Na_2SO_4 and taken to dryness. The residue was purified by silica gel chromatography (hexane/EtOAc, 1:3 v/v) yielding the title compound (5.4 g, 16.2 mmol, 92%) as a colorless oil. ^1H NMR (300 MHz, $CDCl_3$) δ = 3.70-3.65 (m, 20H), 3.38 (t, 4H, *J* = 5.0 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 70.4-70.3 (wide peak), 69.7, 50.4; HRMS: calcd. for $C_{12}H_{24}N_6O_5Na$ $[M+Na]^+$: 355.1700, found: 355.1709.

α,ω-Diazido,α,ω-dideoxy-tetra(ethylene glycol) (3b): Tetra(ethylene glycol) (4.9 g, 25.0 mmol) was transformed into the corresponding diazide using the procedure described for compound **3a**. Silica gel chromatography (toluene/EtOAc, 1:1 v/v) yielded the title compound (6.0 g, 24.4 mmol, 98%) as a slightly yellow liquid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 3.70-3.67 (m, 12H), 3.39 (t, 4H, *J* = 5.0 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 70.8 (wide peak), 70.1, 50.7; HRMS: calcd. for $C_8H_{16}N_6O_3Na$ $[M+Na]^+$: 267.1176, found: 267.1177.

α,ω-Diazido,α,ω-dideoxy-octa(ethylene glycol) (3c): Octa(ethylene glycol) (904 mg, 2.4 mmol) was transformed into the corresponding diazide using the procedure described for compound **3a**. Silica gel chromatography (toluene/EtOAc, 15:85 v/v) yielded the title compound (950 mg, 2.3 mmol, 93%) as a colorless liquid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 3.70-3.65 (m, 28H), 3.39 (t, 4H, *J* = 5.0 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 70.7-70.6 (wide peak), 70.1, 50.7; HRMS: calcd. for $C_{16}H_{32}N_6O_7K$ $[M+K]^+$: 459.1964, found: 459.1933.

α-Amino,ω-azido,α,ω-dideoxy-hexa(ethylene glycol) (4a): A solution of **3a** (5.4 g, 16.2 mmol) in toluene (130 mL) was treated with 2.0M HCl (130 mL) followed by addition of a solution of PPh_3 (4.28 g, 16.3 mmol, 1 eq.) in toluene (15 mL) and the resulting biphasic mixture was vigorously stirred overnight at RT. After separation of both phases, the aqueous layer was washed with toluene (3 x 100 mL). The aqueous layer was next cooled to 0°C and the pH was adjusted to pH 10 by slow addition of 3N NaOH. The alkaline solution was concentrated *in vacuo* and coevaporated once with toluene. The semisolid residue was extracted with EtOAc (3 x 150 mL). After drying of the combined organic layers over Na_2SO_4, the residue was purified by column chromatography (CH_2Cl_2/MeOH, 4:1 v/v with 0.25% NH_4OH) affording the title compound (4.2 g, 13.2 mmol, 81%) as a colorless oil. ^{1}H NMR (300 MHz, CD_3OD) δ = 3.68-3.61 (m, 18H), 3.53 (t, 2H, *J* = 5.4 Hz), 3.37 (t, 2H, *J* = 5.0 Hz), 2.81 (t, 2H, *J* = 5.4 Hz); ^{13}C NMR (75 MHz, CD_3OD) δ = 72.7, 71.3-71.2 (wide peak), 71.0, 70.8, 51.5, 41.7; HRMS: calcd. for $C_{12}H_{27}N_4O_5$ $[M+H]^+$: 307.1976, found: 307.1953.

α-Amino,ω-azido,α,ω-dideoxy-tetra(ethylene glycol) (4b): Desymmetrization of diazide **3b** (5.9 g, 24.0 mmol) towards the corresponding azidoamine was carried out analogously to the procedure described for compound **4a**. Silica gel chromatography (CH_2Cl_2/MeOH, 9:1 v/v with 1.0% NH_4OH) yielded the title compound (4.6 g, 21.2 mmol, 88%) as a pale yellow oil. ^{1}H NMR (300 MHz, CD_3OD) δ = 4.83-4.80 (m, 2H), 3.68-3.61 (m, 10H), 3.51 (t, 2H, *J* = 5.3 Hz), 3.37 (t, 2H, *J* = 5.0 Hz), 2.78 (t, 2H, *J* = 5.3 Hz); ^{13}C NMR (75 MHz, CD_3OD) δ = 73.4, 71.6-71.1 (wide peak), 51.8, 42.1; HRMS: calcd. for $C_8H_{19}N_4O_3$ $[M+H]^+$: 219.1452, found: 219.1405.

α-Amino,ω-azido,α,ω-dideoxy-octa(ethylene glycol) (4c): Desymmetrization of diazide **3c** (927 mg, 2.2 mmol) towards the corresponding azidoamine was carried out analogously to the procedure described for compound **4a**, affording the title compound (778 mg, 2.0 mmol, 89%) as a pale yellow liquid. ^{1}H NMR (300 MHz, CD_3OD) δ = 4.79-4.77 (m, 2H), 3.69-3.61 (m, 26H), 3.53 (t, 2H, *J* = 5.4 Hz), 3.37 (t, 2H, *J* = 5.0 Hz), 2.80 (t, 2H, *J* = 5.4 Hz); ^{13}C NMR (75 MHz, CD_3OD) δ = 73.1, 71.6-71.1 (wide peak), 51.8, 42.0; HRMS: calcd. for $C_{16}H_{35}N_4O_7$ $[M+H]^+$: 395.2500, found: 395.2406.

MTX-PEG$_6$-N$_3$ reagent (6a): A solution of methotrexate-α-*tert*-butylester **5**[92] (1.02 g, 2.0 mmol) and aminoazide **4a** (1.28 g, 4.2 mmol, 2.1 eq.) in DMF was treated with triethylamine (560 µL, 4.0 mmol,

2 eq.) and TPTU (600 mg, 2.0 mmol, 1 eq.) and the resulting reaction mixture was stirred for 48 h at RT. After removal of the solvent under reduced pressure, the residue was purified by silica gel chromatography (CH_2Cl_2/MeOH, 93:7 v/v with 0.4% NH_4OH). A co-eluting contaminant could be removed by precipitation from MeOH:Et_2O yielding the title compound (1.29 g, 1.6 mmol, 80%) as an ochre-yellow solid. HRMS calcd. for $C_{36}H_{55}N_{12}O_9$ $[M+H]^+$: 799.4210, found: 799.4211.

Nearly complete assignment of the ^1H and ^{13}C resonances visible in the spectra could be achieved (see Table III.1).

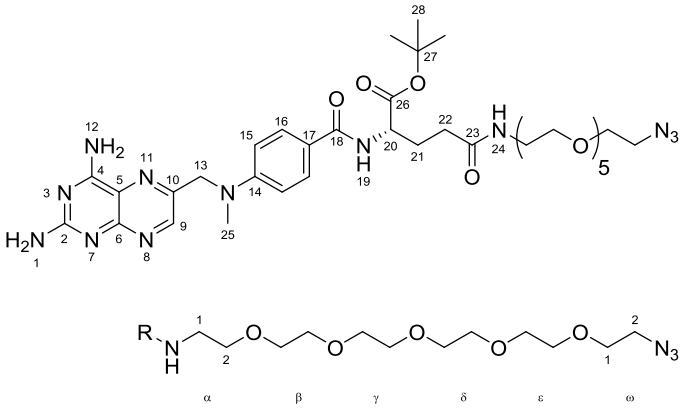


Figure III.16: Labeling scheme of MTX-N_3 6a and its PEG_6 tail.

The methyl resonance of the tBu ester group is easily identified from its integration value; the quaternary carbon is located via HMBC from the tBu protons. This leaves the N-methyl as the only remaining methyl group to be assigned. The N-Me protons correlate to one C_q of the benzene-like ring, one aromatic CH carbon and the CH_2 carbon on the other side of the N-Me group. In the aromatic region three doublets are visible, only two of which correlate mutually in the COSY, identifying these as belonging to the benzene-like aromatic ring. The third doublet has no directly attached carbon and is therefore the amide of the glutamate moiety (amide-1). The only triplet in the aromatic region belongs to the other amide (amide-2; no correlation in the HSQC, triplet expected from the neighbouring CH_2 group). The remaining singlet in the aromatic region is attributed to the only proton directly attached to the pteridine heterocycle. These assigned resonances gave sufficient starting points for further assignment as follows.

The amide carbons C_{18}=O (amide-1) and C_{23}=O (amide-2) were identified at 166.20 and 171.61 ppm, respectively, from $^nJ_{CH}$ correlations to their respective amide proton. Amide-1 also correlates to another C_q at 171.51 ppm, which itself correlates to the $C_{20}H\alpha$ proton from the glutamate moiety in MTX-N_3. This identifies the latter C_q as the ester carbonyl carbon C_{26}. The aromatic CH units were discriminated using a set of $^nJ_{CH}$ correlations from which both benzene C_q's could be identified, and later assigned via connections to the N-Me and N-CH_2 protons and the amide-1 proton. The assignment of the aromatic protons was independently confirmed from the relative intensity and pattern of NOE cross-peaks in the NOESY spectrum. Assignment of the pteridine heterocycle was less straightforward and remains incomplete. This is due in part to the appearance of the NH_2 protons, which are considerably broadened due to intramolecular exchange phenomena (rotation along the C-NH_2 bond), as well as intermolecular exchange with water. As a result, no correlations can be

developed to the directly attached or remote carbons from the NH_2 groups. A broadened C_q carbon at 148.71 ppm was tentatively assigned (see Table III.1).

In all, the above assignments allow to identify all ^{13}C resonances observed above 100 ppm in the ^{13}C APT spectrum. With the exception of the N-Me resonance, all proton resonances between 3 and 4 ppm are contributed by CH_2 type groups, as evidenced by the multiplicity editing in the ^{1}H-^{13}C HSQC spectrum. Only four CH_2 units, labeled a, b, c and d (from left to right) can be separated in this area. Two of these (b and c) are partially overlapping in the ^{1}H spectrum, but can be clearly discriminated in the HSQC. Based on the ^{13}C chemical shifts of these 4 CH_2 moieties, two can be attributed to CH_2-O type units (a, c), as in the PEG chain, and two to CH_2-X type units (b, d), with X designating either the azide or amide function terminating the PEG chain. From the COSY a is seen to couple with b, while c couples with d, indicating that (a,b) and (c,d) are part of the same ethylene unit. Since the amide-2 NH correlates with the carbon directly bound with d, the (c,d) ethylene unit is on the amide side of the PEG-chain, while the (a,b) unit is on the azide side. Thus, the (a,b) unit is identical to the ω-ethylene fragment, while the (c,d) unit is identical to the α-ethylene segment (see labeling scheme). Both CH_2 groups within each of these ethylene moieties were assigned based on chemical shift arguments (CH_2–O $\sim$ highest ppm value).

The remaining proton and carbon resonances are contributed by the PEG chain extending in between the α- and ω-ethylene segments. They extend from 3.54 to 3.44 pm in the ^{1}H spectrum and 69.80 to 69.07 in the ^{13}C spectrum. While some resolution is offered along the ^{13}C dimension, it is insufficient to attempt an assignment of the 4 remaining PEG-repeat units.

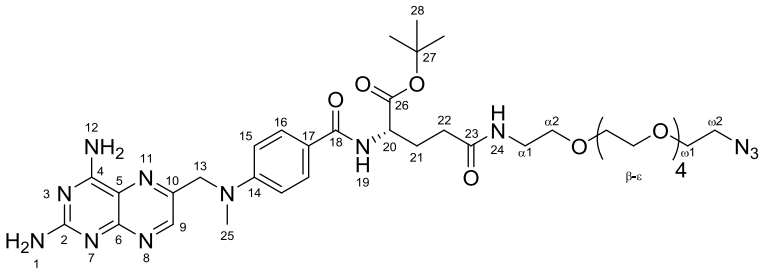


Table III.1: ^{1}H and ^{13}C NMR assignment of MTX-PEG_6-N_3 6a in DMSO-d_6 at 298K (700/176 MHz).

Label	^{1}H (ppm)	^{13}C (ppm)	Multiplicity
1 + 12	(8.49/8.27); (7.55/7.09)	-	br s
2	-	n.a.	
4	-	n.a.	
5	-	n.a.	
6	-	148.71	
9	8.63	148.92	s
10	-	n.a.	
13	4.82	54.48	s
14	-	150.77	
15	6.81	111.06	d
16	7.72	128.9	d
17	-	121.27	
18	-	166.2	
19	8.26	-	d
20	4.21	52.99	m
21	a 1.99 / b 1.89	31.77	m
22	2.19	26.46	m
23	-	171.61	
24	7.89	-	t
25	3.22	52.99	s
26	-	171.51	
27	-	80.27	
28	1.38	27.67	s
α1	3.16 (d)	38.55	m
α2	3.36 (c)	69.09	t
β-ε	3.54 to 3.44	69.80 to 69.07	
ω1	3.58 (a)	69.22	t
ω2	3.37 (b)	49.97	t

MTX-PEG$_4$-N$_3$ reagent (6b): A solution of methotrexate-α-*tert*-butylester **5**[92] (300 mg, 0.59 mmol) and aminoazide **4b** (274 mg, 1.26 mmol, 2.1 eq.) in DMF (10 mL) was treated with triethylamine (1.7 mL, 12.2 mmol, 20.7 eq.) and TPTU (299 mg, 1.01 mmol, 1.7 eq.) and the resulting reaction mixture was stirred for 66 h at RT. After removal of the solvent under reduced pressure, the residue was purified by silica gel chromatography (CH_2Cl_2/MeOH, 93:7 v/v with 0.5% Et_3N). A co-eluting contaminant could be removed by precipitation from MeOH:Et_2O yielding the title compound (209 mg, 0.29 mmol, 49%) as an ochre-yellow solid. LC-HRMS: t_R = 6.34 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{32}H_{47}N_{12}O_7$ $[M+H]^+$: 711.3685, found: 711.3740.

MTX-PEG$_8$-N$_3$ reagent (6c): A solution of methotrexate-α-*tert*-butylester **5**[92] (300 mg, 0.59 mmol) and aminoazide **4c** (548 mg, 1.39 mmol, 2.4 eq.) in DMF (9 mL) was treated with triethylamine (1.7 mL, 12.2 mmol, 20.7 eq.) and TPTU (354 mg, 1.19 mmol, 2.0 eq.) and the resulting reaction mixture was stirred for 71 h at RT. After removal of the solvent under reduced pressure, the residue was purified by silica gel chromatography (CH_2Cl_2/MeOH, 93:7 v/v with 0.5% Et_3N). A co-eluting contaminant could be removed by precipitation from MeOH:Et_2O yielding the title compound (222 mg, 0.25 mmol, 42%) as an ochre solid. LC-HRMS: t_R = 6.52 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{40}H_{63}N_{12}O_{11}$ $[M+H]^+$: 887.4734, found: 887.4756.

***N*-Desmethyltamoxifen hydrochloride:** The synthetic procedure was adapted from Dreaden *et al.*[25] In brief, to an ice cooled solution of tamoxifen (1.12 g, 3.0 mmol) in anhydrous 1,2-dichloroethane (31 mL), α-chloroethyl chloroformate (ACE-Cl) (0.49 mL, 4.5 mmol) was added. The resulting reaction mixture was stirred for 15 min at 0°C and subsequently refluxed for 40 h. After cooling to RT, the mixture was concentrated *in vacuo*. The residue was taken up in methanol (21 mL) and refluxed for 4 h. After cooling to RT, the mixture was concentrated *in vacuo* and the residue was purified by silica gel chromatography (MeOH/CH_2Cl_2, 1:9 v/v) to yield the title compound (1.11 g, 2.82 mmol, 94%) as a white solid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.40 (br s, 2H), 7.35-7.04 (m, 10H), 6.76 (d, 2H, *J* = 8.7 Hz), 6.55 (d, 2H, *J* = 8.7 Hz), 4.02 (t, 2H, *J* = 5.1 Hz), 3.04 (t, 2H, *J* = 5.1 Hz), 2.51 (s, 3H), 2.44 (q, 2H, *J* = 7.5 Hz), 0.92 (t, 3H, *J* = 7.5 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 155.9, 143.7, 142.4, 141.8, 138.2, 136.5, 132.0, 129.8, 129.5, 128.2, 128.0, 126.7, 126.2, 113.7, 63.9, 48.8, 34.1, 29.2, 13.7; HRMS: calcd. for $C_{25}H_{28}NO$ $[M]^+$: 358.2165, found: 358.2189.

***N*-Desmethyl-*N*-propargyltamoxifen (7):** To a suspension of *N*-desmethyltamoxifen hydrochloride (236 mg, 0.6 mmol) and K_2CO_3 (191 mg, 1.38 mmol) in MeCN (6 mL) was added propargyl bromide (81 µL, 80% in toluene, 0.75 mmol). The reaction mixture was heated to 60°C with vigorous stirring for 18 h. After cooling to RT, the mixture was concentrated *in vacuo*. The residue was purified by silica gel chromatography (MeOH/CH_2Cl_2, 1:99 v/v) to yield the title compound (136 mg, 0.34 mmol, 57%) as a yellow oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.37-7.07 (m, 10H), 6.76 (d, 2H, *J* = 8.8 Hz), 6.56 (d, 2H, *J* = 8.8 Hz), 3.93 (t, 2H, *J* = 5.8 Hz), 3.40 (d, 2H, *J* = 2.3 Hz), 2.79 (t, 2H, *J* = 5.8 Hz), 2.45 (q, 2H, *J* = 7.4 Hz), 2.36 (s, 3H), 2.20 (t, 1H, *J* = 2.3 Hz), 0.92 (t, 3H, *J* = 7.4 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 156.8, 143.9, 142.5, 141.5, 138.4, 135.8, 132.0, 129.8, 129.6, 128.2, 128.0, 126.7, 126.1, 113.5, 78.5, 73.5, 65.8, 54.5, 46.1, 42.3, 29.1, 13.7; HRMS: calcd. for $C_{28}H_{30}NO$ $[M+H]^+$: 396.2322, found: 396.2319.

Click-coupled (PEG$_6$) tamoxifen-MFC (8a): Azide **6a** (180 mg, 0.23 mmol) was taken up in a mixture of TFA and CH_2Cl_2 (6 mL, 1:1, v/v) and stirred for 40 min at RT. The reaction mixture was then taken

to dryness, coevaporated twice with toluene and concentrated under high vacuum for 1 h. The residue was taken up in a mixture of water and *tert*-butanol (2 mL, 1:1, v/v) and alkyne **7** (40 mg, 0.1 mmol), $CuSO_4$ (40 µL, 0.5M, 0.2 eq.) and Na ascorbate (200 µL, 0.5M, 1.0 eq.) were added. Finally, the resulting reaction mixture was charged with a catalytic amount of TBTA[99] and heated to 80°C under vigorous stirring. Upon completion of the reaction (96 h), the solution was cooled to RT and concentrated *in vacuo*. The residue was purified by preparative RP-HPLC (10-100% MeCN) yielding the title compound (57 mg, 50 µmol, 50%) as a pale yellow amorphous solid. LC-HRMS: t_R = 6.98 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{60}H_{77}N_{13}O_{10}$ $[M+2H]^{2+}$: 569.7953, found: 569.7931.

For the NMR assignment of the MTX-PEG part of **8a**, we refer to the assignment of reversine-MFC **18** (see further). The labeling scheme of the tamoxifen part of **8a** is depicted below:

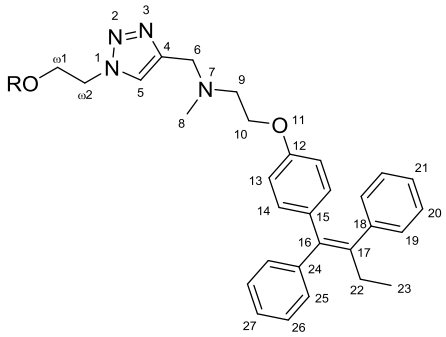


Figure III.17: Labeling scheme of the tamoxifen part of 8a.

From the ω2 CH_2 protons in the PEG chain, there is an HMBC correlation to a CH at 124.0 ppm. With HSQC this is seen to correlate to a ^1H singlet at 7.92 ppm, this is H-5. Another HMBC correlation to C-5 comes from a CH_2 singlet at 3.60 ppm, this is H-6. Starting from H-6, two more HMBC correlations are seen, one to a C_q at 143.2 ppm which is assigned to be C-4 and one to a CH_3 at 42.0 ppm that has an HSQC correlation to a ^1H at 2.17 ppm and is assigned to Me-8. From Me-8, there are NOESY correlations visible to H-6, a CH_2 triplet at 2.63 ppm and a CH_2 triplet at 3.91 ppm; these can be assigned to H-9 and H-10, respectively. H-10 has an HMBC correlation to a C_q at 156.4 ppm, which is assigned to C-12. Another HMBC correlation to C-12 is visible from a CH that integrates for 2 at 6.72 ppm; this is assigned to H-14. From this signal, there is a COSY correlation to a CH that integrates for 2 at 6.59 ppm, this is H-13. From H-13 an HMBC to 156.3 ppm is visible, this is C-15. Thus a consistent picture emerges for the first aromatic moiety.

In the HSQC there is one more CH_3 signal visible, this is a triplet at 0.84 ppm and can be assigned to the Me-23. It shows a COSY correlation to a CH_2 quadruplet at 2.36 ppm, which is H-22. An HMBC correlation is visible from H-23 and H-22 to a C_q at 140.6 ppm, this is C-17. Another HMBC correlation to C-17 comes from the left part of a CH type resonance at 7.12 ppm that integrates for 3 ^1H's and comes from, according to the HSQC spectrum, two carbons. The left part can be assigned to H-19. The right part shows an HMBC correlation to the left part and vice versa, we can thus assign the right part to be H-21. This also agrees with the sum of 3 for this resonance. From these two resonances there is a COSY correlation to the right part of a CH signal at 7.18-7.20 ppm that in total integrates for 4 ^1H's; this can be assigned to H-20. This signal also shows an HMBC correlation to a C_q at 141.7 ppm, C-18. Thus a consistent picture emerges for the second aromatic moiety.

Also, an HMBC to C-18 is visible from H-22. Another HMBC correlation from H-22 goes to a C_q at 137.8 ppm, this is C-16. The left part of the signal at 7.18-7.20 ppm shows an HMBC correlation to C-16 as well and can thus be assigned to H-25. H-25 has a COSY correlation to a CH at 7.37 ppm that integrates for 2 protons and is assigned to H-26. From this signal, an HMBC correlation is visible to a C_q at 143.2 ppm, this is C-24. Also from H-26 a COSY correlation is visible to a CH signal at 7.28 ppm that integrates for 1 proton and is assigned to H-27. Thus a consistent picture emerges for the third aromatic moiety.

Table III.2: ^{1}H and ^{13}C NMR assignment of tamoxifen-MFC 8a in DMSO-d_6 at 298K (700/176 MHz).

Label	^{1}H (ppm)	^{13}C (ppm)	Multiplicity
4	-	143.2	
5	7.92	124.0	s
6	3.62	52.0	s
8	2.17	42.0	s
9	2.63	54.7	t
10	3.91	65.6	t
12	-	156.4	
13	6.59	113.4	d
14	6.72	131.4	d
15	-	156.3	
16	-	137.8	
17	-	140.6	
18	-	141.7	
19	7.12	129.4*	m
20	7.18	128.0*	m
21	7.11	126.3*	m
22	2.36	28.5	q
23	0.84	13.3	t
24	-	143.2	
25	7.20	129.0*	m
26	7.37	128.3	m
27	7.28	126.7	m

* tentative assignment only

Click-coupled (PEG_4) tamoxifen-MFC (8b): Azide **6b** (164 mg, 0.23 mmol) was deprotected and subsequently conjugated to alkyne **7** (40 mg, 0.1 mmol) using the procedure described for conjugate **8a**. Upon completion of the reaction (18 h) and workup, the material was purified by preparative RP-HPLC (10-100% MeCN), yielding the title compound (55 mg, 52 µmol, 52%) as a pale yellow amorphous solid. LC-HRMS: t_R = 7.04 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{56}H_{69}N_{13}O_8$ $[M+2H]^{2+}$: 525.7691, found: 525.7696.

Click-coupled (PEG_8) tamoxifen-MFC (8c): Azide **6c** (204 mg, 0.23 mmol) was deprotected and subsequently conjugated to alkyne **7** (40 mg, 0.1 mmol) using the procedure described for conjugate **8a**. Upon completion of the reaction (18 h) and workup, the material was purified by preparative RP-HPLC (10-100% MeCN), yielding the title compound (78 mg, 63 µmol, 63%) as a pale yellow

amorphous solid. LC-HRMS: t_R = 7.11 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{64}H_{85}N_{13}O_{12}$ $[M+2H]^{2+}$: 613.8215, found: 613.8204.

Methyl (*N*-desmethyltamoxifen-*N*-yl)acetate (9): To an ice cooled solution of methyl bromoacetate (39 µL, 0.42 mmol) in CH_2Cl_2 (1.3 mL), *N*-desmethyltamoxifen hydrochloride (165 mg, 0.42 mmol) and DIPEA (183 µL, 1.05 mmol) were added. The resulting solution was stirred overnight letting the temperature rise to RT. The mixture was diluted with CH_2Cl_2 (6.5 mL), washed with sat. $NaHCO_3$ (2 x 10 mL), dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by silica gel chromatography (MeOH/CH_2Cl_2, 1:99 v/v) yielding the title compound (161 mg, 0.38 mmol, 91%) as a pale white oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.37-7.07 (m, 10H), 6.76 (d, 2H, *J* = 8.7 Hz), 6.53 (d, 2H, *J* = 8.7 Hz), 3.95 (t, 2H, *J* = 5.7 Hz), 3.64 (s, 3H), 3.37 (s, 2H), 2.90 (t, 2H, *J* = 5.7 Hz), 2.45 (q, 2H, *J* = 7.5 Hz), 2.45 (s, 3H), 0.92 (t, 3H, *J* = 7.5 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 171.6, 156.7, 143.9, 142.5, 141.5, 138.4, 135.7, 132.0, 129.8, 129.6, 128.2, 128.0, 126.6, 126.1, 113.5, 66.1, 58.6, 55.7, 51.6, 43.2, 29.1, 13.7; HRMS: calcd. for $C_{28}H_{32}NO_3$ $[M+H]^+$: 430.2377, found: 430.2372.

(*N*-Desmethyltamoxifen-*N*-yl)acetic acid (10): A solution of compound **9** (249 mg, 0.58 mmol) in MeOH (13 mL) was treated with NaOH (0.96 mL, 4.0M). The resulting reaction mixture was stirred for 6 h at 50°C. After cooling to RT, the mixture was neutralized by addition of HCl (1.31 mL, 3.0M), concentrated *in vacuo* and coevaporated twice with toluene. The residue was purified by silica gel chromatography (MeOH/CH_2Cl_2, 12:88 v/v) yielding the title compound (342 mg) as a white amorphous solid. The additional weight could be explained by a significant amount of silica gel present in the material. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.31-7.06 (m, 10H), 6.72 (d, 2H, *J* = 8.0 Hz), 6.44 (d, 2H, *J* = 8.0 Hz), 4.07 (m, 2H), 3.86 (br s, 2H), 3.46 (m, 2H), 2.77 (br s, 3H), 2.43 (q, 2H, *J* = 7.3 Hz), 0.92 (t, 3H, *J* = 7.3 Hz); LC-HRMS: t_R = 8.87 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{27}H_{30}NO_3$ $[M+H]^+$: 416.2220, found: 416.2065.

Amide-coupled tamoxifen-MFC (11): Azide **6a** (160 mg, 0.2 mmol) was taken up in a mixture of TFA and CH_2Cl_2 (7 mL, 1:1, v/v) and stirred for 40 min at RT. The reaction mixture was then taken to dryness, coevaporated twice with toluene and concentrated under high vacuum for 1 h. The residue was taken up in THF (4 mL), treated with a solution of Me_3P in THF (2 mL, 1.0M) and stirred overnight at RT. Subsequently, H_2O (100 µL) was added and stirring was continued for 1 h. The solution was concentrated *in vacuo* and the resulting residue was taken up in a mixture of *tert*-butanol and H_2O (4 mL, 1:1, v/v) and lyophilized. The presence of a significant contaminant bearing an additional methyl group, as judged by LC-HRMS (calcd. for $C_{33}H_{52}N_{10}O_9$ $[M+Me+2H]^{2+}$: 366.1954, found: 366.1926), necessitated an additional purification. Therefore the crude dry material was purified by preparative RP-HPLC (5-100% MeCN) yielding the pure amine (41 mg, 57 µmol, 29%) as an ochre semi solid. LC-HRMS: t_R = 11.40 min (5-15% MeCN, 18 min run); HRMS: calcd. for $C_{32}H_{50}N_{10}O_9$ $[M+2H]^{2+}$: 359.1876, found: 359.1852. Acid **10** (90 mg, ≈ 153 µmol)[100] was taken up in DMF (1.85 mL) and preactivated for 5 min using PyBOP (60 mg, 115 µmol) and DIPEA (150 µL, 861 µmol). The MTX-amine (41 mg, 57 µmol) was taken up in DMF (0.5 mL) and combined with the preactivation mixture and stirred for 4 h. The solution was concentrated *in vacuo* and the residue was purified by preparative RP-HPLC (10-100% MeCN) yielding the title compound (21 mg, 19 µmol, 33%) as a pale yellow amorphous solid. LC-HRMS: t_R = 7.00 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{59}H_{77}N_{11}O_{11}$ $[M+2H]^{2+}$: 557.7897, found: 557.7875.

The NMR characterization of tamoxifen-MFC **11** was split into two parts, beginning with its MTX-PEG moiety. The labeling is depicted below:

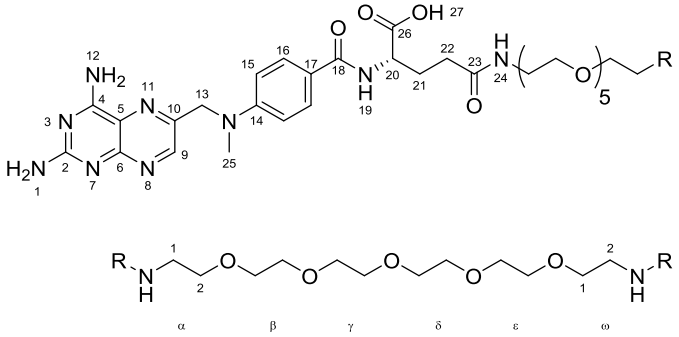


Figure III.18: Labeling scheme of the MTX-PEG part of MFC 11.

For the descriptive assignment of the MTX part of **11** (identical to that of **8a**), we refer to the analysis of reversine-MFC **18** (see further). Nearly complete assignment of the ^{1}H and ^{13}C resonances visible in the spectra could be achieved (see Table III.3).

Table III.3: ^{1}H and ^{13}C NMR assignment of the MTX-PEG part of tamoxifen-MFC 11 in DMSO-d_6 at 298K (700/176 MHz).

Label	^{1}H (ppm)	^{13}C (ppm)	Multiplicity
1 + 12	n.a.	-	
2	-	n.a.	
4	-	n.a.	
5	-	n.a.	
6	-	155.21*	
9	8.55	148.7	s
10	-	145.9	
13	4.76	54.8	s
14	-	150.9	
15	6.81	111.1	d
16	7.69	128.6	d
17	-	121.4*	
18	-	165.8*	
19	8.08	-	d
20	4.18	52.7	m
21	a 2.00 / b 1.88	27.15	m
22	2.16	32.0	m
23	-	171.9	
24	7.91	-	t
25	3.19	39.05	s
26	-	173.5	
27	n.a.	-	
α1	3.16	38.5	m
α2	3.35	69.0	t
β-ω1	3.4-3.55	69.0-70.4	
ω2	3.18	38.0	m

* tentative assignment only

The labeling scheme of the tamoxifen part of tamoxifen-MFC **11** is depicted below:

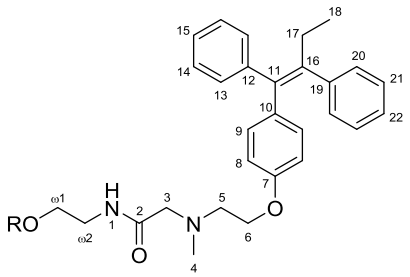


Figure III.19: Labeling scheme of the tamoxifen part of MFC 11.

Me-4 is easily found at 2.26 ppm because it is the only remaining CH_3 group that is a singlet. Starting from this singlet, there are two HMBC correlations, one to a CH_2 at 55.8 ppm and one to a CH_2 at 61.0 ppm. The CH_2 at 55.8 ppm correlates to a proton resonance at 2.68 ppm that is a triplet and can thus be assigned to H-5. The other CH_2 has a proton resonance at 2.98 ppm that is a singlet and can be assigned to H-3. From H-3 an HMBC correlation is visible to a C_q at 169.7 ppm that is assigned to C-2. From H-3 and H-4 a NOESY correlation is visible to a triplet at 7.66 ppm that has no HSQC correlation and can be assigned to H-1.

From H-5 there is a COSY correlation visible to a triplet at 3.91 ppm that can be assigned to H-6. In the HMBC spectrum, a correlation from H-6 to a C_q at 156.6 ppm is visible, this is assigned to C-7. Another two proton resonances have an HMBC correlation to this last one, an aromatic doublet at 6.60 ppm and one at 6.73 ppm. These also have a COSY correlation to each other and can thus be assigned to H-8 and H-9, respectively. Another CH_3 group is present in this part of the molecule, a triplet at 0.82 ppm; this is assigned to H-18. From this resonance, a COSY correlation is visible to a quadruplet at 2.36 ppm, which is assigned to H-17. From both H-18 and H-17 an HMBC correlation is visible to a C_q at 140.7 ppm, this is C-16.

In the aromatic region there are still 4 signals left that integrate all together for 10 protons. The resonance at 7.37 ppm integrates for 2 protons and is assigned to H-14. H-14 has an HMBC correlation to a C_q at 143.2 ppm, which is assigned to C-12. From H-14 there are two COSY correlations, one to the resonance at 7.28 ppm, which is assigned to H-15 and one to the multiplet at 7.19 ppm, from which the left part can be assigned to H-13. The latter has an HMBC correlation to a C_q at 137.9 ppm, C-11. The right part of the multiplet at 7.19 ppm is assigned to H-21 and has an HMBC correlation to a C_q at 141.8 ppm, C-19. H-21 also has a COSY correlation to the remaining aromatic resonance at 7.11 ppm that integrates for 3 protons and is assigned to both H-22 and H-20, which also shows an HMBC correlation to C-16.

Table III.4: ^{1}H and ^{13}C NMR assignment of the tamoxifen part of tamoxifen-MFC 11 in DMSO-d_6 at 298K (700/176 MHz).

Label	^{1}H (ppm)	^{13}C (ppm)	Multiplicity
1	7.66	-	t
2	-	169.7	
3	2.98	61.0	s
4	2.26	43.0	s
5	2.68	55.9	t
6	3.91	65.4	t
7	-	156.6	
8	6.60	113.4	d
9	6.73	131.3	d
10	-	156.6*	
11	-	137.9	
12	-	143.2	
13	7.19	128.9	m
14	7.37	128.3	m
15	7.28	126.6	m
16	-	140.7	
17	2.36	28.5	q
18	0.82	13.3	t
19	-	141.8	
20	7.12	129.4	m
21	7.18	127.9	m
22	7.11	126.2	m

* tentative assignment only

***N*-(4-Nitrophenyl)-*N'*-propargylpiperazine (13):** To a suspension of *N*-(4-nitrophenyl)piperazine **12** (1.0 g, 4.8 mmol) and K_2CO_3 (868 mg, 6.3 mmol) in MeCN (24 mL) was added propargyl bromide (651 µL, 80% in toluene, 6.04 mmol). The reaction mixture was heated to 60°C with vigorous stirring for 18 h. After cooling to RT, the mixture was concentrated *in vacuo*. The residue was purified by silica gel chromatography (MeOH/CH_2Cl_2, 3:97 v/v) to yield the title compound (1.13 g, 4.6. mmol, 95%) as an amorphous yellow solid, which upon crystallization (CH_2Cl_2/hexane) gave fine needles. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 8.12 (d, 2H, *J* = 9.5 Hz), 6.83 (d, 2H, *J* = 9.5 Hz), 3.47 (t, 4H, *J* = 5.3 Hz), 3.39 (d, 2H, *J* = 2.5 Hz), 2.72 (t, 4H, *J* = 5.3 Hz), 2.29 (t, 1H, *J* = 2.5 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 154.8, 138.6, 126.0, 112.8, 78.1, 73.8, 51.3, 47.0, 46.8; HRMS: calcd. for $C_{13}H_{16}N_3O_2$ $[M+H]^+$: 246.1237, found: 246.1250.

***N*-(4-Aminophenyl)-*N'*-propargylpiperazine (14):** Compound **13** (4.2 g, 17.3 mmol) and $SnCl_2$ (18.4 g, 96.8 mmol) were taken up in EtOH (62 mL). The resulting reaction mixture was refluxed at 70°C under vigorous stirring for 9 h. After cooling to RT, the pH was made slightly basic (pH 7-8) by slow addition of sat. $NaHCO_3$. The alkaline solution was transferred to a separation funnel and the aqueous fraction was extracted thrice with EtOAc (60 mL). All EtOAc fractions were pooled, dried over Na_2SO_4 and taken to dryness. The residue was purified by silica gel chromatography (MeOH/CH_2Cl_2, 6:94 v/v) yielding the title compound (1.9 g, 8.9 mmol, 87%) as a beige amorphous solid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 6.80 (d, 2H, *J* = 8.7 Hz), 6.62 (d, 2H, *J* = 8.7 Hz), 3.41 (br s, 2H),

3.34 (d, 2H, *J* = 2.4 Hz), 3.08 (t, 4H, *J* = 5.0 Hz), 2.72 (t, 4H, *J* = 5.0 Hz), 2.28 (t, 1H, *J* = 2.4 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 144.3, 140.2, 118.6, 116.1, 78.8, 73.4, 52.0, 50.8, 46.8; HRMS: calcd. for $C_{13}H_{18}N_3$ $[M+H]^+$: 216.1495, found: 216.1489.

***N*^{6}-Cyclohexyl-2-fluoroadenine (16):** The synthetic procedure was adapted from Chen *et al.*[28] In brief, to a solution of 6-chloro-2-fluoropurine (600 mg, 3.48 mmol) in *n*-butanol (35 mL), cyclohexylamine (398 µL, 3.48 mmol) and *N*,*N*-diisopropylethylamine (DIPEA) (728 µL, 4.18 mmol) were added. The resulting reaction mixture was heated to 80°C for 16 h under vigorous stirring. After cooling to RT, the mixture was concentrated *in vacuo* and the residue was purified by silica gel chromatography ($MeOH/CH_2Cl_2$, 3:97 v/v) to yield the title compound (519 mg, 2.21 mmol, 64%) as a white solid. ^{1}H NMR (300 MHz, CD_3OD) δ = 8.00 (s, 1H), 4.04 (m, 1H), 2.13-1.98 (m, 2H), 1.89-1.75 (m, 2H), 1.74-1.62 (m, 1H), 1.56-1.20 (m, 5H); HRMS: calcd. for $C_{11}H_{15}N_5F$ $[M+H]^+$: 236.1306, found: 236.1284.

***N*^{6}-Cyclohexyl-2-(4-(4-propargylpiperazin-1-yl)-*N*-anilinyl)-adenine (17):** *N*^{6}-Cyclohexyl-2-fluoroadenine (235 mg, 1.0 mmol) and compound **14** (431 mg, 2.0 mmol) were taken up in EtOH (3 mL) and the resulting reaction mixture was heated (µW) to 130°C for 5 h. After cooling to RT, the mixture was transferred directly into a silica gel column ($MeOH/CH_2Cl_2$, 6:94 v/v). The thus obtained semi pure product was purified further by precipitation (CH_2Cl_2/hexane) yielding the title compound (157 mg, 0.36 mmol, 36%) as a beige amorphous solid. (More product could be isolated from the mother liquor, albeit in a significantly lower purity as the first crop.) ^{1}H NMR (300 MHz, $CDCl_3$) δ = 12.82 (br s, 1H), 7.42 (d, 2H, *J* = 8.9 Hz), 6.91 (d, 2H, *J* = 8.9 Hz), 6.74 (s, 1H), 6.65 (s, 1H), 5.56 (d, 1H, *J* = 6.3 Hz), 4.09 (m, 1H), 3.36 (d, 2H, *J* = 2.1 Hz), 3.18 (t, 4H, *J* = 4.8 Hz), 2.74 (t, 4H, *J* = 4.8 Hz), 2.28 (t, 1H, *J* = 2.1 Hz), 2.13-2.02 (m, 2H), 1.83-1.71 (m, 2H), 1.70-1.59 (m, 1H), 1.50-1.14 (m, 5H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 157.3, 154.5, 150.6, 147.9, 135.9, 132.6, 123.3, 117.5, 114.6, 78.8, 73.5, 52.0, 50.0, 49.3, 47.0, 33.4, 25.8, 25.1; HRMS: calcd. for $C_{24}H_{31}N_8$ $[M+H]^+$: 431.2666, found: 431.2689.

Reversine-MFC (18): Azide **6a** (160 mg, 0.2 mmol) was taken up in a mixture of TFA and CH_2Cl_2 (5 mL, 1:1, v/v) and stirred for 40 min at RT. The reaction mixture was then taken to dryness, coevaporated twice with toluene and concentrated under high vacuum for 1 h. The residue was taken up in a mixture of water and *tert*-butanol (3 mL, 1:1, v/v) and alkyne **17** (43 mg, 0.1 mmol), $CuSO_4$ (20 µL, 0.5M, 0.1 eq.) and Na ascorbate (100 µL, 0.5M, 0.5 eq.) were added. Finally, the resulting reaction mixture was charged with a catalytic amount of TBTA[99] and Et_3N and heated to 80°C under vigorous stirring for 72 h. After cooling to RT, the solution was concentrated *in vacuo* and the residue was purified by preparative RP-HPLC (10-100% MeCN) yielding the title compound (36 mg, 31 µmol, 31%) as a pale yellow amorphous solid. LC-HRMS: t_R = 5.17 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{56}H_{79}N_{20}O_9$ $[M+3H]^{3+}$: 391.8774, found: 391.7788.

The NMR characterization of reversine-MFC **18** was split into two parts, beginning with its MTX-PEG moiety. Nearly complete assignment of the ^{1}H and ^{13}C resonances visible in the spectra could be achieved (see Table III.5). The labeling is depicted below:

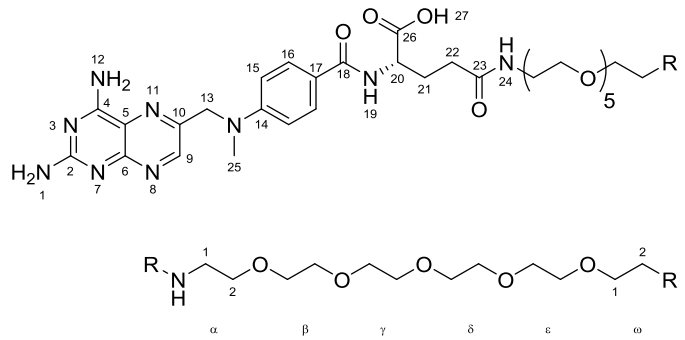


Figure III.20: Labeling scheme of the common MTX-PEG part of MFCs 18 and 8a.

Since there is only one CH_3 group in conjugate **18**, Me-25 can, with its integration value, easily be assigned to the singlet at 3.19 ppm. Starting from this singlet there are two HMBC correlations, one to a CH_2 at 54.8 ppm and one to a C_q at 150.6 ppm, which is C-14. The CH_2 signal has a proton resonance at 4.76 ppm and is a singlet, thus it can be assigned to be H-13. It also shows a NOESY correlation to both Me-25 and a signal in the aromatic region which can be assigned to H-9. Also, H-13 shows an HMBC correlation to a C_q at 145.9 ppm, this is C-10. There is also a HMBC correlation visible from H-9 to a C_q, which is thus tentatively assigned to C-6.

From Me-25 and H-13 there is a NOESY correlation to an aromatic doublet at 6.83 ppm, which can be assigned to the two H-15 protons. From this resonance, a COSY correlation to a doublet at 7.71 ppm is visible, this is H-16. H-15 has an HMBC correlation to 121.1 ppm, which is assigned to C-17, and H-16 has an HMBC correlation to 150.6 ppm which was already assigned as C-14. H-16 additionally has a small HMBC to a C_q at 166.0 ppm that is tentatively assigned to C-18.

From H-16 there is a NOESY correlation to a resonance at 8.18 ppm that does not have an HSQC correlation and can therefore be assigned to the H-19 amine. From this position, a COSY correlation is visible to a resonance at 4.25 ppm, which is according to the HSQC a CH and can be assigned to H-20. From H-20 there are two COSY correlations to 2.04 and 1.90 ppm, respectively. Looking at the integrations and the HSQC spectrum, these can be assigned to the two protons H-21. From these resonances, a COSY correlation is visible towards a resonance at 2.18 ppm, which is assigned to H-22 based on HSQC and the integral value. Both resonances from H-21 and H-22 have a NOESY correlation to the amine H-19. From H-22 there is also a NOESY correlation to a triplet at 7.88 ppm, which has an integral of 1 and no HSQC correlation and can be assigned to the H-24 amine.

From H-24 a COSY correlation to a multiplet at 3.16 ppm is visible that has an HSQC correlation to a CH_2 and an integral value of two and can be assigned to H-α1. From this one, a COSY correlation to a triplet at 3.35 ppm with an HSQC correlation to a CH_2 is visible, that can be assigned to H-α2. From H-24, H-21 and H-α1 an HMBC correlation is visible towards a C_q at 171.81 ppm, that can be assigned to C-23. From H-α2, HMBC correlations to 69.6 ppm are visible; these correlate to multiple CH_2's between 3.40 and 3.55 ppm and are thus assigned to the CH_2's β to ε. Another resonance at 3.82 ppm (triplet) also shows an HMBC to these CH_2 carbons; this resonance is a CH_2 and can be assigned to H-ω1. From the latter, there is a COSY correlation to a triplet at 4.50 ppm that has a HSQC correlation to a CH_2, which is H-ω2.

The H-1 and H-12 amines are present in the region 6.5-8 ppm; this is in agreement with the integrations. However, they cannot be distinguished. $O_{27}H$ is assigned at 12.29 ppm. Finally, a small HMBC correlation is visible from H-20 and H-21 to a C_q at 173.7 ppm; this is assigned to C-26.

Table III.5: ^{1}H and ^{13}C NMR assignment of the MTX-PEG part of reversine-MFC 18 in DMSO-d_6 at 298K (700/176 MHz).

Label	^{1}H (ppm)	^{13}C (ppm)	Multiplicity
1 + 12	n.a.	-	
2	-	n.a.	
4	-	n.a.	
5	-	n.a.	
6	-	155.21*	
9	8.55	148.9	s
10	-	145.9	
13	4.76	54.54	s
14	-	150.8	
15	6.82	110.72	d
16	7.71	128.53	d
17	-	121.2*	
18	-	166.0*	
19	8.18	-	d
20	4.25	52.11	m
21	a 2.04 / b 1.90	26.41	m
22	2.18	31.7	m
23	-	171.7	
24	7.88	-	t
25	3.19	38.83	s
26	-	173.7	
27	12.29	-	s
α1	3.16	38.21	m
α2	3.35	68.76	t
β-ε	3.4-3.55	69.6	
ω1	3.82	68.5	t
ω2	4.50	48.9	t

* tentative assignment only

The labeling scheme of the reversine part of reversine-MFC **18** is depicted below:

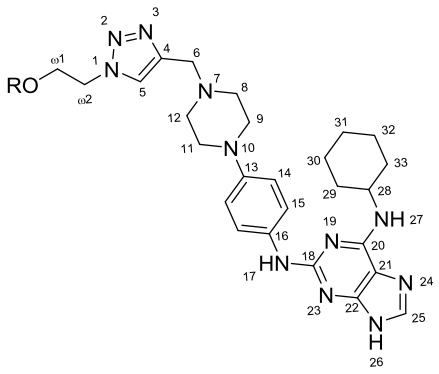


Figure III.21: Labeling scheme of the reversine part of MFC 18.

From the ω2 CH_2 protons in the PEG chain, there is an HMBC correlation to a CH at 123.7 ppm. With HSQC this is seen to correlate to a ^1H singlet at 7.98 ppm, this is H-5. Another HMBC correlation to C-5 comes from a CH_2 singlet at 3.60 ppm, this is H-6. Starting from H-6, two more HMBC correlations are seen: one to a C_q at 143.1 ppm which is assigned to be C-4; and one to a CH_2 at 52.11 ppm that has an HSQC correlation to 2.54 ppm and integrates for 4 protons, and can be assigned to the isochronous H-8 and H-12. From this position there is a COSY correlation to a signal at 3.01 ppm that integrates for 4 protons and can be assigned to the isochronous H-9 and H-11. From this last resonance, a NOESY correlation is visible to two CH doublets, each integrating for two protons, at 6.80 and 7.63 ppm, which can be assigned to H-14 and H-15, respectively. An HMBC correlation is visible from H-14 to a C_q at 134.4 ppm that is assigned to C-16. Also from H-15 an HMBC correlation is visible to a C_q at 145.1 ppm that is assigned to C-13. A NOESY correlation is visible from H-15 to a NH at 8.48 ppm that is assigned to the H-17 amine.

Because of the large amount of nitrogen atoms and quaternary carbons, assignment from this point onwards was difficult. A TOCSY correlation was found between 5 CH_2 resonances that integrated together for 10 protons and a CH resonance at 4.05 ppm, which is assigned to H-28 (the resonance is however broadened and no HSQC correlation was found). Distinction between the 5 CH_2 signals to assign them to 29-33 is difficult and only done tentatively. A NOESY correlation is visible from H-28 and the CH_2 at 1.33 ppm to a NH signal at 7.04 ppm that is assigned to the H-27 amine.

H-25 is only assigned tentatively by means of exclusion; there is only one singlet left in the aromatic region and this is therefore very likely to be H-25. The remaining quaternary carbons C-22, C-21, C-20 and C-18 are not assigned due to lack of HMBC correlations. Also the amine H-26 is not assigned. Nevertheless, the integrity of the chemical link between the PEG chain and the clicked moiety could be established without ambiguity.

Table III.6: ^{1}H and ^{13}C NMR assignment of the reversine part of reversine-MFC 18 in DMSO-d_6 at 298K (700/176 MHz).

Label	^{1}H (ppm)	^{13}C (ppm)	Multiplicity
4	-	143.1	
5	7.98	123.7	s
6	3.60	52.34	s
8+12	2.54	52.11	br t
9+11	3.01	49.1	br t
13	-	145.1	
14	6.80	115.7	d
15	7.63	119.0	d
16	-	134.4	
17	8.48	-	br s
18	-	n.a.	
20	-	n.a.	
21	-	n.a.	
22	-	n.a.	
25	8.16*	n.a.	s
26	n.a.	-	
27	7.04	-	m
28	4.05		m
29	1.33*	24.9	m
30	1.33*	24.9	m
31	1.15+1.63*	25.0	m
32	1.76*	24.9	m
33	1.33+1.93*	32.3	m

* tentative assignment only

4-((trimethylsilyl)ethynyl)phenol (20): To a solution of 4-iodophenol (2.2 g, 10.0 mmol) in triethylamine (30 mL) were added $PdCl_2(PPh_3)_2$ (70 mg, 0.1 mmol, 1 mol%), CuI (19 mg, 0.1 mmol, 1 mol%) and trimethylsilylacetylene (2.82 mL, 20.0 mmol) and the solution was heated to 80°C under vigorous stirring. After 6 h, the reaction mixture was cooled to RT, filtered and concentrated *in vacuo*. The dark colored residue was purified by silica gel chromatography (MeOH/CH_2Cl_2, 1:99 v/v) to yield the title compound (1.83 g, 9.6 mmol, 96%) as a light brown solid, which upon recrystallization (heptanes) gave off white needles. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.35 (d, 2H, *J* = 8.7 Hz), 6.74 (d, 2H, *J* = 8.7 Hz), 5.13 (s, 1H), 0.23 (s, 9H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 155.7, 133.7, 115.5, 115.3, 105.1, 92.6, 0.1; HRMS calcd. for $C_{11}H_{13}OSi$ $[M-H]^-$: 189.0741, found: 189.0733.

Alkyne-functionalized FK506 (22): Alcohol **19**[50] (104 mg, 0.1 mmol) and *p*-(TMS-ethynyl)phenol (**20**) (38 mg, 0.2 mmol) were placed in a flask under a nitrogen atmosphere. To the dry material was added a solution of PPh_3 in toluene (1.3 mL, 0.1M, 1.3 eq.). The solution was cooled on ice and dropwise a solution of DEAD in toluene (1.3 mL, 0.1M, 1.3 eq.) was added and the solution was stirred overnight. The reaction mixture was transferred directly into a silica gel column and the product (106 mg, 87.7 µmol, 88%) was obtained after elution (EtOAc/toluene 0-25%, v/v). HRMS calcd. for $C_{66}H_{113}N_2O_{13}Si_3$ $[M+NH_4]^+$: 1225.7545, found: 1225.7505. The product is taken up in MeCN (4 mL) and cooled on ice. To the solution are added acetic acid (30 µL, 526 µmol, 6 eq.) and TBAF

(525 µL, 1.0M in THF, 525 µmol, 6 eq.) and the mixture was stirred overnight. The solution was cooled on ice again and HF·pyridine (100 µL) was added and stirring was continued for 6 h. The solution was poured into sat. $NaHCO_3$ (20 mL) and transferred to a separation funnel. The aqueous fraction was extracted with EtOAc (4 x 25 mL) and the organic fractions were pooled, dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by silica gel chromatography (MeOH/CH_2Cl_2 0-10%, v/v) to yield the title compound (73 mg, 80.4 µmol, 80% based on **19**) as an off white foam. ^{1}H NMR (700 MHz, $SO(CD_3)_2$) (mixture of rotamers) δ = 7.38 (2H, d), 6.87 (2H, d), 6.54 (0.5H, s), 5.25 (0.5H, d), 5.11 (1H, m), 5.07 (0.5H, d), 5.04 (0.5H, dd), 4.80 (1.5H, br m), 4.70 (0.5H, d), 4.63 (1H, t), 4.42 (0.5H, m), 4.21 (0.5H, d), 4.00 (1H, d), 3.93-3.78 (3H, m), 3.74 (0.5H, dt), 3.68 (0.5H, m), 3.57-3.36 (2.5H, m), 3.34-3.13 (15H, br m), 2.91 (1.5H, m), 2.74 (0.5H, m), 2.62 (0.5H, m), 2.44 (0.5H, m), 2.36 (0.5H, dd), 2.29-0.9 (36.5H, br m), 0.9-0.7 (9H, m), 0.62 (1.5H, d); HRMS calcd. for $C_{51}H_{73}NO_{13}Na$ $[M+Na]^+$: 930.4974, found: 930.5010.

The NMR spectra are quite complex and a full assignment has not been attempted, given that the synthesis of **22** starts from commercially available FK506. A lot of resonances are apparent. The labeling scheme for **22** is presented hereafter.

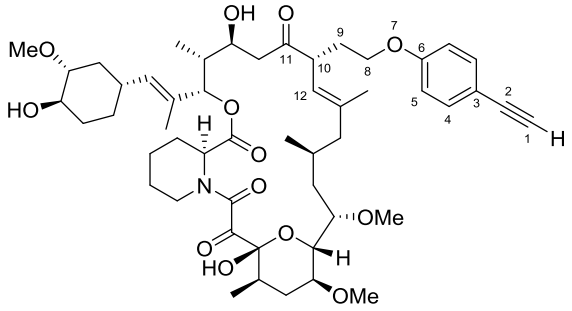


Figure III.22: Labeling scheme for the relevant part of alkyne-functionalized FK506 (22)

Analysis starts by noting the alkyne proton to carbon correlations in the HSQC, which have a typical position and appearance due to the unusual $^1J_{CH}$ (~250 Hz) and $^2J_{CH}$ (~50 Hz) scalar coupling constant. As a result of the large size of the latter, a cross-peak is visible in the HSQC spectrum representing a ***–C$_q$***Ξ**C–*H*** correlation. The corresponding carbons are at 83.79 (C_q) and 79.17 (C-H) ppm. Their identity is further confirmed by noting the residual ^{1}J scalar coupling doublet artefacts that occur at the frequency positions of the alkyne C_q (Figure III.23). Because it is correlated to the alkyne proton via ^2JC_qH coupling of 51.4 Hz, the ^{2}JCH splitting is not removed by the low-pass filter set at 140 Hz. Thus both the chemical shift and the appearance confirm the assignment to the alkyne functionality in **22**.

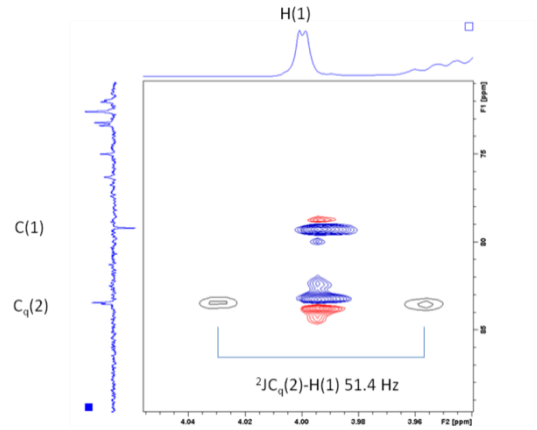


Figure III.23: Superposition of a detail of the ^{1}H-{^{13}C} HSQC spectrum (red/blue for -/+ intensity) with the ^{1}H-{^{13}C} HMBC spectrum (black) showing the residual ^{2}JCH splitting artefact at the level of the Cq(2) carbon. The antiphase nature of the correlation in the HSQC results from the large mismatch between the set and actual $^n J_{CH}$ couplings.

Continuing from the alkyne proton unit, an additional ^{n}JCH correlation involving a C_q and CH carbon can be identified at 113.61 and 133.19 ppm, respectively. These correspond to carbon 3 and 4 of the benzene-like moiety introduced on the FK506 basic structure. The associated proton shift of H-4 is 7.38 ppm, and features a triplet multiplicity. The latter at first appears odd, since only a doublet is expected. A strong COSY correlation is visible to the aromatic proton centered at 6.88 ppm, which therefore corresponds to proton 5, the second proton of the aromatic cycle. The latter signal shows a double doublet-like structure, which is also not expected (Figure III.24, left).

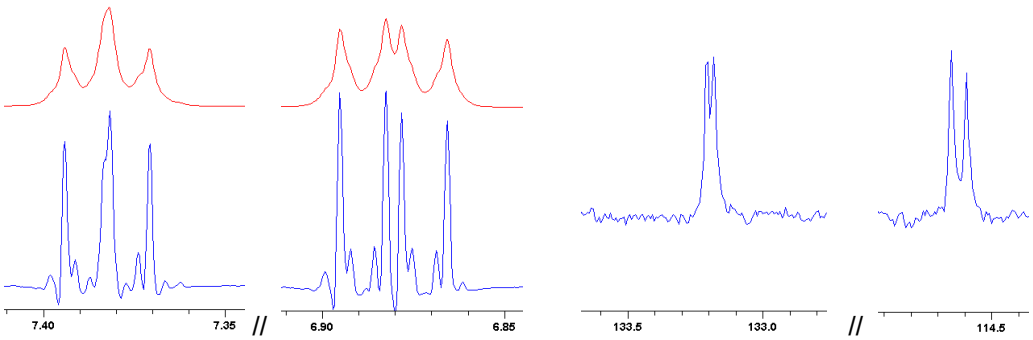


Figure III.24: Details of the ^{1}H (left) and ^{13}C APT (right) spectra showing the proton and carbon resonances corresponding to positions 4 and 5 in **22**. In the ^{1}H spectrum, the bottom spectra represent a resolution enhanced version of the top spectra, obtained through Gaussian window multiplication.

Gaussian resolution enhancement of the 1D proton spectrum shows the presence of two equally intense doublets with overlapping lines, causing the apparent triplet in the regularly processed 1D proton spectrum. Also, the ^{13}C APT shows that the carbon resonances associated with H-4 and H-5 each consist of two equally intense CH type resonances (Figure III.24, right). Because the lines are equally intense, one could conclude that the proton and carbon resonances at position 4 and 5 of **22** are non-equivalent, for instance as a result of hindered rotation which results in slow or no exchange on the NMR time scale. However, the carbon at C-6 also appears as two equally intense resonances

ruling out the possibility of slow rotation around the C6-O7 bond as the origin of the non-equivalence.
We noticed that splitting of the expected resonances in two equally intense ones, both in the ^{1}H and ^{13}C spectrum, occurred frequently and that the chemical shift separation between both increased as one progressed towards the macrocycle. Based upon literature reports describing the assignment of FK506, we propose that this phenomenon is explained by the presence of *cis-trans* isomers around the tertiary amide bond, causing two stereoisomers in slow or no exchange.[101] For FK506 in $CDCl_3$, for example, the *cis-trans* ratio comprises 2:1. FK506 was reported as insoluble in DMSO-d_6, thus there is no information on the expected *cis-trans* ratio in DMSO. It would appear that the modification to the FK506 structure as in **22** results in solubility in DMSO-d_6 and a 1:1 ratio of both stereoisomers. However, this needs confirmation via a complete conformational analysis of **22**, which was not attempted here. Unless mentioned otherwise, chemical shifts reported hereafter generally correspond to the midpoint or only one of the closely spaced resonances in the proton and carbon spectra.

Both H-4 and H-5 are connected to a quaternary carbon at 158.85 ppm which represent the C_q at the foot of the alkoxy fragment, i.e. position 6. In the NOESY spectrum H-5 shows strong nOe correlations to protons at 3.95 and 3.85 ppm. These two protons correlate to a single CH_2 type carbon at 65.37 ppm in the multiplicity edited HSQC, identifying the CH_2 at position 8. This assignment is confirmed by long range ^{n}JCH correlations to $C_q(6)$ at 158.85 ppm.

The area of the proton spectrum in the vicinity of the -CH_2-O at position 8 is heavily overlapped, complicating analysis. The TOCSY suggests additional correlations to four other protons at 4.71, 3.69, 2.08 and 1.71 ppm, respectively. Cross-peaks connect the CH_2 at position 8 directly with the resonances at 2.08 and 1.71, identifying the latter as the CH_2 unit at position 9, with its carbon at 30.26 ppm. This carbon also correlates to the protons at position 8. Both CH_2's (8 and 9) are found to correlate in the HMBC to two CH units located at 49.19/3.69 and 49.29/3.73 ppm, respectively. Again, both protons together integrate for a single 1H compared to the aromatic signals, indicating that these should be considered as the same proton at position 10, but in separate isomers. This is confirmed from the TOCSY and COSY. Finally, the -C(9)H_2-C(10)H- fragment shows ^{n}JCH to the keto-carbonylregion, with two peaks at 209.30 and 210.24 ppm, respectively, corresponding to C-11. With this, the macrocycle of the FK506 structure has been reached.

Based upon the analysis, the identity of the modification to the FK506 basic structure could be unambiguously determined, supporting the structure of **22** as presented.

Table III.7: Partial ^{1}H and ^{13}C NMR assignment of alkyne-functionalized FK506 (22) in DMSO-d_6 at 298K (700/176 MHz) according to the labeling scheme above.

Label	^{1}H (ppm)	^{13}C (ppm)	Multiplicity
1	4.00	79.17	d
2	-	83.79	
3	-	113.61	
4_1	7.39	133.21	d
4_2	7.38	133.18	d
5_1	6.89	114.65	d
5_2	6.87	114.59	d
6_1	-	158.89	
6_2	-	158.82	
8	3.93/3.84	65.37	m/m
9	2.08/1.71	30.26	m/m
10_1	3.73	49.29	dt
10_2	3.69	49.19	ddd
11_1	-	210.24	
11_2	-	209.30	
12_1	4.80	122.17	d
12_2	4.70	121.32	m (overlapped)

* the subscripts 1 and 2 refer to the two different rotamers present in solution. The label was attributed based upon the relative position of the peaks (highest ppm first), not by assignment to a specific rotamer.

FK506-MFC (23): Azide **6a** (47 mg, 59 µmol) was taken up in a mixture of TFA and CH_2Cl_2 (1.5 mL, 1:1, v/v) and stirred for 40 min at RT. The reaction mixture was then taken to dryness, coevaporated twice with toluene and concentrated under high vacuum for 1 h. The residue was taken up in a mixture of water and *tert*-butanol (1 mL, 1:1, v/v) and alkyne **22** (30 mg, 33 µmol), $CuSO_4$ (33 µL, 1.0M, 1.0 eq.) and Na ascorbate (165 µL, 1.0M, 5.0 eq.) were added. The resulting reaction mixture was stirred for 16 h at RT. The solution was concentrated *in vacuo* and the residue was purified by preparative RP-HPLC (10-90% MeCN) yielding the title compound (18 mg, 11 µmol, 33%) as a pale yellow amorphous solid. HRMS calcd. for $C_{83}H_{120}N_{13}O_{22}$ $[M+H]^+$: 1650.8665, found: 1650.9032.

III.5.2. Molecular biology

Performed by the Cytokine Receptor Laboratory, Department of Medical Protein Research, VIB, Ghent & Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University (Prof. Jan Tavernier, Dr. Sam Lievens).

The pCLL-eDHFR receptor-DHFR vector was cloned by transferring the *E. coli* DHFR insert from pSEL-eDHFR[84] into the pCLL backbone[102] using SacI and NotI restriction sites. The pCLG-eDHFR receptor-DHFR plasmid was similarly produced by transferring the *E.* coli DHFR insert into the pCLG backbone.[103] The empty prey construct pMG2 has been previously described.[88] The FKBP12 prey construct pMG2-FKBP12 was generated by amplifying the FKBP12 coding sequence and cloning it into the EcoRI and NotI restriction sites of pMG2.[104] The prey plasmids pMG1-EFHA1, pMG1-ESR1 and pMG1-TTK were created by Gateway transfer of the full size EFHA1, ESR1 and TTK ORFs, obtained as an entry clone in the hORFeome collection,[105] into the Gateway compatible pMG1 prey destination vector as described earlier.[88]

Cells were transfected using a standard calcium phosphate protocol as previously described.[88] In binary MASPIT assays, HEK293T cells were seeded in black tissue-culture treated 96-well plates at 10.000 cells/well in 100 µL culture medium (DMEM supplemented with 10% fetal calf serum), and grown at 37°C, 8% CO_2. Twenty-four hours later, cells were transfected with combinations of receptor-DHFR and prey constructs and the pXP2d2-rPAP1-luciferase reporter. Twenty-four hours after transfection, cells were either left unstimulated or treated with 100 ng/mL leptin, with or without addition of bait fusion compound. Another twenty-four hours later, luciferase activity was assayed using the Luciferase Assay System kit (Promega).

Microtiterplates (384-well) for array screening, containing dried mixtures of prey and pXP2d2-rPAP1-luciferase reporter plasmids together with transfection reagent and additional components, were prepared as described, providing quadruplicate wells for each prey.[88] The collection used in this chapter contained a subset of 1879 full-length human ORF preys selected from the human ORFeome collection (Full list available in Supporting Information Table 1 of reference [105]). To screen the plates, HEK293T cells were first seeded in T175 flasks at a density of $7 \cdot 10^6$ cells/flask in 35 mL culture medium. After twenty-four hours, cells were transfected with the pCLL-eDHFR plasmid. Twenty-four hours after transfection, cells were detached and added to the array screening plates at 5.000 cells/well in 15 µL medium. After twenty-four hours, duplicate wells were supplemented with 15 µL MTX-FK506 containing medium (final concentration 1 µM) or with 15 µL medium containing MTX-FK506 and leptin (final concentration 1 µM and 100 ng/mL, respectively). Non-regiomeric MTX-FK506 conjugate was used in this experiment. Twenty-four hours after stimulation, luciferase activity was measured. Average values of the duplicate wells were normalized for the plate median value. Filters for data cleanup and to remove aspecifically binding preys were set as earlier described.[88]

References

[1] C. K. Osborne, *N. Engl. J. Med.* **1998**, *339*, 1609-1618.
[2] V. C. Jordan, *Nat. Rev. Drug Discov.* **2003**, *2*, 205-213.
[3] M. D. Johnson, H. Zuo, K. H. Lee, J. P. Trebley, J. M. Rae, R. V. Weatherman, Z. Desta, D. A. Flockhart, T. C. Skaar, *Breast Cancer Res. Treat.* **2004**, *85*, 151-159.
[4] S. Gauthier, B. Caron, J. Cloutier, Y. L. Dory, A. Favre, D. Larouche, J. Mailhot, C. Ouellet, A. Schwerdtfeger, G. Leblanc, C. Martel, J. Simard, Y. Mérand, A. Bélanger, C. Labrie, F. Labrie, *J. Med. Chem.* **1997**, *40*, 2117-2122.
[5] P. Y. Maximov, C. B. Myers, R. F. Curpan, J. S. Lewis-Wambi, V. C. Jordan, *J. Med. Chem.* **2010**, *53*, 3273-3283.
[6] M. E. Lieberman, J. Gorski, V. C. Jordan, *J. Biol. Chem.* **1983**, *258*, 4741-4745.
[7] S. Mosselman, J. Polman, R. Dijkema, *FEBS Lett.* **1996**, *392*, 49-53.
[8] E. J. Folkerd, M. Dowsett, *J. Clin. Oncol.* **2010**, *28*, 4038-4044.
[9] W. Zhou, J. M. Slingerland, *Nat. Rev. Cancer* **2014**, *14*, 26-38.
[10] M. Droog, K. Beelen, S. Linn, W. Zwart, *Eur. J. Pharmacol.* **2013**, *717*, 47-57.
[11] J. Frasor, J. M. Danes, B. Komm, K. C. N. Chang, C. R. Lyttle, B. S. Katzenellenbogen, *Endocrinology* **2003**, *144*, 4562-4574.
[12] E. P. Gelmann, *J. Natl. Cancer Inst.* **1996**, *88*, 224-226.
[13] S. Mandlekar, A. N. T. Kong, *Apoptosis* **2001**, *6*, 469-477.
[14] A. A. Colletta, J. R. Benson, M. Baum, *Breast Cancer Res. Treat.* **1994**, *31*, 5-9.
[15] Y. Kang, R. Cortina, R. R. Perry, *J. Natl. Cancer Inst.* **1996**, *88*, 279-284.
[16] C. Ferlini, G. Scambia, M. Marone, M. Distefano, C. Gaggini, G. Ferrandina, A. Fattorossi, G. Isola, P. B. Panici, S. Mancuso, *Br. J. Cancer* **1999**, *79*, 257-263.
[17] S. Mabuchi, M. Ohmichi, A. Kimura, Y. Ikebuchi, K. Hisamoto, E. Arimoto-Ishida, Y. Nishio, K. Takahashi, K. Tasaka, Y. Murata, *Endocrinology* **2004**, *145*, 1302-1313.
[18] Y. Nagahara, I. Shiina, K. Nakata, A. Sasaki, T. Miyamoto, M. Ikekita, *Cancer Sci.* **2008**, *99*, 608-614.
[19] A. K. Shiau, D. Barstad, P. M. Loria, L. Cheng, P. J. Kushner, D. A. Agard, G. L. Greene, *Cell* **1998**, *95*, 927-937.
[20] A. M. Brzozowski, A. C. W. Pike, Z. Dauter, R. E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G. L. Greene, J. A. Gustafsson, M. Carlquist, *Nature* **1997**, *389*, 753-758.
[21] V. C. Jordan, B. Gosden, *Mol. Cell. Endocrinol.* **1982**, *27*, 291-306.
[22] D. W. Robertson, J. A. Katzenellenbogen, J. R. Hayes, B. S. Katzenellenbogen, *J. Med. Chem.* **1982**, *25*, 167-171.
[23] G. Dayan, M. Lupien, A. Auger, S. I. Anghel, W. Rocha, S. Croisetiere, J. A. Katzenellenbogen, S. Mader, *Mol. Pharmacol.* **2006**, *70*, 579-588.
[24] P. J. Burke, T. H. Koch, *J. Med. Chem.* **2004**, *47*, 1193-1206.
[25] E. C. Dreaden, S. C. Mwakwari, Q. H. Sodji, A. K. Oyelere, M. A. El-Sayed, *Bioconjugate Chem.* **2009**, *20*, 2247-2253.
[26] B. E. Gryder, M. K. Rood, K. A. Johnson, V. Patil, E. D. Raftery, L.-P. D. Yao, M. Rice, B. Azizi, D. F. Doyle, A. K. Oyelere, *J. Med. Chem.* **2013**, *56*, 5782-5796.
[27] V. Sreekanth, S. Bansal, R. K. Motiani, S. Kundu, S. K. Muppu, T. D. Majumdar, K. Panjamurthy, S. Sengupta, A. Bajaj, *Bioconjugate Chem.* **2013**, *24*, 1468-1484.
[28] S. B. Chen, Q. S. Zhang, X. Wu, P. G. Schultz, S. Ding, *J. Am. Chem. Soc.* **2004**, *126*, 410-411.
[29] L. Anastasia, M. Sampaolesi, N. Papini, D. Oleari, G. Lamorte, C. Tringali, E. Monti, D. Galli, G. Tettamanti, G. Cossu, B. Venerando, *Cell Death Differ.* **2006**, *13*, 2042-2051.
[30] L. Anastasia, G. Pelissero, B. Venerando, G. Tettamanti, *Cell Death Differ.* **2010**, *17*, 1230-1237.
[31] S. B. Chen, S. C. Takanashi, Q. S. Zhang, W. Xiong, S. T. Zhu, E. C. Peters, S. Ding, P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10482-10487.
[32] A. F. Straight, C. M. Field, T. J. Mitchison, *Mol. Biol. Cell* **2005**, *16*, 193-201.
[33] M. Vicente-Manzanares, X. F. Ma, R. S. Adelstein, A. R. Horwitz, *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 778-790.
[34] W. Kolch, *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 827-837.
[35] J. A. Engelman, J. Luo, L. C. Cantley, *Nat. Rev. Genet.* **2006**, *7*, 606-619.
[36] A. M. D'Alise, G. Amabile, M. Iovino, F. P. Di Giorgio, M. Bartiromo, F. Sessa, F. Villa, A. Musacchio, R. Cortese, *Mol. Cancer Ther.* **2008**, *7*, 1140-1149.

[37] S. Santaguida, A. Tighe, A. M. D'Alise, S. S. Taylor, A. Musacchio, *J. Cell Biol.* **2010**, *190*, 73-87.
[38] C. Ditchfield, V. L. Johnson, A. Tighe, R. Ellston, C. Haworth, T. Johnson, A. Mortlock, N. Keen, S. S. Taylor, *J. Cell Biol.* **2003**, *161*, 267-280.
[39] N. Keen, S. Taylor, *Nat. Rev. Cancer* **2004**, *4*, 927-936.
[40] G. Amabile, A. M. D'Alise, M. Iovino, P. Jones, S. Santaguida, A. Musacchio, S. Taylor, R. Cortese, *Cell Death Differ.* **2008**, *16*, 321-330.
[41] X. D. Liu, M. Winey, *Annu. Rev. Biochem..* **2012**, *81*, 561-585.
[42] F. Girdler, K. E. Gascoigne, P. A. Eyers, S. Hartmuth, C. Crafter, K. M. Foote, N. J. Keen, S. S. Taylor, *J. Cell Sci.* **2006**, *119*, 3664-3675.
[43] S. C. Hua, T. C. Chang, H. R. Chen, C. H. Lu, Y. W. Liu, S. H. Chen, H. I. Yu, Y. P. Chang, Y. R. Lee, *Pharm. Res.* **2012**, *29*, 1990-2005.
[44] All FK506-related chemistry described in this thesis was performed by Dr. M. D. P. Risseeuw.
[45] L. J. Scott, K. McKeage, S. J. Keam, G. L. Plosker, *Drugs* **2003**, *63*, 1247-1297.
[46] M. W. Harding, A. Galat, D. E. Uehling, S. L. Schreiber, *Nature* **1989**, *341*, 758-760.
[47] S. L. Schreiber, *Science* **1991**, *251*, 283-287.
[48] J. Liu, J. D. Farmer, W. S. Lane, J. Friedman, I. Weissman, S. L. Schreiber, *Cell* **1991**, *66*, 807-815.
[49] E. J. Licitra, J. O. Liu, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 12817-12821.
[50] M. N. Pruschy, D. M. Spencer, T. M. Kapoor, H. Miyake, G. R. Crabtree, S. L. Schreiber, *Chem. Biol.* **1994**, *1*, 163-172.
[51] J. P. Griffith, J. L. Kim, E. E. Kim, M. D. Sintchak, J. A. Thomson, M. J. Fitzgibbon, M. A. Fleming, P. R. Caron, K. Hsiao, M. A. Navia, *Cell* **1995**, *82*, 507-522.
[52] P. A. Clemons, B. G. Gladstone, A. Seth, E. D. Chao, M. A. Foley, S. L. Schreiber, *Chem. Biol.* **2002**, *9*, 49-61.
[53] A. Corsini, F. M. Maggi, A. L. Catapano, *Pharmacol. Res.* **1995**, *31*, 9-27.
[54] J. A. Tobert, *Nat. Rev. Drug Discov.* **2003**, *2*, 517-526.
[55] H. V. Ganga, H. B. Slim, P. D. Thompson, *Am. Heart J.* **2014**, *168*, 6-15.
[56] M. A. Omar, J. P. Wilson, *Ann. Pharmacother.* **2002**, *36*, 288-295.
[57] C. M. Ballantyne, A. Corsini, M. H. Davidson, H. Holdaas, T. A. Jacobson, E. Leitersdorf, W. Marz, J. P. D. Reckless, E. A. Stein, *Arch. Intern. Med.* **2003**, *163*, 553-564.
[58] L. R. Pierce, D. K. Wysowski, T. P. Gross, *J. Am. Med. Assoc.* **1990**, *264*, 71-75.
[59] G. J. Magarian, L. M. Lucas, C. Colley, *Arch. Intern. Med.* **1991**, *151*, 1873-1874.
[60] A. Corsini, S. Bellosta, R. Baetta, R. Fumagalli, R. Paoletti, F. Bernini, *Pharmacol. Ther.* **1999**, *84*, 413-428.
[61] J. A. Staffa, J. Chang, L. Green, *N. Engl. J. Med.* **2002**, *346*, 539-540.
[62] *Lancet* **2003**, *361*, 793-793.
[63] M. Evans, A. Rees, *Drug Saf.* **2002**, *25*, 649-663.
[64] D. Mosshammer, E. Schaeffeler, M. Schwab, K. Morike, *Br. J. Clin. Pharmacol.* **2014**, *78*, 454-466.
[65] J. K. Liao, U. Laufs, in *Annual Review of Pharmacology and Toxicology, Vol. 45*, Annual Reviews, Palo Alto, **2005**, pp. 89-118.
[66] C. Y. Wang, P. Y. Liu, J. K. Liao, *Trends Mol. Med* **2008**, *14*, 37-44.
[67] F. L. Zhang, P. J. Casey, *Annu. Rev. Biochem.* **1996**, *65*, 241-269.
[68] P. A. Konstantinopoulos, M. V. Karamouzis, A. G. Papavassiliou, *Nat. Rev. Drug Discov.* **2007**, *6*, 540-555.
[69] S. M. Ostrowski, B. L. Wilkinson, T. E. Golde, G. Landreth, *J. Biol. Chem.* **2007**, *282*, 26832-26844.
[70] S. Rao, D. C. Porter, X. M. Chen, T. Herliczek, M. Lowe, K. Keyomarsi, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 7797-7802.
[71] J. Greenwood, L. Steinman, S. S. Zamvil, *Nat. Rev. Immunol.* **2006**, *6*, 358-370.
[72] G. Mundy, R. Garrett, S. Harris, J. Chan, D. Chen, G. Rossini, B. Boyce, M. Zhao, G. Gutierrez, *Science* **1999**, *286*, 1946-1949.
[73] K. C. Hsieh, C. L. Kao, C. W. Feng, Z. H. Wen, H. F. Chang, S. C. Chuang, G. J. Wang, M. L. Ho, S. M. Wu, J. K. Chang, H. T. Chen, *Org. Lett.* **2014**, *16*, 4376-4379.
[74] N. P. Whitehead, M. J. Kim, K. L. Bible, M. E. Adams, S. C. Froehner, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, doi: 10.1073/pnas.1509536112.
[75] E. S. Istvan, J. Deisenhofer, *Science* **2001**, *292*, 1160-1164.
[76] G. E. Stokker, W. F. Hoffman, A. W. Alberts, E. J. Cragoe, A. A. Deana, J. L. Gilfillan, J. W. Huff, F. C. Novello, J. D. Prugh, *J. Med. Chem.* **1985**, *28*, 347-358.
[77] C. Chidley, H. Haruki, M. G. Pedersen, E. Muller, K. Johnsson, *Nat. Chem. Biol.* **2011**, *7*, 375-383.

[78] J. Drews, *Science* **2000**, *287*, 1960-1964.
[79] J. P. Overington, B. Al-Lazikani, A. L. Hopkins, *Nat. Rev. Drug Discov.* **2006**, *5*, 993-996.
[80] K. Szardenings, B. Li, L. Ma, M. Wu, *Drug Discovery Today: Technol.* **2004**, *1*, 9-15.
[81] T. Ito, H. Ando, T. Suzuki, T. Ogura, K. Hotta, Y. Imamura, Y. Yamaguchi, H. Handa, *Science* **2010**, *327*, 1345-1350.
[82] G. C. Terstappen, C. Schlupen, R. Raggiaschi, G. Gaviraghi, *Nat. Rev. Drug Discov.* **2007**, *6*, 891-903.
[83] S. Eyckerman, A. Verhee, J. V. der Heyden, I. Lemmens, X. V. Ostade, J. Vandekerckhove, J. Tavernier, *Nat. Cell Biol.* **2001**, *3*, 1114-1119.
[84] M. Caligiuri, L. Molz, Q. Liu, F. Kaplan, J. P. Xu, J. Z. Majeti, R. Ramos-Kelsey, K. Murthi, S. Lievens, J. Tavernier, N. Kley, *Chem. Biol.* **2006**, *13*, 711-722.
[85] K. Venkatesan, J.-F. Rual, A. Vazquez, U. Stelzl, I. Lemmens, T. Hirozane-Kishikawa, T. Hao, M. Zenkner, X. Xin, K.-I. Goh, M. A. Yildirim, N. Simonis, K. Heinzmann, F. Gebreab, J. M. Sahalie, S. Cevik, C. Simon, A.-S. de Smet, E. Dann, A. Smolyar, A. Vinayagam, H. Yu, D. Szeto, H. Borick, A. Dricot, N. Klitgord, R. R. Murray, C. Lin, M. Lalowski, J. Timm, K. Rau, C. Boone, P. Braun, M. E. Cusick, F. P. Roth, D. E. Hill, J. Tavernier, E. E. Wanker, A.-L. Barabasi, M. Vidal, *Nat. Meth.* **2009**, *6*, 83-90.
[86] P. Braun, M. Tasan, M. Dreze, M. Barrios-Rodiles, I. Lemmens, H. Yu, J. M. Sahalie, R. R. Murray, L. Roncari, A.-S. de Smet, K. Venkatesan, J.-F. Rual, J. Vandenhaute, M. E. Cusick, T. Pawson, D. E. Hill, J. Tavernier, J. L. Wrana, F. P. Roth, M. Vidal, *Nat. Meth.* **2009**, *6*, 91-97.
[87] X. Yang, J. S. Boehm, X. Yang, K. Salehi-Ashtiani, T. Hao, Y. Shen, R. Lubonja, S. R. Thomas, O. Alkan, T. Bhimdi, T. M. Green, C. M. Johannessen, S. J. Silver, C. Nguyen, R. R. Murray, H. Hieronymus, D. Balcha, C. Fan, C. Lin, L. Ghamsari, M. Vidal, W. C. Hahn, D. E. Hill, D. E. Root, *Nat. Meth.* **2011**, *8*, 659-661.
[88] S. Lievens, N. Vanderroost, J. Van der Heyden, V. Gesellchen, M. Vidal, J. Tavernier, *J. Proteome Res.* **2009**, *8*, 877-886.
[89] X. E. Wells, V. J. Bender, C. L. Francis, H. M. He-Williams, M. K. Manthey, M. J. Moghaddam, W. G. Reilly, R. G. Whittaker, *Drug Dev. Res.* **1999**, *46*, 302-308.
[90] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem.-Int. Edit.* **2002**, *41*, 2596-2599.
[91] S. S. Iyer, A. S. Anderson, S. Reed, B. Swanson, J. G. Schmidt, *Tet. Lett.* **2004**, *45*, 4285-4288.
[92] C. L. Francis, Q. Yang, N. K. Hart, F. Widmer, M. K. Manthey, H. M. He-Williams, *Aust. J. Chem.* **2002**, *55*, 635-645.
[93] F. D. Bellamy, K. Ou, *Tet. Lett.* **1984**, *25*, 839-842.
[94] The trimethylsilyl protecting group on the acetylene group is retained, as 4-ethynylphenol (contrary to its methoxy congener 4-ethynylanisole) is highly unstable. Alkylation of the phenol moiety to an ether group eliminates the instability, thus enabling removal of the TMS group.
[95] M. Krause, X. Ligneau, H. Stark, M. Garbarg, J.-C. Schwartz, W. Schunack, *J. Med. Chem.* **1998**, *41*, 4171-4176.
[96] Both fluoride sources were necessary, as no single set of conditions in our hands proved successful to remove both the alkynylsilane and the silylether groups.
[97] The non-regiopure conjugate was generated analogously to its regioselective γ congener by only using a non-regioselective MTX-PEG azide, prepared as described in Scheme III.2.
[98] Design of a fluorescence-activated cell-sorting-based mammalian protein-protein interaction trap: S. Lievens, J. Van der Heyden, E. Vertenten, J. Plum, J. Vandekerckhove, J. Tavernier, *Methods Mol. Biol.* **2004**, *263*, 293-310.
[99] T. R. Chan, R. Hilgraf, K. B. Sharpless, V. V. Fokin, *Org. Lett.* **2004**, *6*, 2853-2855.
[100] The molar amount of acid is a rough estimation given the significant amount of silica gel remaining in this material after chromatography.
[101] D. F. Mierke, P. Schmieder, P. Karuso, H. Kessler, *Helv. Chim. Acta* **1991**, *74*, 1027-1047.
[102] P. Ulrichts, F. Peelman, R. Beyaert, J. Tavernier, *FEBS Lett.* **2007**, *581*, 629-636.
[103] J. P. Yan, Q. Li, S. Lievens, J. Tavernier, J. X. You, *J. Virol.* **2010**, *84*, 76-87.
[104] T. Montoye, I. Lemmens, D. Catteeuw, S. Eyckerman, J. Tavernier, *Blood* **2005**, *105*, 4264-4271.
[105] P. Lamesch, N. Li, S. Milstein, C. Y. Fan, T. Hao, G. Szabo, Z. J. Hu, K. Venkatesan, G. Bethel, P. Martin, J. Rogers, S. Lawlor, S. McLaren, A. Dricot, H. Borick, M. E. Cusick, J. Vandenhaute, I. Dunham, D. E. Hill, M. Vidal, *Genomics* **2007**, *89*, 307-315.

CHAPTER IV

TMP- AND SNAP-TAG APPROACH

Alternative Reagents for Methotrexate as Immobilizing Anchor Moieties in the Optimization of MASPIT: Synthesis and Biological Evaluation

The content of this chapter was derived from:
D. J. H. De Clercq, M. D. P. Risseeuw, I. Karalic, A.-S. De Smet, D. Defever, J. Tavernier, S. Lievens, S. Van Calenbergh, *ChemBioChem* **2015**, *16*, 834-843.

S. Lievens, S. Gerlo, I. Lemmens, D. J. H. De Clercq, M. D. P. Risseeuw, N. Vanderroost, A.-S. De Smet, E. Ruyssinck, E. Chevet, S. Van Calenbergh, J. Tavernier, *Mol. Cell. Proteomics* **2014**, *13*, 3332-3342.

IV. Alternative Reagents for Methotrexate as Immobilizing Anchor Moieties in the Optimization of MASPIT: Synthesis and Biological Evaluation

IV.1. Introduction

Methods that allow high-throughput identification of the cellular targets of bioactive small molecules are invaluable assets in pharmaceutical research. They are useful in mechanism-of-action studies of hits identified by phenotypic screening, which is increasingly being applied in both academic and industrial programs.[1] Additionally, they can uncover unexpected targets of established drugs that could contribute to their therapeutic efficacy or cause unwanted side effects (polypharmacology).[2] Finally, such methods can also lead to the identification of new therapeutic applications of existing drugs within the scope of drug repositioning projects.[3] Over the past decade, numerous case studies within the target profiling field that apply, for example, activity-based protein profiling (ABPP)[4] or compound-centric chemical proteomics methods (CCCP)[5] have proved successful. However, despite these success stories, target deconvolution often remains an important bottleneck in drug discovery research, as a generally applicable methodology is still lacking.[6]

MASPIT (mammalian small molecule-protein interaction trap) is the three-hybrid component of the MAPPIT technology platform,[7],[8] which enables the identification of interactions between small organic compounds and their cytosolic target proteins in living human cells (Figure IV.1).[9] A prerequisite for successful MASPIT analysis is the availability of appropriate synthetic probes. We previously presented a scalable synthesis of a versatile methotrexate (MTX) reagent that allows the rapid γ-selective conjugation to alkyne-functionalized bioactive small molecules to yield MTX fusion compounds (MFCs) appropriate for MASPIT.[10] Here, we take the next step and discuss our efforts to optimize the MASPIT system's sensitivity based on chemical dimerizers with tamoxifen (TAM) as the model 'bait'. The latter is a selective estrogen receptor (ER) modulator that has been part of the standard therapy for ER-positive breast cancer treatment since the 1970s.[11] However, TAM has been reported to induce apoptosis even in ER-negative cancer cells, thus suggesting that it can also operate by modulating alternative targets.[12] As yet, the exact mechanism of action underlying the apparent promiscuity of this blockbuster drug remains elusive. Taking this together and building on prior experience in constructing and evaluating various TAM-MFCs,[10] we judged this bait would be a particularly interesting test case for our optimization work.

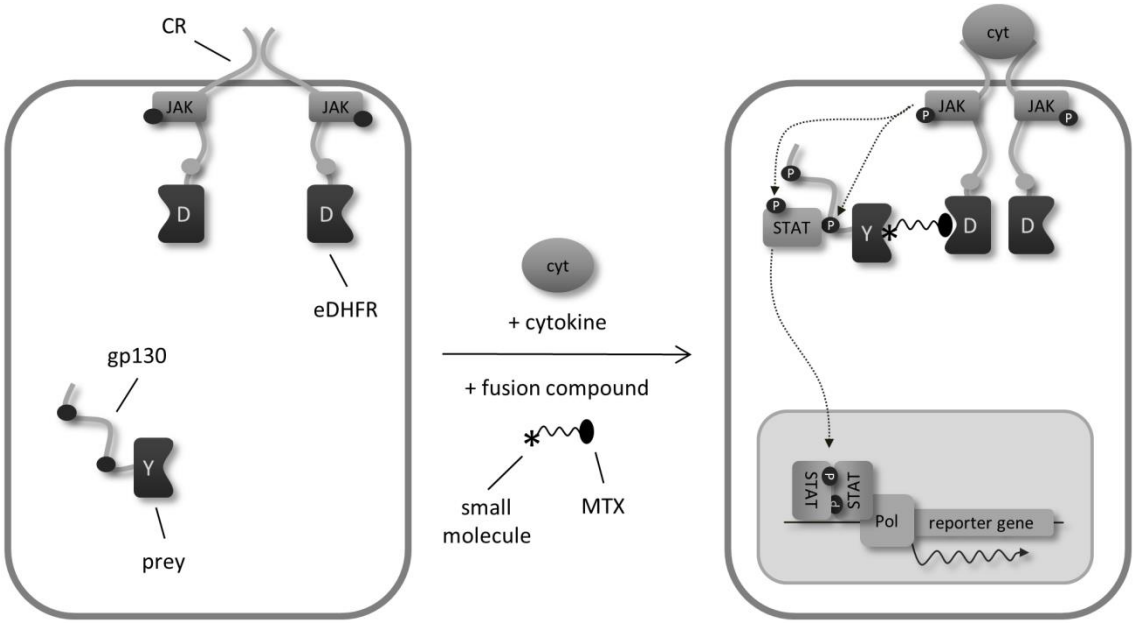


Figure IV.1: Outline of the MASPIT system. *E. coli* dihydrofolate reductase (eDHFR) is fused to a cytokine receptor (CR) that is rendered signaling-deficient by mutating STAT3 recruitment sites in its cytoplasmic tail (grey dots). A prey protein is tethered to a gp130 CR fragment containing functional STAT3 transcription factor docking sites (black dots). When a fusion compound consisting of a small molecule of interest (asterisk) coupled to methotrexate (MTX) is added to the cells, MTX binds to eDHFR, resulting in the compound of interest being displayed as bait. Upon administration of the appropriate cytokine ligand (cyt), the CR-eDHFR chimeric receptor undergoes a conformational change, activating the associated JAK2 kinases through crossphosphorylation (P). Interaction between the small-molecule bait and the prey-gp130 fusion protein brings the latter into proximity of the activated JAK kinases, reconstituting a functional JAK-STAT signaling pathway. Sequential phosphorylation of STAT3 docking sites on the gp130 chain (P), STAT3 recruitment, and STAT3 phosphorylation (P) ultimately leads to activated STAT3 dimers that induce the expression of a luciferase reporter gene coupled to a STAT3-dependent promoter.

First, we tried to circumvent any potential limitations related to the tight binding of MTX to endogenous human dihydrofolate reductase (DHFR), which might titrate out a portion of the fusion compound and induce cellular toxicity through perturbation of the endogenous folate metabolism. As an *Escherichia coli* enzyme is employed in MASPIT, we explored trimethoprim (TMP; Figure IV.2A) as an alternative prokaryote-specific DHFR ligand. Whereas MTX binds both prokaryotic and mammalian DHFR with similar affinity, TMP displays a 12 000-fold binding preference for *E. coli* over human DHFR (K_i = 80 pM vs. 960 nM), thus reflecting its use as a selective antibiotic in the clinic.[13] Cornish and Sheetz have previously demonstrated the compatibility of the TMP-tag with mammalian systems for intracellular live-cell imaging.[14]

Furthermore, so as to stabilize the ternary complex (CR-fusion compound-prey chimera) in order to improve the system's sensitivity, we introduced the concept of covalent bonding into the MASPIT assay, which currently relies on reversible interactions on both ends of the MFCs. As a starting point, we selectively and covalently immobilized the fusion compound to the CR by using a SNAP-tag-based system.[15] This strategy is centered around the human DNA-repair protein O^6-alkylguanine-DNA alkyltransferase (hAGT). Johnsson and co-workers have exploited the low substrate specificity of this enzyme to covalently label SNAP-tag-fused proteins *in vivo* with a ligand of interest by conjugating the latter to the *para* position of O^6-benzylguanine (BG).[15] By fusing hAGT to the cytoplasmic domain

of the CR and synthesizing the appropriate BG fusion compound (BGFC), we planned to present a covalently coupled bait small molecule (Figure IV.2B).

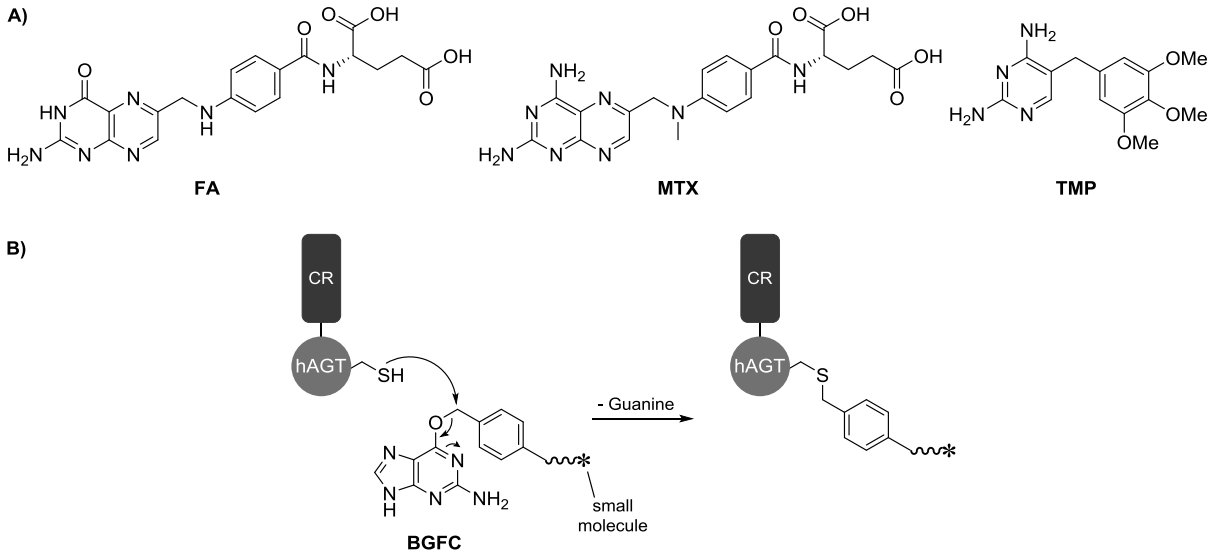
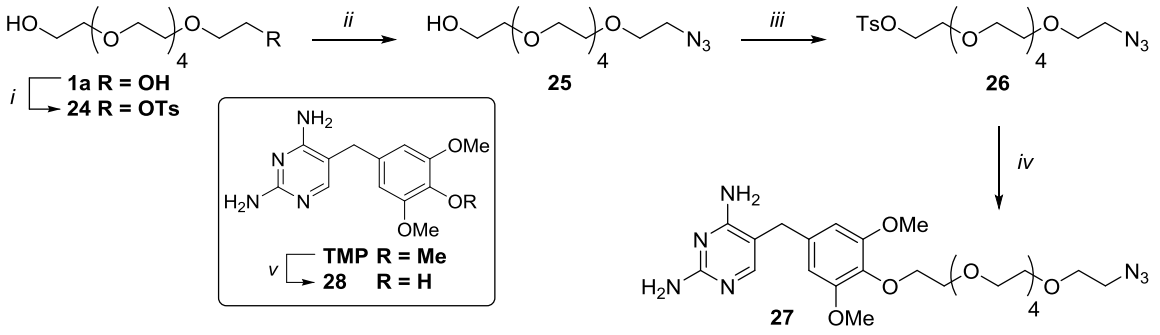


Figure IV.2: A) Chemical structures of folic acid (FA) and the DHFR inhibitors methotrexate and trimethoprim. **B)** Mechanism of hAGT-mediated covalent immobilization of the O^6-benzylguanine fusion compound (BGFC) to the cytokine receptor (CR).

IV.2. Results and Discussion

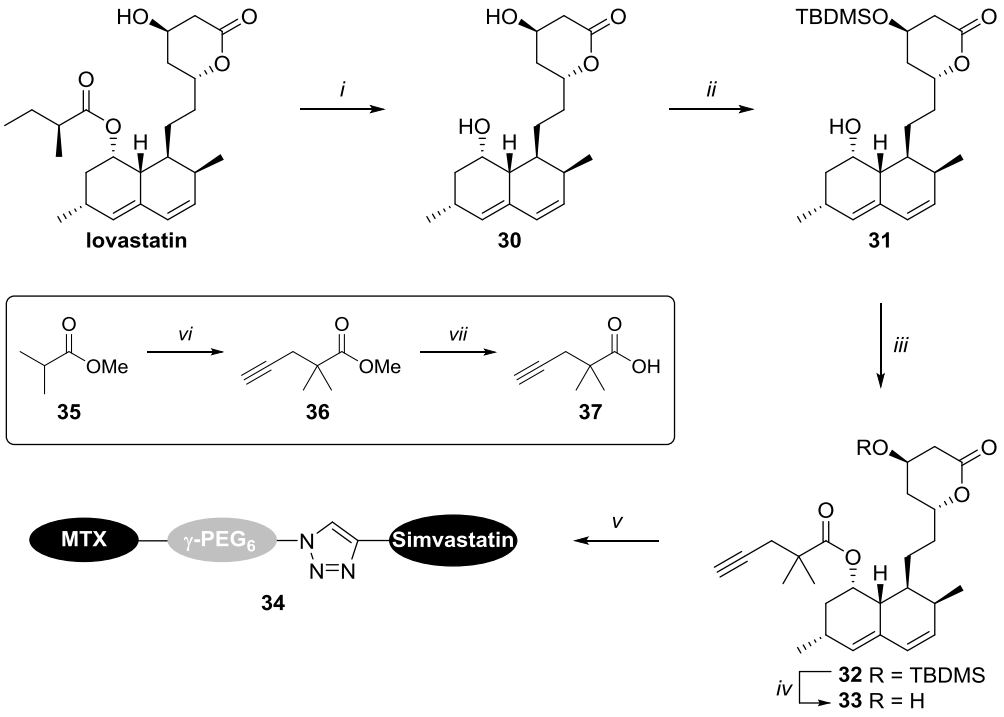
IV.2.1. TMP-tag approach

We sought to develop a scalable synthesis of a versatile TMP reagent with an azide ligation handle as an alternative for the earlier MTX congener. A synthetic route towards a first-generation TMP-azido reagent is depicted in Scheme IV.1. The synthesis began with the generation of a tosyl/azido bifunctionalized hexa(ethylene glycol) linker (**26**).[16] This spacer was used to alkylate phenol **28**, which was obtained by acidic hydrolysis of TMP,[17] to afford the desired ligation handle **27**. However, purification of **27** required RP-HPLC to remove a side product formed by alkylation at the benzhydrilic position. Presumably, this double alkylated TMP analogue was formed by oxidative conversion of phenol **28** to a reactive pyrimidine iminoquinone methide intermediate.[18] The latter can be converted through resonance stabilization to the corresponding *para*-quinone methide form, which has a prochiral activated exocyclic methylene group. This side reaction rendered this alkylation unsuitable for scale-up. Therefore, we examined alternative alkylation conditions, varying the base (DBU, Na_2CO_3), leaving group (I, OMs), solvent (DMF), temperature (RT) and reaction time (16-48 h), but none yielded a more favorable regioselectivity profile.



Scheme IV.1: Synthesis of the first-generation TMP-N_3 reagent. [i] TsCl, Et_3N, CH_2Cl_2, 0°C; [ii] NaN_3, DMF, 60°C, 64% two steps; [iii] TsCl, Et_3N, CH_2Cl_2, 0°C, 91%; [iv] **28**, K_2CO_3, acetone, Δ, 39%; [v] first aq. HBr (48%), Δ; then NaOH (50%), 30%.

Subsequently, first-generation trimethoprim fusion compounds (TFCs) **29a-c** were prepared by ligating azido reagent **27** via CuAAC[19] to the alkyne-functionalized tamoxifen,[10] reversine[10] and simvastatin,[20] respectively (Figure IV.3). The synthesis of the first two alkynylated bait analogues was previously described in Chapter 3. The latter was prepared in several steps from lovastatin (Scheme IV.2). *tert*-Butyldimethylsilyloxymonacolin J (**31**) and the required 2,2-dimethylpent-4-ynoic acid **37** were obtained according to a protocol by Sorensen *et al*.[21] and Welsch *et al*.,[22] respectively. Analogously to a procedure by Hong *et al*.[23] the latter acid was converted *in situ* to the corresponding acid bromide, which was used to acylate **31** to afford alkyne **33** after desilylation.



Scheme IV.2: Synthesis of alkyne-functionalized simvastatin and its corresponding MFC. [i] first LiOH, Δ; then CaH_2, toluene, Δ (Soxhlet), 44%; [ii] TBDMSCl, imidazole, DMF, 93%; [iii] **37**, PPh_3, NBS, $PhNMe_2$, CH_2Cl_2, 0°C to RT, 80%; [iv] HCl, THF, dioxane, 0°C to RT, 76%; [v] first **6a**, TFA, CH_2Cl_2; then **33**, $CuSO_4$, Na ascorbate, Et_3N, TBTA, DMF, H_2O, 150°C (μW), 44%; [vi] first LDA, THF, -78°C; then propargyl bromide, -78°C to RT, 71%; [vii] first NaOH, MeOH, 50°C; then HCl, 94%.

MASPIT evaluation of these TMP conjugates (**29a-c**) against their established primary target preys showed decreased EC_{50} values with respect to the corresponding MFCs (Figure IV.3). These data imply a successful improvement in the system's sensitivity by introducing TMP as an alternative immobilizing anchor moiety, despite the fact that it exhibits lower affinity for eDHFR than MTX (K_i = 1.0 pM).[13]

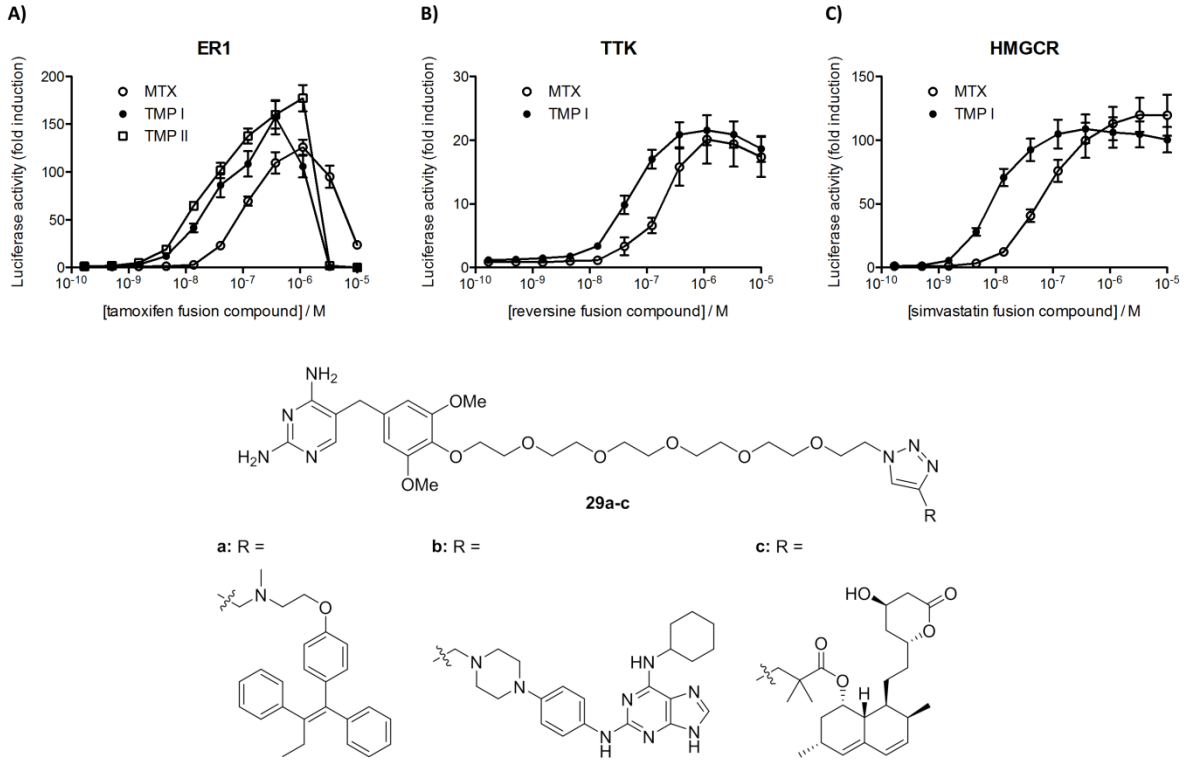
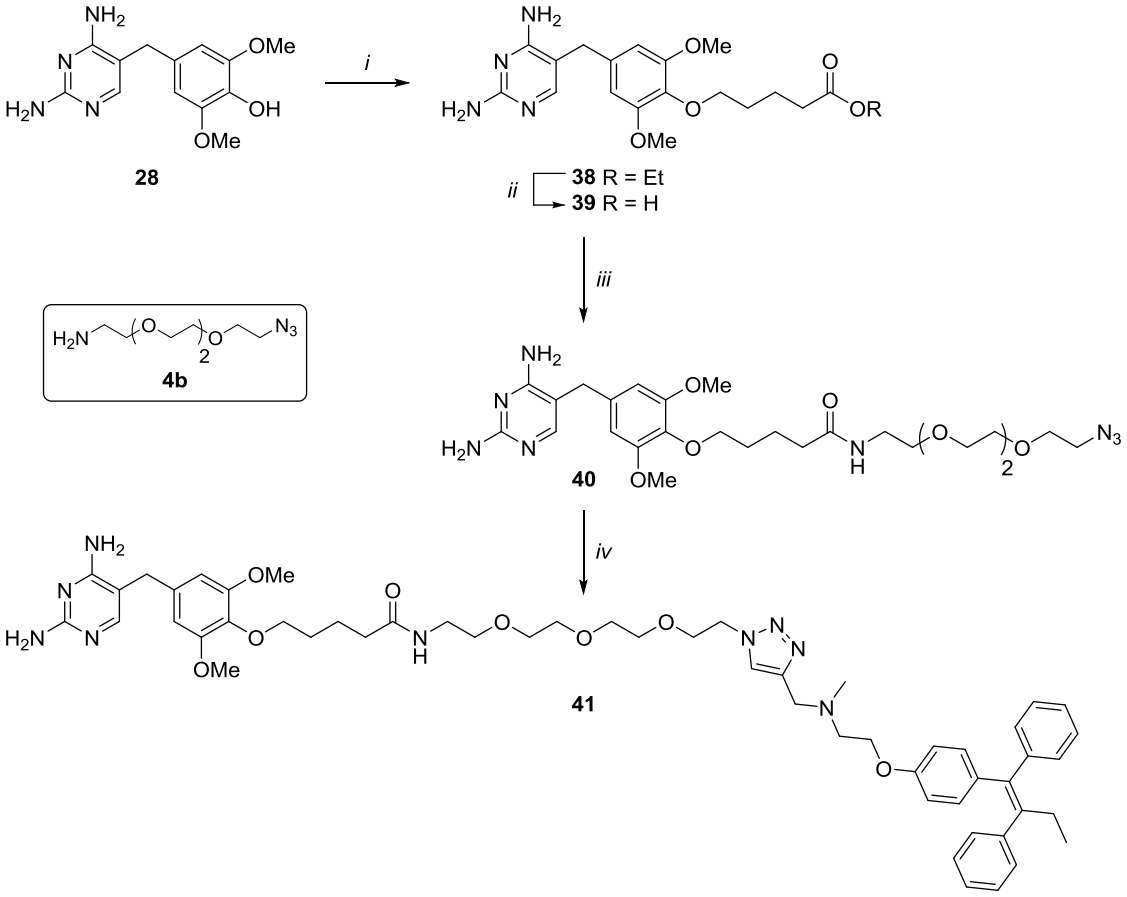


Figure IV.3: Evaluation of small-molecule TFCs 29a-c (TMP I) and 41 (TMP II) against the corresponding MFCs (MTX) for binding to their primary target protein in MASPIT. A) Tamoxifen-ER1; **B)** reversine-TTK; **C)** simvastatin-HMGCR. Luciferase signals are expressed as fold induction relative to a control sample treated with cytokine without fusion compound. In A), the signal decrease at the highest concentrations is due to the cellular toxicity of the tamoxifen fusion compound.

In a next step, we explored a different alkylation strategy starting from phenol **28** (Scheme IV.3). Initial alkylations using ethyl 5-bromovalerate in combination with K_2CO_3 consistently provided an intractable mixture of overalkylated products. Varying the solvent (DMF, acetone), temperature (40°C, Δ) and reaction time (6-60h) had negligible effects on the regioselectivity and reaction progress/yield. However, encouraged by the significant positive trend we observed earlier for **27** upon switching to a softer counterion for the carbonate base (K_2CO_3 vs. Na_2CO_3) and further guided by a procedure by Chen *et al.*,[24] we ultimately opted for Cs_2CO_3 which successfully gave access to acid **39**. The latter was subsequently condensed with PEG_4 azidoamine **4b**[10] in the presence of 2-(2-pyridon-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TPTU) to obtain a second-generation TMP-azido reagent in satisfactory yield on a multigram scale. This reagent has the same spacer length, with respect to the number of atoms and chemical bonds, as the original **27** and was analogously click-coupled to the tamoxifen model bait to generate the second-generation TAM-TFC **41**.



Scheme IV.3: Synthesis of the second-generation TMP-N_3 reagent and TAM-TFC. [i] Ethyl 5-bromovalerate, Cs_2CO_3, DMF, 70°C, 66%; [ii] first NaOH, MeOH; then HCl, 92%; [iii] **4b**, TPTU, Et_3N, DMF, 78%; [iv] alkyne-functionalized tamoxifen,[10] $CuSO_4$, Na ascorbate, TBTA, H_2O/*t*BuOH, 80°C, 56%.

Both generations of TAM-TFCs in MASPIT yielded approximately coinciding curves, again shifted towards lower EC_{50} values compared to the original MFC (Figure IV.3A). Hereby, we demonstrated the biological equivalence of the scalable **40** to **27**, and confirmed the increase in sensitivity for TFCs. Moreover, next to its superior behavior in the MASPIT assay, the newly developed second-generation TMP reagent offers a number of important advantages over the existing MTX anchor from a chemical perspective. The TMP reagent lacks chirality and shows increased solubility, thus making it in general more practical. Furthermore, the reagent has enhanced stability in CuAAC reactions, whereas, under typical conditions, the MTX-azido reagent suffered from significant degradation to the corresponding amine. This is possibly due to residual trifluoroacetic acid in the material from cleavage of the α-*tert*-butylester precursor.[10] In the case of TMP, which does not require these protecting-group manipulations, CuAACs can be performed without the formation of degradation or by-products, thereby facilitating purification of the final conjugates.

Previously, Cornish[25] and Bertozzi[26] obtained contradictory outcomes when comparing the effectiveness of MTX- and TMP-based chemical inducers of dimerization (CIDs) in the same yeast three-hybrid system. For the dexamethasone ligand, Cornish found that the TMP-based CID did not induce transcription activation as efficiently as the corresponding MTX probe, whereas Bertozzi concluded exactly the opposite for the SLF bait (a synthetic analogue of FK506). They postulated that the disparity in activity between the two CID anchors might be attributed, for example, to different

cell-permeability properties. To our knowledge, this aspect has not been experimentally clarified in the context of compound profiling strategies so far. Hence, in an effort to elucidate the origin of the superior performance of TMP fusions in the assay, we studied the uptake of MTX- versus TMP-linked fluorophores in HEK293T cells, the cell line employed in MASPIT. The known MTX-azido and optimized TMP-azido reagents were fused to an alkyne-functionalized BODIPY analogue[27] by CuAAC to yield **42** and **43**, respectively (Figure IV.4).

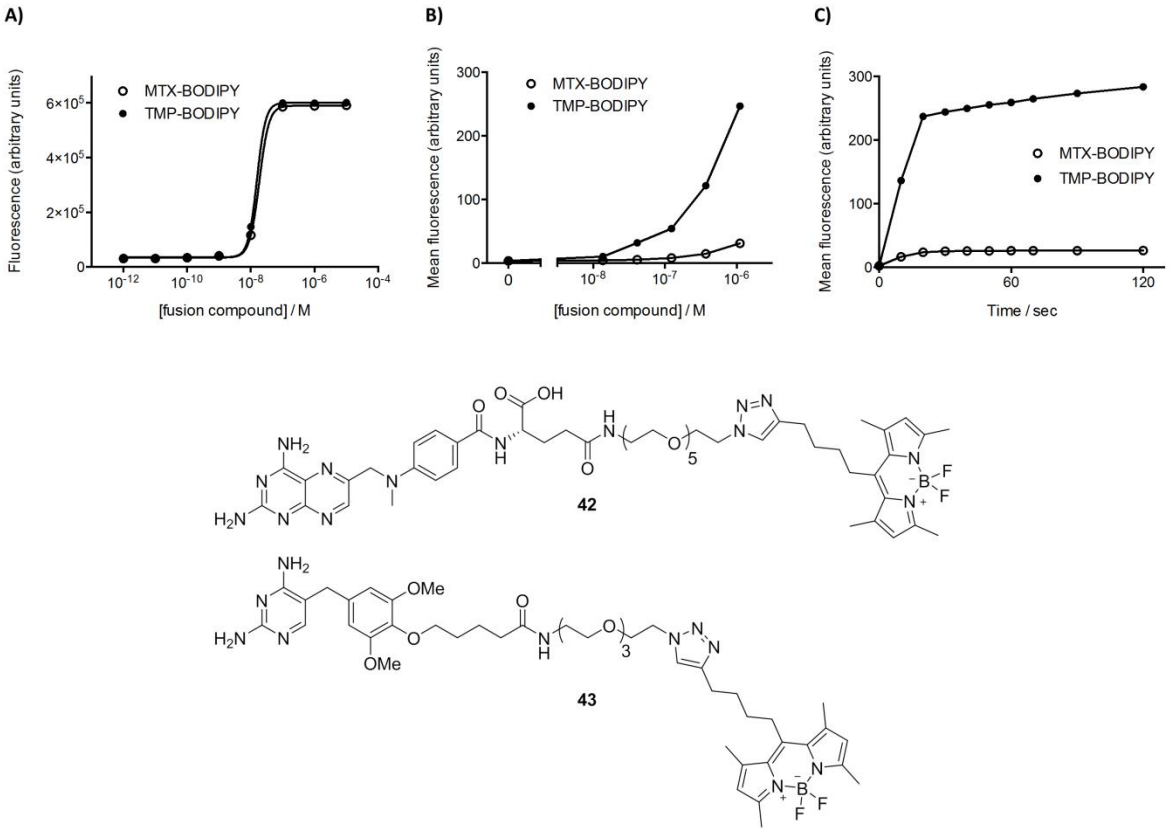


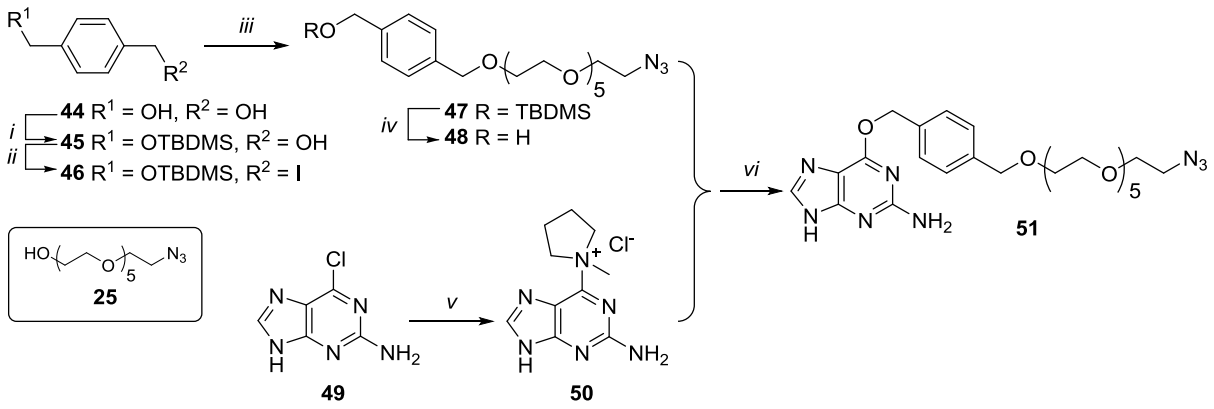
Figure IV.4: Evaluation of the membrane permeability of BODIPY conjugates of MTX (42) and TMP (43). A) Dose-response curves showing the inherent fluorescence of the conjugates. **B)** Dose-response curves of the cellular uptake of the BODIPY conjugates. The graph shows the mean fluorescence as measured by FACS analysis of cells stained with increasing concentrations of the conjugates. **C)** Kinetic analysis of cellular staining with the MTX- or TMP-BODIPY fusion molecules. Mean fluorescence measured by FACS is plotted against time after the addition of 1 µM of either conjugate.

Subsequently, the permeability of both conjugates was tracked in a fluorescence-activated cell-sorting (FACS) experiment. First, we measured the fluorescence intensity of both fusion molecules in solution in order to exclude the possibility that the inherent fluorescence of the BODIPY fluorophore was affected by fusion to MTX or TMP. As the concentration-fluorescence curves of both BODIPY fusions closely overlap, this does not seem to be the case (Figure IV.4A). Next, HEK293T cells were incubated for a fixed time (15 min) with increasing concentrations of either BODIPY fusion molecule. The mean fluorescence of the viable cell subset was measured by FACS, and showed a dose-dependent increase that was significantly higher for the TMP-linked fluorophore than for the MTX fusion molecule (Figure IV.4B). Additionally, the rate of dye fusion molecule uptake was followed as a function of time at a set concentration of 1 µM. Cells turned out to take up the BODIPY conjugates

rapidly, with maximal fluorescence being reached within 30 s (Figure IV.4C). Consistent with the dose-response analysis, both the rate of conjugate uptake and the plateau level of fluorescence were markedly higher for the TMP-BODIPY fusion. These results indicate that the increased sensitivity obtained when using TMP fusion compounds in MASPIT can be attributed, at least partly, to a significantly higher membrane permeability than for MTX-linked compounds, although we cannot exclude that this trend might be affected by the nature of the bait.

IV.2.2. SNAP-tag approach

To implement the SNAP-tag strategy, we needed a general O^6-benzylguanine (BG) reagent, *para*-substituted with a PEG linker and with a terminal azide ligation handle. Therefore, the azido spacer **25** was first benzylated and then purinated with **46**[28] and **50**,[29] respectively (Scheme IV.4). Subsequent CuAAC between the resulting BG-based building block **51** and alkyne-functionalized tamoxifen readily afforded the desired TAM-BGFC **52**. However, the BG-PEG_6-N_3 reagent suffers from intrinsic thermal degradation issues that result in the elimination of the terminal azidoethyl group and make it less attractive than the optimized TMP reagent. In order to enable BGFC incorporation in the MASPIT system, in the plasmid encoding the CR fusion protein, the eDHFR coding sequence was replaced by a DNA fragment encoding a hAGT mutant that exhibits increased activity towards BG and which has been optimized with regard to mammalian codon usage.



Scheme IV.4: Synthesis of the BG-PEG_6-N_3 reagent. [i] TBDMSCl, imidazole, DMF, 42%; [ii] PPh_3, I_2, imidazole, Et_2O/MeCN, 79%; [iii] **25**, NaH, DMF, 0°C, 59%; [iv] HF.pyr, THF, 80%; [v] 1-methylpyrrolidine, DMF, 63%; [vi] KO*t*Bu, DMF, 25%.

The SNAP-tag MASPIT version was evaluated for the interaction between tamoxifen and its primary target, ER1. Cells expressing both the CR-hAGT and ER1 prey fusion proteins and treated with the cytokine ligand and increasing concentrations of the tamoxifen-BGFC **52** exhibited a dose-dependent increase in luciferase reporter activity (Figure IV.5). This indicated that a ternary complex containing the hAGT and ER1 fusion proteins and the tamoxifen-BGFC is indeed formed, likely involving a covalent bond between hAGT and the BG conjugate. However, the induction window turned out to be significantly lower (sixfold vs. 177-fold) and the EC_{50} markedly higher (2.4×10^{-7} M vs. 2.5×10^{-8} M) than in the case of the TMP-tag setup. Note, we tested a variety of experimental conditions (varying CR-hAGT and ER1 prey expression levels and time between BGFC addition and luciferase readout) for the interaction between the tamoxifen-BGFC and the ER1 prey, as well as several other tamoxifen

(off- or on-) target proteins recently identified by our group, and in none of these cases did the SNAP-tag approach perform better than the MTX- or TMP-tag-based assay. There might be a number of reasons for this, including titration of the supplied BGFC by endogenous AGT[15] and CR-hAGT fusion protein degradation upon alkyl transfer.[30]

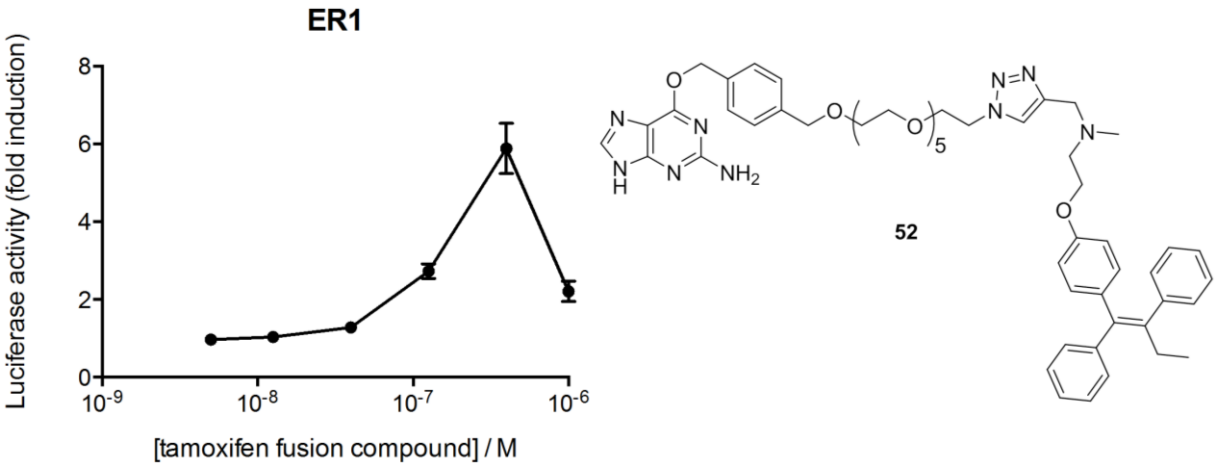


Figure IV.5: Evaluation of the interaction of TAM-BGFC 52 with ER1 in the SNAP-tag MASPIT setup. The graph shows the fold change in luciferase activity for a concentration gradient of TAM-BGFC relative to a control treated with the cytokine ligand without TAM-BGFC. The signal decrease at the highest concentration is due to the cellular toxicity of the tamoxifen fusion compound.

IV.3. Conclusions

In this chapter, we evaluated two approaches, i.e. TMP- and SNAP-tag, to increase the sensitivity of MASPIT. Unexpectedly, the implementation of the SNAP-tag approach, in which the bait fusion molecule is coupled covalently to the CR chimera, did not yield the hypothesized increase in sensitivity. On the other hand, we have demonstrated a clear improvement in the system's sensitivity by introducing trimethoprim as an alternative immobilizing anchor moiety. This improvement is possibly due to its significantly higher membrane permeability than that of MTX-based fusion compounds. In addition, we presented a scalable synthesis of a versatile TMP reagent that proved superior to the original MTX probe with respect to solubility and stability under various reaction conditions.

IV.4. Experimental Section

IV.4.1. Synthesis

General: All reactions were performed under nitrogen and at ambient temperature, unless stated otherwise. Reagents and solvents were purchased from Sigma-Aldrich, Acros Organics, or TCI Europe, and used as received. Reactions were monitored by thin-layer chromatography on TLC aluminum sheets (Macherey-Nagel, Alugram Sil G/UV_{254}) with detection by spraying with a solution of $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$ (25 g/L) and $(NH_4)_4Ce(SO_4)_4{\cdot}2H_2O$ (10 g/L) in H_2SO_4 (10%) or $KMnO_4$ (20 g/L) and K_2CO_3 (10 g/L) in water followed by charring, or Ninhydrin (2% in ethanol). Column chromatography was performed manually with 60 Å silica gel (Grace, Davisil, 40-63 µm) and/or automatically on a Grace Reveleris X2 flash system equipped with disposable silica gel cartridges (Grace, Reveleris). LC-MS analyses were carried out on a Waters Alliance 2695 XE separation Module by using a Phenomenex Luna reversed-phase C18 column (100×2.00 mm, 3 µm) and a gradient system of HCOOH in H_2O (0.1 %, *v/v*)/HCOOH in CH_3CN (0.1 %, *v/v*) at a flow rate of 0.4 mL/min. High-resolution spectra were recorded on a Waters LCT Premier XE Mass spectrometer. ^1H and ^{13}C NMR spectra were measured on a Varian Mercury-300BB (300/75 MHz) spectrometer. NMR solvents were purchased from Euriso-top. Chemical shifts (δ) are given in ppm relative to tetramethylsilane (^1H NMR) or $CDCl_3$, CD_3OD or $SO(CD_3)_2$ (^{13}C NMR) as internal standards. Coupling constants are given in Hz. Preparative TLC purification was carried out on glass-backed Uniplate TLC plates (Analtech, Silica gel GF, UV254, 20×20 cm, 2000 µm). Preparative HPLC purifications were carried out by using a Laprep preparative RP-HPLC system equipped with a Phenomenex Luna C18 column (21.20×250 mm, 5 µm) with a gradient system of HCOOH in H_2O (0.2 %, *v/v*)/CH_3CN at a flow rate of 17.5 mL/min. Microwave experiments were performed using a Milestone Microsynth under fiberoptic internal temperature control.

α-Azido,α-deoxyhexa(ethylene glycol) (25): Triethylamine (8.4 mL, 60.4 mmol) and TsCl (5.62 g, 29.5 mmol) were added to an ice-cooled solution of hexa(ethylene glycol) (9.58 g, 33.9 mmol) in anhydrous CH_2Cl_2 (150 mL). This solution was stirred overnight, and the temperature was allowed to rise to RT. The reaction mixture was concentrated *in vacuo*, and the crude tosylate was taken up in DMF (250 mL). Sodium azide (6.71 g, 103.2 mmol) was added, and the mixture was vigorously stirred overnight at 60°C. The mixture was concentrated *in vacuo*, and the residue was purified by silica gel chromatography (MeOH/CH_2Cl_2 2:98, v/v) to yield the title compound (5.80 g, 18.9 mmol, 64%) as a pale yellow viscous liquid. ^1H NMR (300 MHz, $CDCl_3$) δ = 3.74-3.69 (m, 2H), 3.69-3.64 (m, 18H), 3.62-3.58 (m, 2H), 3.39 (t, *J* = 5.0 Hz, 2H), 3.05 ppm (br s, 1H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 72.5, 70.5-70.3 (wide peak), 70.2, 69.9, 61.5, 50.5 ppm; HRMS: calcd. for $C_{12}H_{25}N_3O_6Na$: 330.1636 $[M+Na]^+$, found: 330.1649.

α-Azido,α-deoxy,ω-*p*-toluenesulfonylhexa(ethylene glycol) (26): Alcohol **25** (1.53 g, 5.0 mmol) was dissolved in anhydrous CH_2Cl_2 (25 mL), and the solution was cooled on ice. Triethylamine (2.1 mL, 15.1 mmol) and TsCl (2.38 g, 12.5 mmol) were added. The resulting mixture was stirred overnight, and the temperature was allowed to rise to RT. The mixture was concentrated *in vacuo*, and the residue was purified by silica gel chromatography (MeOH/CH_2Cl_2 1-3%, v/v) to yield the title compound (2.10 g, 4.5 mmol, 91%) as a brownish-yellow viscous liquid. ^1H NMR (300 MHz, $CDCl_3$) δ = 7.79 (d, *J* = 8.4 Hz, 2H), 7.35 (d, *J* = 8.4 Hz, 2H), 4.15 (t, *J* = 5.0 Hz, 2H), 3.70-3.56 (m, 20H), 3.37 (t, *J* =

5.0 Hz, 2H), 2.44 ppm (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 144.6, 132.8, 129.7, 127.7, 70.5-70.2 (wide peak), 69.8, 69.1, 68.4, 50.4, 21.4 ppm; HRMS: calcd. for $C_{19}H_{35}N_4O_8S$: 479.2170 $[M+NH_4]^+$, found: 479.2176.

First-generation TMP-N_3 reagent (27): K_2CO_3 (138 mg, 1.00 mmol) and PEG linker **26** (369 mg, 0.80 mmol) were added to a solution of phenol **28** (138 mg, 0.50 mmol) in acetone (5 mL). The resulting mixture was refluxed at 75°C for 60 h with vigorous stirring. After being cooled to RT, the mixture was concentrated *in vacuo*. The residue was repeatedly purified by silica gel chromatography (MeOH/CH_2Cl_2 0-15%, v/v) to give a semipure product, which was further purified to homogeneity by preparative RP-HPLC (10-100% MeCN) to yield the title compound (111.2 mg, 197 µmol, 39%) as a clear, colorless oil. ^{1}H NMR (300 MHz, CD_3OD) δ = 7.35 (s, 1H), 6.54 (s, 2H), 4.05 (t, *J* = 4.8 Hz, 2H), 3.79 (s, 6H), 3.77-3.71 (m, 2H), 3.71-3.60 (m, 20H), 3.35 ppm (t, *J* = 5.0 Hz, 2H); ^{13}C NMR (75 MHz, CD_3OD) δ = 164.9, 161.3, 154.8, 151.8, 136.4, 136.0, 108.8, 106.9, 73.4, 71.5-71.4 (wide peak), 71.34, 71.27, 71.0, 56.6, 51.8, 34.3 ppm; LC-HRMS: t_R = 5.90 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{25}H_{40}N_7O_8$: 566.2933 $[M+H]^+$, found: 566.2919.

TMP-OH (28): The synthetic procedure was adapted from Long *et al*.[17] A large scale synthesis was initiated from 100 g commercially available trimethoprim, subdivided into four equal, sequential synthesis batches. In brief, trimethoprim (25 g, 86.1 mmol) was taken up in aq. HBr (311 mL, 48%, w/w) and heated to 100°C under vigorous stirring. After maintaining this temperature for 28 min, the solution was immediately quenched with ice-cooled aq. NaOH (75 mL, 50%, w/w). The resulting mixture was allowed to cool to RT to give fine precipitates, which were filtered, washed with Et_2O (2 x 75 mL) and dried overnight under high vacuum. Subsequently, the crude precipitates from the four batches were combined and taken up in ice-cooled water (600 mL), after which the pH was adjusted to 7 by the addition of NaOH (4M). The resulting mixture was redissolved by heating to 125°C for 15 min under vigorous stirring. After this time, the solution was allowed to cool for only 5 min at RT and the freshly formed precipitates were directly filtered and dried overnight under high vacuum. This recrystallization procedure was twice more repeated to yield the title compound (28.2 g, 102.1 mmol, 30%) as an off white solid. ^{1}H NMR (300 MHz, $SO(CD_3)_2$) δ = 8.16 (br s, 1H), 7.47 (s, 1H), 6.49 (s, 2H), 6.07 (s, 2H), 5.71 (s, 2H), 3.70 (s, 6H), 3.48 ppm (s, 2H); ^{13}C NMR (75 MHz, $SO(CD_3)_2$) δ = 162.23, 162.16, 155.4, 147.9, 133.8, 130.0, 106.2, 106.1, 56.0, 32.8 ppm; HRMS: calcd. for $C_{13}H_{17}N_4O_3$: 277.1295 $[M+H]^+$, found: 277.1297.

First-generation tamoxifen TFC (29a): Azide **27** (166.5 mg, 294 µmol, 2.3 eq.) was taken up in H_2O/*t*BuOH (2.5 mL, 1:1, v/v), and alkyne-functionalized tamoxifen (50.0 mg, 126 µmol, 1.0 eq.), $CuSO_4$ (50.4 µL, 0.5M, 0.2 eq.), and Na ascorbate (252 µL, 0.5M, 1.0 eq.) were added. The resulting mixture was charged with a catalytic amount of TBTA[31] and heated to 80°C for 16 h with vigorous stirring. After being cooled to RT, the solution was concentrated *in vacuo*. The residue was purified by preparative RP-HPLC (10-100% MeCN) to yield the title compound (90.9 mg, 95 µmol, 75%) as a white amorphous solid. LC-HRMS: t_R = 6.96 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{53}H_{70}N_8O_9$: 481.2627 $[M+2H]^{2+}$, found: 481.2622.

First-generation reversine TFC (29b): Azide **27** (67.9 mg, 120 µmol, 1.7 eq.) was taken up in H_2O/*t*BuOH (2.1 mL, 1:1, v/v), and alkyne-functionalized reversine (30.3 mg, 70 µmol, 1.0 eq.), $CuSO_4$ (28.0 µL, 0.5M, 0.2 eq.), and Na ascorbate (140 µL, 0.5M, 1.0 eq.) were added. The resulting mixture

was charged with a catalytic amount of both Et_3N and TBTA[31] and heated to 80°C with vigorous stirring. Upon completion of the reaction (88 h), the solution was cooled to RT and concentrated *in vacuo*. The residue was purified by preparative RP-HPLC (10-100% MeCN) to yield the title compound (36.6 mg, 37 µmol, 53%) as a white amorphous solid. LC-HRMS: t_R = 5.30 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{49}H_{72}N_{15}O_8$: 332.8557 $[M+3H]^{3+}$, found: 332.8524.

First-generation simvastatin TFC (29c): Azide **27** (84.8 mg, 150 µmol, 1.5 eq.) was taken up in H_2O/DMF (3.3 mL, 1:1, v/v), and alkyne-functionalized simvastatin (42.9 mg, 100 µmol, 1.0 eq.), $CuSO_4$ (40.0 µL, 0.5M, 0.2 eq.), and Na ascorbate (200 µL, 0.5M, 1.0 eq.) were added. The resulting mixture was charged with a catalytic amount of TBTA[31] and heated to 75°C for 16 h with vigorous stirring. After being cooled to RT, the solution was concentrated *in vacuo*. The residue was purified by preparative RP-HPLC (30-50% MeCN) to yield the title compound (16.7 mg, 17 µmol, 17%) as a white amorphous solid. LC-HRMS: t_R = 7.39 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{51}H_{77}N_7O_{13}$: 497.7784 $[M+2H]^{2+}$, found: 497.7700.

Monacolin J (30): The synthetic procedure was adapted from Sorensen *et al*.[21] In brief, lovastatin (12.0 g, 29.7 mmol) was suspended in aq. LiOH (600 mL, 1M) and refluxed overnight under vigorous stirring. After being cooled to RT, the pH was adjusted to 1-2 by the addition of HCl (12M). The acidic solution was extracted with CH_2Cl_2 (3 x 600 mL) and the organic fractions were pooled, dried over Na_2SO_4 and concentrated *in vacuo* to give 12.7 g crude residue. Next, the latter was dissolved in toluene (700 mL) and lactonized by refluxing with a CaH_2 Soxhlet apparatus for 1 h. After being cooled to RT, the reaction mixture was concentrated *in vacuo* and the residue was purified by silica gel chromatography (EtOAc/hexane 75-80%, v/v). The thus obtained semi pure product was purified further by precipitation (EtOAc/hexane) yielding the title compound (4.2 g, 13.2 mmol, 44%) as a white solid. ^{1}H NMR (300 MHz, CD_3OD) δ = 5.95 (d, 1H, *J* = 9.6 Hz), 5.75 (dd, 1H, *J* = 9.6 Hz, *J* = 6.0 Hz), 5.49-5.44 (m, 1H), 4.77-4.62 (m1 H), 4.29-4.20 (m, 2H), 2.73 (dd, 1H, *J* = 17.7 Hz, *J* = 4.5 Hz), 2.53 (ddd, 1H, *J* = 17.7 Hz, *J* = 3.3 Hz, *J* = 1.7 Hz), 2.45-2.32 (m, 2H), 2.14 (dd, 1H, *J* = 12.0 Hz, *J* = 2.7 Hz), 2.02-1.73 (m, 7H), 1.64-1.48 (m, 1H), 1.48-1.33 (m, 1H), 1.19 (d, 3H, *J* = 7.5 Hz), 0.90 ppm (d, 3H, *J* = 7.2 Hz); ^{13}C NMR (75 MHz, CD_3OD) δ = 173.7, 133.9, 133.0, 130.7, 130.1, 78.3, 65.7, 63.3, 39.6, 39.2, 37.5, 37.1, 36.6, 34.0, 32.1, 29.1, 25.1, 23.6, 14.3 ppm; HRMS: calcd. for $C_{19}H_{29}O_4$: 321.2060 $[M+H]^+$, found: 321.2060.

***tert*-Butyldimethylsilyloxymonacolin J (31):** The synthetic procedure was adapted from Sorensen *et al*.[21] In brief, to a solution of compound **30** (1.60 g, 5.0 mmol) and imidazole (1.70 g, 25.0 mmol) in DMF (70 mL), *tert*-butyldimethylsilyl chloride (TBDMSCl) (1.88 g, 12.5 mmol) was added and the resulting reaction mixture was stirred overnight at RT. The solution was poured into water (140 mL) and transferred to a separation funnel. The aqueous fraction was extracted with Et_2O (3 x 250 mL) and the organic fractions were pooled, dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by silica gel chromatography (EtOAc/toluene 15:85, v/v) to yield the title compound (2.03 g, 4.67 mmol, 93%) as a white solid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 5.98 (d, 1H, *J* = 9.6 Hz), 5.80 (dd, 1H, *J* = 9.6 Hz, *J* = 6.0 Hz), 5.57-5.51 (m, 1H), 4.73-4.61 (m, 1H), 4.33-4.26 (m, 1H), 4.26-4.20 (m, 1H), 2.62 (dd, 1H, *J* = 17.4 Hz, *J* = 4.2 Hz), 2.55 (ddd, 1H, *J* = 17.4 Hz, *J* = 3.6 Hz, *J* = 1.3 Hz), 2.50-2.31 (m, 2H), 2.17 (dd, 1H, *J*= 12.0 Hz, *J* = 2.7 Hz), 1.96-1.63 (m, 7H), 1.57-1.43 (m, 2H), 1.37 (br s, 1H), 1.19 (d, 3H, *J* = 7.5 Hz), 0.90 (d, 3H, *J* = 7.2 Hz), 0.884 (s, 9H), 0.078 (s, 3H), 0.073 ppm (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 170.6, 133.8, 131.5, 130.2, 128.6, 76.6, 65.4, 63.7, 39.5, 39.0, 37.1, 36.6, 36.0, 33.2,

31.0, 27.5, 25.8, 24.5, 24.0, 18.1, 14.1, -4.7 ppm; HRMS: calcd. for $C_{25}H_{43}O_4Si$: 435.2925 $[M+H]^+$, found: 435.2941.

TBS-protected alkyne-functionalized simvastatin (32): To an ice-cooled solution of 2,2-dimethylpent-4-ynoic acid **37** (278 mg, 2.20 mmol) in CH_2Cl_2 (5 mL), PPh_3 (1.44 g, 5.5 mmol) and NBS (0.63 g, 3.5 mmol) were added. The resulting reaction mixture was allowed to stir for 30 min at RT. After this time, the mixture was cooled on ice again and *N*,*N*-dimethylaniline (0.3 mL, 2.38 mmol) and compound **31** (869 mg, 2.00 mmol) were added. The resulting solution was stirred overnight letting the temperature rise to RT. The reaction mixture was successively washed with 1.0M HCl, water, sat. $NaHCO_3$, water and brine (20 mL), dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by silica gel chromatography (EtOAc/toluene 10-12.5%, v/v) to yield the title compound (624 mg, 1.15 mmol, 58%, 80% based on recovered starting material) as a pale yellowish oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 5.98 (d, 1H, *J* = 9.6 Hz), 5.77 (dd, 1H, *J* = 9.6 Hz, *J* = 6.0 Hz), 5.53-5.49 (m, 1H), 5.38-5.33 (m, 1H), 4.65-4.54 (m, 1H), 4.33-4.27 (m, 1H), 2.62 (dd, 1H, *J* = 17.4 Hz, *J* = 4.5 Hz), 2.54 (ddd, 1H, *J* = 17.4 Hz, *J* = 3.3 Hz, *J* = 1.2 Hz), 2.46-2.39 (m, 3H), 2.39-2.32 (m, 1H), 2.27 (dd, 1H, *J* = 12.0 Hz, *J* = 2.7 Hz), 2.06-1.91 (m, 3H), 1.91-1.79 (m, 2H), 1.74-1.59 (m, 2H), 1.54-1.29 (m, 3H), 1.26 (s, 3H), 1.25 (s, 3H), 1.08 (d, 3H, *J* = 7.5 Hz), 0.890 (d, 3H, *J* = 6.9 Hz), 0.885 (s, 9H), 0.084 (s, 3H), 0.081 ppm (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 176.1, 170.1, 132.9, 131.5, 129.6, 128.3, 81.0, 76.3, 70.7, 68.7, 63.5, 42.2, 39.2, 37.4, 36.77, 36.75, 33.0, 32.6, 30.6, 29.4, 27.2, 25.7, 24.7, 24.4, 24.1, 23.1, 17.9, 13.8, -4.91, -4.92 ppm; HRMS: calcd. for $C_{32}H_{51}O_5Si$: 543.3500 $[M+H]^+$, found: 543.3502.

Alkyne-functionalized simvastatin (33): To an ice-cooled solution of compound **32** (390 mg, 0.72 mmol) in a mixture of THF and 1,4-dioxane (7 mL, 95:5, v/v), concentrated HCl (482 µL) was added and the resulting reaction mixture was stirred for 5 h at RT. The solution was poured into sat. $NaHCO_3$ (30 mL) and transferred to a separation funnel. The aqueous fraction was extracted with CH_2Cl_2 (3 x 30 mL) and the organic fractions were pooled, dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by silica gel chromatography (EtOAc/toluene 1:1, v/v) yielding the title compound (234 mg, 0.55 mmol, 76%) as a yellowish/beige foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 5.99 (d, 1H, *J* = 9.6 Hz), 5.78 (dd, 1H, *J* = 9.6 Hz, *J* = 6.0 Hz), 5.53-5.49 (m, 1H), 5.40-5.35 (m, 1H), 4.69-4.58 (m, 1H), 4.41-4.33 (m, 1H), 2.74 (dd, 1H, *J* = 17.6 Hz, *J* = 5.0 Hz), 2.62 (ddd, 1H, *J* = 17.6 Hz, *J* = 3.8 Hz, *J* = 1.5 Hz), 2.49-2.41 (m, 3H), 2.39-2.33 (m, 1H), 2.27 (dd, 1H, *J* = 12.0 Hz, *J* = 2.7 Hz), 2.05 (t, 1H, *J* = 2.7 Hz), 2.01-1.93 (m, 3H), 1.90-1.77 (m, 2H), 1.76-1.64 (m, 2H), 1.55-1.30 (m, 3H), 1.26 (s, 3H), 1.25 (s, 3H), 1.08 (d, 3H, *J* = 7.2 Hz), 0.89 ppm (d, 3H, *J* = 6.9 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 176.6, 170.6, 133.1, 131.6, 129.8, 128.5, 81.3, 76.5, 71.0, 69.0, 62.8, 42.5, 38.8, 37.6, 36.8, 36.2, 33.1, 32.9, 30.8, 29.7, 27.4, 24.9, 24.7, 24.5, 23.3, 14.0 ppm; HRMS: calcd. for $C_{26}H_{36}O_5Na$: 451.2455 $[M+Na]^+$, found: 451.2466.

Simvastatin-MFC (34): The earlier prepared α-*tert*-butylester protected MTX-PEG_6-N_3 reagent **6a** (160 mg, 0.2 mmol) was taken up in a mixture of TFA and CH_2Cl_2 (5 mL, 1:1, v/v) and stirred for 40 min at RT. The reaction mixture was then taken to dryness, coevaporated twice with toluene and concentrated under high vacuum for 1 h. The residue was dissolved in 2.8 mL of DMF, taken up in a mixture of water and DMF (3 mL, 1:1, v/v) and alkyne **33** (43 mg, 0.1 mmol), $CuSO_4$ (400 µL, 0.05M, 0.2 eq.) and Na ascorbate (200 µL, 0.5M, 1.0 eq.) were added. Finally, the resulting reaction mixture was charged with a catalytic amount of both Et_3N and TBTA[31] and heated to 150°C for 15 min (microwave). After being cooled to RT, the solution was concentrated *in vacuo*. The residue was

purified by preparative RP-HPLC (10-70% MeCN) yielding the title compound (51 mg, 44 µmol, 44%) as a pale yellow amorphous solid. LC-HRMS: t_R = 4.69 min (30-100% MeCN, 15 min run); HRMS: calcd. for $C_{58}H_{84}N_{12}O_{14}$: 586.3110 $[M+2H]^{2+}$, found: 586.3058.

Methyl 2,2-dimethylpent-4-ynoate (36): To a stirred solution of diisopropylamine (14.7 mL, 104.0 mmol) in anhydrous THF (50 mL) under a nitrogen atmosphere at -78°C was added dropwise *n*-BuLi (56.2 mL, 1.6M in *n*-hexane, 89.9 mmol). The resulting solution was warmed to 0°C, stirred for 20 min and then cooled to -78°C. To this solution of LDA was added dropwise methyl isobutyrate (**35**) (9.87 mL, 86.1 mmol) at -78°C. This mixture was next brought to 0°C, stirred for 20 min and cooled to -78°C again. Finally, propargyl bromide (10.2 mL, 80% in toluene, 94.4 mmol) was added and the resulting solution was stirred overnight letting the temperature rise to RT. The reaction mixture was poured into brine (20 mL) and transferred to a separation funnel. The aqueous fraction was extracted with Et_2O (3 x 40 mL) and the organic fractions were pooled, dried over Na_2SO_4 and filtered. Vacuum distillation of the filtrate yielded the title compound (8.57 g, 61.1 mmol, 71%) as a colorless liquid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 3.70 (s, 3H), 2.44 (d, 2H, *J* = 2.6 Hz), 2.01 (t, 1H, *J* = 2.6 Hz), 1.28 ppm (s, 6H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 177.2, 81.1, 70.6, 52.2, 42.2, 29.7, 24.6 ppm.

2,2-Dimethylpent-4-ynoic acid (37): An ice-cooled solution of ester **36** (8.57 g, 61.1 mmol) in MeOH (100 mL) was treated with NaOH (30 mL, 4.0M). The resulting reaction mixture was warmed to RT and subsequently heated to 50°C overnight under vigorous stirring. After being cooled to RT, the mixture was acidified to pH 1 by addition of 4M HCl. The acidic solution was next extracted with EtOAc (3 x 250 mL) and the organic fractions were pooled, dried over Na_2SO_4 and concentrated *in vacuo* to yield the title compound (7.21 g, 57.2 mmol, 94%) as a pale yellowish oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 11.60 (br s, 1H), 2.47 (d, 2H, *J* = 2.6 Hz), 2.04 (t, 1H, *J* = 2.6 Hz), 1.32 ppm (s, 6H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 183.7, 80.8, 70.9, 42.1, 29.3, 24.4 ppm. HRMS: calcd. for $C_7H_9O_2$: 125.0608 $[M-H]^-$, found: 125.0594.

Ethyl 5-(*p*-trimethoprimoxy)valerate (38): Cs_2CO_3 (11.80 g, 36.2 mmol, 2.0 eq.) and ethyl 5-bromovalerate (2.87 mL, 18.1 mmol, 1.0 eq.) were added to a solution of phenol **28** (5.00 g, 18.1 mmol, 1.0 eq.) in DMF (250 mL). The resulting solution was heated to 70°C for 7 h. After being cooled to RT, the mixture was concentrated *in vacuo*, and the residue was purified by silica gel chromatography ($MeOH/CH_2Cl_2$ 0-13%, v/v) to yield the title compound (4.81 g, 11.9 mmol, 66%) as an off white solid. ^{1}H NMR (300 MHz, CD_3OD) δ = 7.51 (s, 1H), 6.50 (s, 2H), 4.11 (q, *J* = 7.1 Hz, 2H), 3.89 (t, *J* = 6.0 Hz, 2H), 3.76 (s, 6H), 3.62 (s, 2H), 2.38 (t, *J* = 7.2 Hz, 2H), 1.87-1.75 (m, 2H), 1.75-1.64 (m, 2H), 1.23 ppm (t, *J* = 7.1 Hz, 3H); ^{13}C NMR (75 MHz, CD_3OD) δ = 175.5, 164.4, 163.1, 155.8, 154.8, 136.7, 136.2, 108.1, 106.7, 73.6, 61.4, 56.5, 34.7, 34.5, 30.4, 22.7, 14.6 ppm; HRMS: calcd. for $C_{20}H_{29}N_4O_5$: 405.2133 $[M+H]^+$, found: 405.2113.

5-(*p*-Trimethoprimoxy)valeric acid (39): A solution of ester **38** (370 mg, 0.92 mmol, 1.0 eq.) in MeOH (9 mL) was treated with NaOH (1.0 mL, 4.0M). The resulting mixture was stirred overnight at RT, then neutralized by the addition of HCl (1.33 mL, 3.0M); this gave fine precipitates. The latter were filtered, washed with a minimal amount of cold MeOH and dried overnight under high vacuum to afford the title compound (317 mg, 0.84 mmol, 92%) as an off white solid. ^{1}H NMR (300 MHz, $SO(CD_3)_2$) δ = 7.51 (s, 1H), 6.54 (s, 2H), 6.09 (s, 2H), 5.71 (s, 2H), 3.77 (t, *J* = 6.0 Hz, 2H), 3.71 (s, 6H), 3.52 (s, 2H), 2.26 (t, *J* = 6.9 Hz, 2H), 1.71-1.54 ppm (m, 4H); ^{13}C NMR (75 MHz, $SO(CD_3)_2$) δ = 174.8,

162.23, 162.18, 155.6, 152.9, 135.7, 134.8, 105.85, 105.76, 71.9, 55.8, 33.8, 33.0, 29.1, 21.3 ppm; HRMS: calcd. for $C_{18}H_{23}N_4O_5$: 375.1674 $[M-H]^-$, found: 375.1669.

Second-generation TMP-N_3 reagent (40): TPTU (2.72 g, 9.2 mmol, 1.2 eq.) and Et_3N (10.6 mL, 76.3 mmol) were added to a solution of acid **39** (2.87 g, 7.6 mmol, 1.0 eq.) in DMF (33.5 mL). The resulting preactivation mixture was stirred for 5 min at RT. Subsequently, a solution of spacer **4b** (1.58 g, 7.2 mmol, 0.95 eq.) and Et_3N (10.6 mL, 76.3 mmol) in DMF (4.5 mL) was added dropwise to this mixture. The resulting reaction mixture was stirred for 5 h at RT, then concentrated *in vacuo*, and the residue was purified by silica gel chromatography (MeOH/CH_2Cl_2 0-12%, v/v) to yield the title compound (3.27 g, 5.7 mmol, 78%) as a beige oil. ^{1}H NMR (300 MHz, CD_3OD) δ = 7.29 (s, 1H), 6.55 (s, 2H), 3.92 (t, *J* = 6.2 Hz, 2H), 3.80 (s, 6H), 3.68-3.57 (m, 12H), 3.54 (t, *J* = 5.6 Hz, 2H), 3.40-3.32 (m, 4H), 2.28 (t, *J* = 7.2 Hz, 2H), 1.87-1.65 ppm (m, 4H); ^{13}C NMR (75 MHz, CD_3OD) δ = 176.1, 164.9, 160.8, 154.9, 150.9, 136.8, 135.4, 109.0, 106.9, 73.8, 71.50, 71.49, 71.4, 71.2, 71.0, 70.5, 56.6, 51.7, 40.3, 36.6, 34.2, 30.5, 23.6 ppm; HRMS: calcd. for $C_{26}H_{41}N_8O_7$: 577.3093 $[M+H]^+$, found: 577.3090.

Second-generation tamoxifen TFC (41): Azide **40** (132.6 mg, 230 µmol, 2.3 eq.) was taken up in H_2O/*t*BuOH (1.6 mL, 1:1, v/v), and alkyne-functionalized tamoxifen (39.6 mg, 100 µmol, 1.0 eq.), $CuSO_4$ (40 µL, 0.5M, 0.2 eq.), and Na ascorbate (200 µL, 0.5M, 1.0 eq.) were added. The resulting mixture was charged with a catalytic amount of TBTA[31] and heated to 80°C for 16 h with vigorous stirring. After being cooled to RT, the solution was concentrated *in vacuo*. The residue was purified by preparative RP-HPLC (10-100% MeCN) to yield the title compound (54.6 mg, 56 µmol, 56%) as a white amorphous solid. LC-HRMS: t_R = 6.98 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{54}H_{71}N_9O_8$: 486.7707 $[M+2H]^{2+}$, found: 486.7708.

BODIPY MFC (42): The earlier prepared α-*tert*-butylester protected MTX-PEG_6-N_3 reagent **6a** (79.9 mg, 100 µmol) was taken up in TFA/CH_2Cl_2 (2.2 mL, 1:1, v/v), and the solution was stirred for 40 min at RT. The mixture was then evaporated, coevaporated twice with toluene, and concentrated under high vacuum for 1 h. The residue was taken up in H_2O/DMF (2.6 mL, 1:1, v/v), and $CuSO_4$ (120 µL, 0.5M, 0.6 eq.), Na ascorbate (600 µL, 0.5M, 3.0 eq.), and a catalytic amount of TBTA[31] were added. Finally, the pH of the resulting reaction mixture was adjusted to 8 by the addition of Et_3N (168 µL, 12 eq.), and BODIPY-alkyne[27] (108.3 mg, 330 µmol, 3.3 eq.) was added. The reaction mixture was stirred for 24 h at RT and concentrated *in vacuo*. The residue was purified by preparative RP-HPLC (10-100% MeCN) to yield the title compound (63.7 mg, 59 µmol, 59%) as an orange amorphous solid. LC-HRMS: t_R = 7.08 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{51}H_{71}N_{14}O_9BF_2$: 536.2789 $[M+2H]^{2+}$, found: 536.2791.

BODIPY TFC (43): Azide **40** (57.7 mg, 100 µmol, 1.0 eq.) was taken up in H_2O/DMF (2.6 mL, 1:1, v/v), and BODIPY-alkyne[27] (75.5 mg, 230 µmol, 2.3 eq.), $CuSO_4$ (80 µL, 0.5M, 0.4 eq.), and Na ascorbate (400 µL, 0.5M, 2.0 eq.) were added. The resulting mixture was charged with a catalytic amount of TBTA,[31] stirred for 29 h at RT, and concentrated *in vacuo*. The residue was purified by preparative TLC (MeOH/CH_2Cl_2 1:4, v/v with 1.0% NH_4OH) to yield the title compound (37.0 mg, 41 µmol, 41%) as an orange amorphous solid. LC-HRMS: t_R = 7.22 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{45}H_{65}N_{10}O_7BF_2$: 453.2544 $[M+2H]^{2+}$, found: 453.2533.

4-(((TBDMS)oxy)methyl)benzyl alcohol (45): To a solution of 1,4-benzenedimethanol (**44**) (14.00 g, 101.33 mmol, 1.45 eq.) in DMF (500 mL) were added imidazole (6.81 g, 100.03 mmol, 1.43 eq.) and TBDMSCl (10.55 g, 70.00 mmol, 1.0 eq.). The resulting reaction mixture was stirred at RT for 16 h and then concentrated *in vacuo.* To the residue were added water (400 mL) and Et_2O (400 mL) and the biphasic mixture was agitated until the residue was completely dissolved and transferred to a separation funnel. The organic layer was separated and the aqueous fraction was washed with Et_2O (2 x 200 mL). All Et_2O fractions were pooled, dried on Na_2SO_4 and taken to dryness. The residue was purified by silica gel chromatography (EtOAc/hexane, 2:8 v/v) yielding the title compound (6.70 g, 26.54 mmol, 38%) as a colorless oil.
Furthermore, the double silylated product, i.e. 1,4-bis(((TBDMS)oxy)methyl)benzene, was isolated as well, and subsequently mono-deprotected as follows. To a solution of this bis-silyl-protected diol (8.06 g, 21.98 mmol, 1.0 eq.) in THF (200 mL) was added tetra-*n*-butylammonium fluoride (TBAF) (10.1 mL, 1.0M in THF). The resulting reaction mixture was stirred at RT for 1 h and then concentrated *in vacuo.* The residue was purified by silica gel chromatography (EtOAc/hexane 2:8, v/v) yielding additional compound **45** (0.78 g, 3.09 mmol, 31%) as a colorless oil. Taken together, starting from 14.0 g 1,4-benzenedimethanol, 7.48 g (29.63 mmol) of the title compound was obtained in 42% overall yield. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.25 (d, 2H, *J* = 8.4 Hz), 7.21 (d, 2H, *J* = 8.4 Hz), 4.69 (s, 2H), 4.48 (s, 2H), 3.32 (br s, 1H), 0.94 (s, 9H), 0.09 ppm (s, 6H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 140.5, 139.7, 127.0, 126.2, 64.9, 64.7 (weak), 26.0, 18.4, -5.2 ppm; HRMS: calcd. for $C_{14}H_{23}OSi$: 235.1513 $[M+H-H_2O]^+$, found: 235.1519.

4-(((TBDMS)oxy)methyl)benzyl iodide (46): To a solution of triphenylphosphine (5.90 g, 22.50 mmol, 1.5 eq.) and imidazole (2.00 g, 30.0 mmol, 2.0 eq.) in a mixture of Et_2O/MeCN (84 mL, 3:1, v/v) was added diiodine (I_2) (5.70 g, 22.50 mmol, 1.5 eq.) and vigorous stirring was continued until a yellow suspension formed. A solution of benzyl alcohol **45** (3.84 g, 15.00 mmol, 1.0 eq.) in 16 mL of the same solvent mixture was then added. The resulting reaction mixture was stirred for 1 h at RT, filtered and concentrated *in vacuo*. To the residue were added Et_2O (100 mL) and a saturated sodium thiosulfate solution (100 mL) and the biphasic mixture was agitated until the residue was completely dissolved and transferred to a separation funnel. The organic layer was separated and the aqueous fraction was washed with Et_2O (2 x 100 mL). All Et_2O fractions were pooled, dried on Na_2SO_4 and taken to dryness. The residue was purified by silica gel chromatography (EtOAc/hexane 5:95, v/v) yielding the title compound (4.28 g, 11.81 mmol, 79%) as a reddish-pink oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.34 (d, 2H, *J* = 8.0 Hz), 7.24 (d, 2H, *J* = 8.0 Hz), 4.70 (s, 2H), 4.46 (s, 2H), 0.94 (s, 9H), 0.10 ppm (s, 6H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 141.5, 138.0, 128.8, 126.6, 64.8, 26.1, 18.6, 6.0, -5.1 ppm.

α-Azido,α-deoxy,ω-4-(((TBDMS)oxy)methyl)benzylhexa(ethylene glycol) (47): Azido spacer **25** (3.63 g, 11.8 mmol, 1.05 eq.) was dissolved in DMF (100 mL), and the solution was cooled on ice under nitrogen. Sodium hydride (60% dispersion in mineral oil, 450 mg, 11.25 mmol, 1.0 eq.) was added to this solution, and the mixture was stirred for 10 min. A solution of compound **46** (4.28 g, 11.8 mmol, 1.05 eq.) in DMF (20 mL) was then added, and the resulting mixture was stirred overnight; the temperature was allowed to rise to RT. The mixture was then concentrated *in vacuo*, resuspended in EtOAc (300 mL), and filtered. The filtrate was concentrated *in vacuo*, and the residue was purified by silica gel chromatography (MeOH/CH_2Cl_2 0-8%, v/v) to yield the title compound (3.58 g, 6.6 mmol, 59%) as a clear, colorless oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.30 (br s, 4H), 4.73 (s, 2H), 4.55 (s, 2H), 3.69-3.59 (m, 22H), 3.38 (t, *J* = 5.0 Hz, 2H), 0.94 (s, 9H), 0.09 ppm (s, 6H); ^{13}C NMR (75 MHz, $CDCl_3$) δ =

140.8, 136.9, 127.7, 126.1, 73.1, 70.7-70.5 (wide peak), 70.0, 69.3, 64.8, 50.7, 26.0, 18.4, -5.2 ppm; HRMS: calcd. for $C_{26}H_{51}N_4O_7Si$: 559.3522 $[M+NH_4]^+$, found: 559.3511.

α-Azido,α-deoxy,ω-4-(hydroxymethyl)benzylhexa(ethylene glycol) (48): A hydrogen fluoride-pyridine complex (HF·pyr; 0.9 mL, 38.5M, 34.7 mmol, 16.6 eq.) was added dropwise to a solution of compound **47** (1.13 g, 2.1 mmol, 1.0 eq.) in anhydrous THF (30 mL) in a plastic flask. The resulting mixture was gently stirred at RT for 6 h, then neutralized with NaOH (10 mL, 4.0M). The basic solution was concentrated *in vacuo*, resuspended in MeOH (40 mL), and filtered. The filtrate was dried over Na_2SO_4, concentrated *in vacuo*, and finally coevaporated twice with toluene to remove any traces of pyridine. The residue was purified by silica gel chromatography (MeOH/CH_2Cl_2 0-8%, v/v) to yield the title compound (716 mg, 1.68 mmol, 80%) as a clear, colorless oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 7.30 (br s, 4H), 4.60 (s, 2H), 4.52 (s,2H), 3.67-3.57 (m, 22H), 3.34 (t, *J* = 5.0 Hz, 2H), 3.20 ppm (br s, 1H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 140.6, 137.1, 127.6, 126.6, 72.7,70.4-70.2 (wide peak), 69.7, 69.1, 64.3, 50.4 ppm; HRMS: calcd. for $C_{20}H_{37}N_4O_7$: 445.2657 $[M+NH_4]^+$, found: 445.2673.

1-(2-Aminopurin-6-yl)-1-methylpyrrolidinium chloride (50): 6-Chloroguanine (**49**) (5.0 g, 29.5 mmol) was suspended in DMF (200 mL) and stirred for 1 h at RT. Next, 1-methylpyrrolidine (9.2 mL, 88.5 mmol) was added and the resulting reaction mixture was stirred for two days at RT. After this time, acetone (100 mL) was added to precipitate the product. The solution was filtered and the solid was washed with acetone and dried under high vacuum to yield the title compound (4.72 g, 18.6 mmol, 63%) as an off white solid. ^{1}H NMR (300 MHz, $SO(CD_3)_2$) δ = 13.34 (br s, 1H), 8.34 (s, 1H), 7.10 (s, 2H), 4.65-4.54 (m, 2H), 4.02-3.90 (m, 2H), 3.65 (s, 3H), 2.33-2.15 (m, 2H), 2.15-1.98 ppm (m, 2H); ^{13}C NMR (75 MHz, D_2O) δ = 158.9, 157.9, 151.9, 143.5, 116.3, 64.9, 51.8, 21.3 ppm; HRMS: calcd. for $C_{10}H_{15}N_6$: 219.1353 $[M]^+$, found: 219.1347.

BG-PEG$_6$-N$_3$ reagent (51): Benzyl alcohol **48** (711 mg, 1.66 mmol, 1.0 eq.) was dissolved in DMF (4.2 mL) under nitrogen. KO*t*Bu (746 mg, 6.65 mmol, 4.0 eq.) was added to this solution, and the mixture was stirred for 30 min at RT, turning dark red. After this time, compound **50** (423 mg, 1.66 mmol, 1.0 eq.) was added slowly, and the resulting mixture was stirred for 22 h at RT. The mixture was concentrated *in vacuo*, and the residue was purified by silica gel chromatography (MeOH/CH_2Cl_2 0-10%, v/v) to yield the title compound (227 mg, 0.41 mmol, 25%) as a pale yellow oil. ^{1}H NMR (300 MHz, $SO(CD_3)_2$) δ = 7.82 (br s, 1H), 7.48 (d, *J* = 7.8 Hz, 2H), 7.35 (d, *J* = 7.8 Hz, 2H), 6.27 (br s, 2H), 5.48 (s, 2H), 4.50 (s, 2H), 3.61-3.48 ppm (m, 22H); ^{13}C NMR (75 MHz, $SO(CD_3)_2$) δ = 159.9 (weak), 159.7, 155.2 (weak), 138.4, 137.8 (weak), 135.9, 128.4, 127.5, 113.5 (weak), 71.7, 69.86-69.78 (wide peak), 69.70, 69.2, 69.1, 66.5, 50.0 ppm; LC-HRMS: t_R = 6.47 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{25}H_{37}N_8O_7$: 561.2780 $[M+H]^+$, found: 561.2735.

Tamoxifen BGFC (52): Azide **51** (227.3 mg, 405 µmol, 3.2 eq.) was taken up in H_2O/*t*BuOH (2.5 mL, 1:1, v/v), and alkyne-functionalized tamoxifen (50.0 mg, 126 µmol, 1.0 eq.), $CuSO_4$ (50.4 µL, 0.5M, 0.2 eq.), and Na ascorbate (252 µL, 0.5M, 1.0 eq.) were added. The resulting mixture was charged with a catalytic amount of TBTA[31] and heated to 80°C for 16 h with vigorous stirring. After being cooled to RT, the solution was concentrated *in vacuo*. The residue was purified by preparative RP-HPLC (30-100% MeCN) to yield the title compound (58.1 mg, 61 µmol, 48%) as an off white amorphous solid. LC-HRMS: t_R = 7.89 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{53}H_{67}N_9O_8$: 478.7551 $[M+2H]^{2+}$, found: 478.7526.

IV.4.2. Molecular biology

Performed by the Cytokine Receptor Laboratory, Department of Medical Protein Research, VIB, Ghent & Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University (Prof. Jan Tavernier, Dr. Sam Lievens).

Plasmid constructs: The CR-eDHFR chimeric protein consisting of the full-size mouse leptin receptor F3 mutant fused to *E. coli* DHFR is encoded by the pCLL-eDHFR plasmid described previously.[10] The pCLG-hAGT plasmid encodes the CR-hAGT fusion protein, which is made up of the extracellular, transmembranal, and membrane-proximal portions of the mouse leptin receptor tethered to a mutant hAGT protein. This plasmid was generated by amplifying the hAGT coding sequence of the pSNAPf vector (New England Biolabs) by using forward primer 5'-CCCGA GCTCA ATGGA CAAAG ACTGC GAAAT G-3' and reverse primer 5'-GGGGC GGCCG CTTAA CCCAG CCCAG GCTTG CCCAG-3' to introduce a stop codon downstream of hAGT. The amplicon was cloned into the pCLG vector backbone, which has been described previously,[32] by using SacI and NotI restriction enzymes. The ER1 prey, consisting of full-size ER1 fused N-terminally to a gp130 receptor fragment containing multiple STAT3 recruitment sites, was generated by Gateway recombinatorial cloning (Invitrogen) into a Gateway-compatible pMG1C vector. To generate this destination vector, first the pMet7-Flag Gateway destination vector was made by amplifying the Gateway cassette from the Gateway-compatible pMG1 vector[32] by using forward primer 5'-CCCCA ATTGA CAAGT TTGTA CAAAA AAGC-3' and reverse primer 5'-GGGTC TAGAC TACTT ATCGT CGTCA TCCTT GTAAT CTTTA ATTAA AACCA CTTTG TACAA GAAAG C-3' (the latter containing the Flag epitope coding sequence), and inserting this amplicon into the pMet7 expression vector[7] by ligation of the MfeI-XbaI restriction enzyme product into the EcoRI-XbaI cut pMet7 vector. Next, the gp130 coding sequence of the pMG1 vector[32] was amplified by using forward primer 5'-CCCTT AATTA ACGGA GGGAG TATCT CGACC GTGGT ACACA GTG-3' and reverse primer 5'-GGGTT AATTA ACCCC TGAGG CATGT AGCCG C-3', which was inserted into the PacI site of the pMet7-Flag Gateway destination vector. Finally, the ER1 open reading frame was transferred into the pMG1C destination vector from the ER1 entry clone of the hORFeome collection.[33] The pMG1-TTK prey plasmid encoding the C-terminal fusion of full-size TTK to the gp130 fragment[10] and the pMG1-HMGCR plasmid coding for the C-terminal fusion of the statin-binding cytoplasmic domain to the gp130 fragment[20] have been described elsewhere, as has the STAT3-dependent firefly luciferase reporter pXP2d2-rPAPI-luciferase.[7]

MASPIT assays: HEK293T cells were cultured in 96-well microtiter plates in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, incubated at 37°C, under 8% CO_2, and transfected with pCLL-eDHFR or pCLG-hAGT CR fusion protein plasmid (10 ng per 96 wells), pMG1C-ER1, pMG1-TTK, or pMG1-HMGCR prey plasmid (100 ng per 96 wells) and pXP2d2-rPAPI-luciferase (5 ng per 96 wells) by applying a standard calcium phosphate transfection method, as described earlier.[32] Twenty-four hours after transfection, cells were stimulated with leptin (100 ng/mL final concentration) alone or in combination with the indicated concentration of fusion compound. After another 24 h, luciferase activity was measured by using the Luciferase Assay System kit (Promega) on an Envision plate reader (PerkinElmer). Luciferase data are the average of three technical replicates, error bars indicate standard deviation.

Fluorescence measurement: The fluorescence of the MTX- and TMP-BODIPY fluorophores was measured by preparing a serial dilution of the molecules in PBS in 96-well microtiter plates, and analyzing the plates in an Envision plate reader equipped with a FITC filter set. Fluorescence data are the average of three replicates, error bars indicate standard deviation. The curve in Figure IV.4A was fit by using four-parameter nonlinear regression in GraphPad Prism.

FACS analysis: HEK293T cells at a density of 10^7/mL were incubated with propidium iodide (3 μM), and the indicated concentration of MTX- or TMP-BODIPY (1 μM final concentration in the kinetic experiment in Figure IV.4C) for the indicated time (15 min in the dose-response experiment in Figure IV.4B). Samples were analyzed by using a FACSCalibur instrument (BD Biosciences), measuring mean BODIPY fluorescence of the viable cells (propidium iodide negative).

References

[1] D. C. Swinney, J. Anthony, *Nat. Rev. Drug Discov.* **2011**, *10*, 507-519.
[2] A. L. Hopkins, *Nat. Chem. Biol.* **2008**, *4*, 682-690.
[3] G. Jin, S. T. C. Wong, *Drug Discovery Today* **2014**, *19*, 637-644.
[4] B. F. Cravatt, A. T. Wright, J. W. Kozarich, *Annu. Rev. Biochem.* **2008**, *77*, 383-414.
[5] U. Rix, G. Superti-Furga, *Nat. Chem. Biol.* **2009**, *5*, 616-624.
[6] S. Ziegler, V. Pries, C. Hedberg, H. Waldmann, *Angew. Chem.-Int. Edit.* **2013**, *52*, 2744-2792.
[7] S. Eyckerman, A. Verhee, J. V. der Heyden, I. Lemmens, X. V. Ostade, J. Vandekerckhove, J. Tavernier, *Nat. Cell Biol.* **2001**, *3*, 1114-1119.
[8] S. Lievens, F. Peelman, K. De Bosscher, I. Lemmens, J. Tavernier, *Cytokine Growth Factor Rev.* **2011**, *22*, 321-329.
[9] M. Caligiuri, L. Molz, Q. Liu, F. Kaplan, J. P. Xu, J. Z. Majeti, R. Ramos-Kelsey, K. Murthi, S. Lievens, J. Tavernier, N. Kley, *Chem. Biol.* **2006**, *13*, 711-722.
[10] M. D. P. Risseeuw, D. J. H. De Clercq, S. Lievens, U. Hillaert, D. Sinnaeve, F. Van den Broeck, J. C. Martins, J. Tavernier, S. Van Calenbergh, *ChemMedChem* **2013**, *8*, 521-526.
[11] C. K. Osborne, *N. Engl. J. Med.* **1998**, *339*, 1609-1618.
[12] E. P. Gelmann, *J. Natl. Cancer Inst.* **1996**, *88*, 224-226.
[13] J. R. Appleman, N. Prendergast, T. J. Delcamp, J. H. Freisheim, R. L. Blakley, *J. Biol. Chem.* **1988**, *263*, 10304-10313.
[14] L. W. Miller, Y. Cai, M. P. Sheetz, V. W. Cornish, *Nat. Meth.* **2005**, *2*, 255-257.
[15] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, *Nat. Biotech.* **2002**, *21*, 86-89.
[16] M. J. Hynes, J. A. Maurer, *Angew. Chem.-Int. Edit.* **2012**, *51*, 2151-2154.
[17] M. J. C. Long, Y. Pan, H.-C. Lin, L. Hedstrom, B. Xu, *J. Am. Chem. Soc.* **2011**, *133*, 10006-10009.
[18] W. G. Lai, N. Zahid, J. P. Uetrecht, *J. Pharmacol. Exp. Ther.* **1999**, *291*, 292-299.
[19] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem.-Int. Edit.* **2002**, *41*, 2596-2599.
[20] S. Lievens, S. Gerlo, I. Lemmens, D. J. H. De Clercq, M. D. P. Risseeuw, N. Vanderroost, A.-S. De Smet, E. Ruyssinck, E. Chevet, S. Van Calenbergh, J. Tavernier, *Mol. Cell. Proteomics* **2014**, *13*, 3332-3342.
[21] J. L. Sorensen, K. Auclair, J. Kennedy, C. R. Hutchinson, J. C. Vederas, *Org. Biomol. Chem.* **2003**, *1*, 50-59.
[22] T. Welsch, H.-A. Tran, B. Witulski, *Org. Lett.* **2010**, *12*, 5644-5647.
[23] Hong, C.I.; Kim, J.W.; Shin, H.J.; Kang, T.W.; Cho, D.O. WO 01/45484 A2, 28.06.2001, Chong Kun Dang Pharmaceutical Corp., Seoul (KR).
[24] Z. Chen, C. Jing, S. S. Gallagher, M. P. Sheetz, V. W. Cornish, *J. Am. Chem. Soc.* **2012**, *134*, 13692-13699.
[25] S. S. Gallagher, L. W. Miller, V. W. Cornish, *Anal. Biochem.* **2007**, *363*, 160-162.
[26] J. L. Czlapinski, M. W. Schelle, L. W. Miller, S. T. Laughlin, J. J. Kohler, V. W. Cornish, C. R. Bertozzi, *J. Am. Chem. Soc.* **2008**, *130*, 13186-13187.
[27] M. Verdoes, U. Hillaert, B. I. Florea, M. Sae-Heng, M. D. P. Risseeuw, D. V. Filippov, G. A. van der Marel, H. S. Overkleeft, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 6169-6171.
[28] G. Gorins, L. Kuhnert, C. R. Johnson, L. J. Marnett, *J. Med. Chem.* **1996**, *39*, 4871-4878.
[29] M. Kindermann, N. George, N. Johnsson, K. Johnsson, *J. Am. Chem. Soc.* **2003**, *125*, 7810-7811.
[30] A. Keppler, H. Pick, C. Arrivoli, H. Vogel, K. Johnsson, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9955-9959.
[31] T. R. Chan, R. Hilgraf, K. B. Sharpless, V. V. Fokin, *Org. Lett.* **2004**, *6*, 2853-2855.
[32] S. Lievens, N. Vanderroost, J. Van der Heyden, V. Gesellchen, M. Vidal, J. Tavernier, *J. Proteome Res.* **2009**, *8*, 877-886.
[33] P. Lamesch, N. Li, S. Milstein, C. Fan, T. Hao, G. Szabo, Z. Hu, K. Venkatesan, G. Bethel, P. Martin, J. Rogers, S. Lawlor, S. McLaren, A. Dricot, H. Borick, M. E. Cusick, J. Vandenhaute, I. Dunham, D. E. Hill, M. Vidal, *Genomics* **2007**, *89*, 307-315.

CHAPTER V

THE PHOTOCROSSLINKER APPROACH

The Photocrosslinker Approach: Towards Covalency on the Bait-End of the Dimerizers

V. The Photocrosslinker Approach: Towards Covalency on the Bait-End of the Dimerizers

V.1. Introduction

Further optimization efforts to stabilize the ternary complex (CR-CID-Prey) were aimed at introducing covalent bonding on the bait-end of the dimerizer starting from TMP as immobilizing anchor moiety. For this purpose, we envisioned that photoactivatable crosslinkers would be particularly suitable for capturing transient and low-affinity bait-prey interactions (Figure V.1). Indeed, such photocrosslinkers convert upon photoactivation to highly reactive species, capable of introducing covalent bonds between interacting molecules.[1] In the context of the MASPIT assay, this strategy thus might give rise to increased signal output. However, a potential disadvantage of the approach includes undesired or detrimental nonspecific crosslinking events with cellular components. To the best of our knowledge, no prior literature data on the implementation of photocrosslinkers in the context of three-hybrid systems for compound profiling exists.

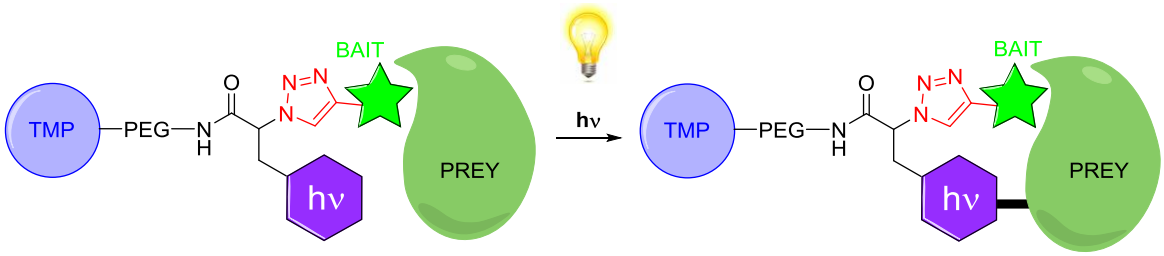


Figure V.1: Schematic representation of the photocrosslinker optimization approach. Upon interaction with a bait, covalent immobilization of a prey chimera on the TMP fusion compound could be attained by activation of the photocrosslinker (hν) with light of the appropriate wavelength.

The initial report in the field of photoaffinity labeling (PAL) was published in 1962 by Westheimer and colleagues, describing the use of a singlet carbene species, generated by photolysis of a diazoacetyl group, to inactivate chymotrypsin.[2] Since then substantial research on the development of novel photogenerated reagents has been conducted,[3],[4] yet the current panel of photocrosslinking groups that has demonstrated utility in cellular systems is rather limited. These comprise benzophenones, aryl azides and diazirines (Figure V.2), whose characteristics and photochemistry will be discussed in the following subsections.[1],[5]

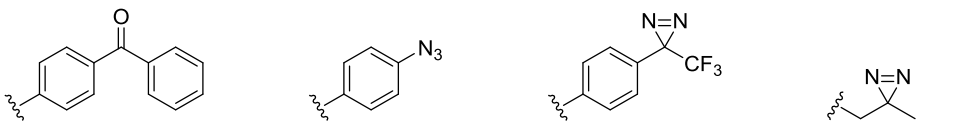


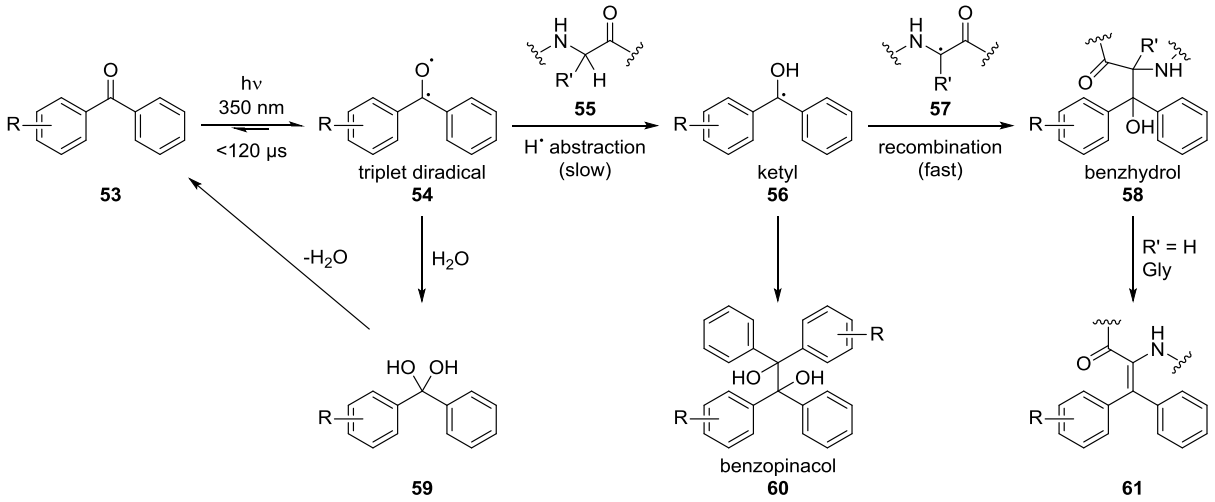
Figure V.2: Chemical structures of photocrosslinking functional groups for use in biological systems. Figure adapted from: *Molecular Biosystems* 2008, 4 (6), 473-480.

In order to function in a cellular context, photocrosslinkers must meet a number of requirements.[1],[5] First, the photochemical precursors must be stable under the various conditions of the cellular environment, while being bioorthogonal to all functional groups present in the sample.

Furthermore, they should only be activated upon irradiation with the appropriate wavelength of light, which should not damage cellular components, i.e. λ_{act} > 300 nm. Second, the lifetime of the generated reactive species needs to be shorter than that of the studied macromolecular complexes so as to limit nonspecific labeling. However, this desired specificity of short-lived active species is often at the expense of their crosslinking efficiency and vice versa. Additionally, the photogenerated species preferably react with any chemical entity in close proximity, regardless of its nature, thereby forming stable covalent adducts. Finally, to minimize perturbation of the studied interaction, a relatively small photophore is required. Taken together, the relative importance of each of these criteria can be anticipated to be application-specific.

V.1.1. Benzophenone (BP)

The chemistry of benzophenones after photolysis is outlined in Scheme V.1.[1],[5] Upon activation by 350-365 nm light, benzophenones (**53**) reversibly convert to a triplet state benzhydril diradical (**54**). In the absence of an appropriate reaction partner, this reactive species can persist for as long as 120 µs before relaxing back to its ground state. The triplet diradical abstracts in a first slow step a hydrogen radical from a neighboring X-H bond. In general, C-H bonds are more prone to react than O-H bonds, especially those that form stabilized carbon radicals, such as benzylic positions, tertiary carbon centers, and amino acid α-positions (**55**). In particular, benzophenones have been shown to exhibit a strong preference for methionine ε-H's.[6] Hydrogen abstraction by the diradical affords a ketyl (**56**) and an alkyl radical (**57**), which subsequently recombine in a second fast step to obtain the final benzhydrol covalent adduct (**58**). In the case of glycine, water can be eliminated from the benzhydrol yielding the corresponding cross-conjugated olefin **61**. Another side reaction that can occur is the homodimerization of ketyl **56** resulting in the formation of benzopinacol **60**. However, typically only trace amounts of the latter are retrieved, given the significant relative difference in kinetics for the tandem abstraction and recombination reactions. Finally, the reactive diradical species (**54**) can be quenched by the insertion of water, which is omnipresent in a cellular environment, giving the corresponding hydrate (**59**). Yet, the latter swiftly dehydrates to the thermodynamically favored ketone (**53**), which can ultimately recover the diradical after UV irradiation.



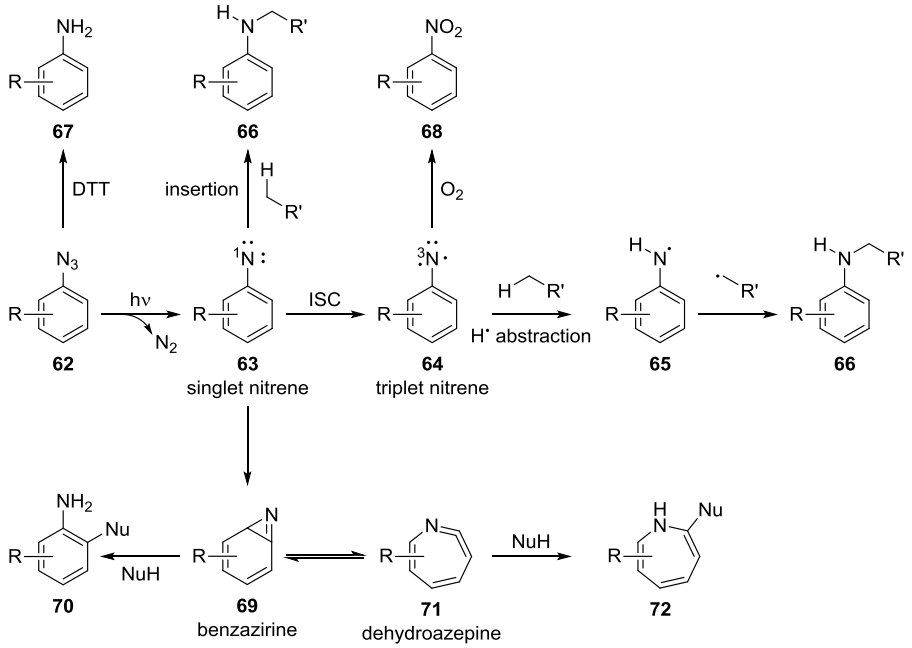
Scheme V.1: Possible reaction pathways of benzophenones after photolysis. Scheme adapted from: *Topics in Current Chemistry* 2012, 324, 85-113.

This 'recycling' capability represents one of the major assets of the benzophenone moiety in terms of crosslinking efficiency, yet concomitantly increases the change of nonspecific labeling (cf. supra).[5] The latter event is also facilitated by the relative bulkiness of this photocrosslinker, as the resulting steric hindrance can lead to discrimination between reaction sites. Moreover, its relatively large size can also directly cause a detrimental perturbation of binding events. On the other hand, benzophenones benefit from the fact that their photogenerated species are more reactive towards C-H bonds compared to nitrenes, and less prone to intramolecular rearrangements than carbenes (cf. infra).[5] Additionally, this photophore can be excited at relatively long wavelengths of 350-365 nm (i.e., well above the earlier mentioned 300 nm cutoff), hence minimizing the incidence of nucleic acid or protein damage.[7],[1] However, the photolysis often requires more prolonged time (> 30 min) to attain reasonable crosslinking efficiency.[8]

V.1.2. Aryl azide (AA)

Aryl azides (**62**) are photoactivated by irradiation with 250 nm light, which triggers the initial formation of a singlet nitrene (**63**) upon release of molecular nitrogen (N_2) (Scheme V.2.).[5] Displaying electrophilic properties, this nitrene species readily undergoes insertion reactions with nearby C-H bonds to yield a covalent adduct (**66**). However, due to its highly energetic and reactive character, the singlet nitrene has a short lifetime (~0.1 ms) and it is rapidly converted into other, lower-energy intermediates such as the corresponding triplet nitrene (**64**) and benzazirine (**69**).[3],[9] The former is generated by intersystem crossing and reacts analogous to activated benzophenones (cf. Section V.1.1) as a diradical via a two-step hydrogen radical abstraction-recombination cascade to afford the same adduct **66**. The benzazirine can further rearrange into dehydroazepine (**71**), and both species can react, as long-lived electrophiles, only with nucleophiles to obtain addition products **70** and **72**, respectively. Therefore, these rearrangements are typically associated with losses in crosslinking efficiency and increase the risk of nonspecific labeling events.[1] Two alternative side reactions might take place upon application of aryl azides in a biological context, i.e. triplet nitrene scavenging by molecular oxygen (a notorious triplet quencher), giving the corresponding nitro

species (**68**);[10] and reduction of the starting material (**62**) to the amine (**67**) by thiols (e.g. dithiothreitol).[11]



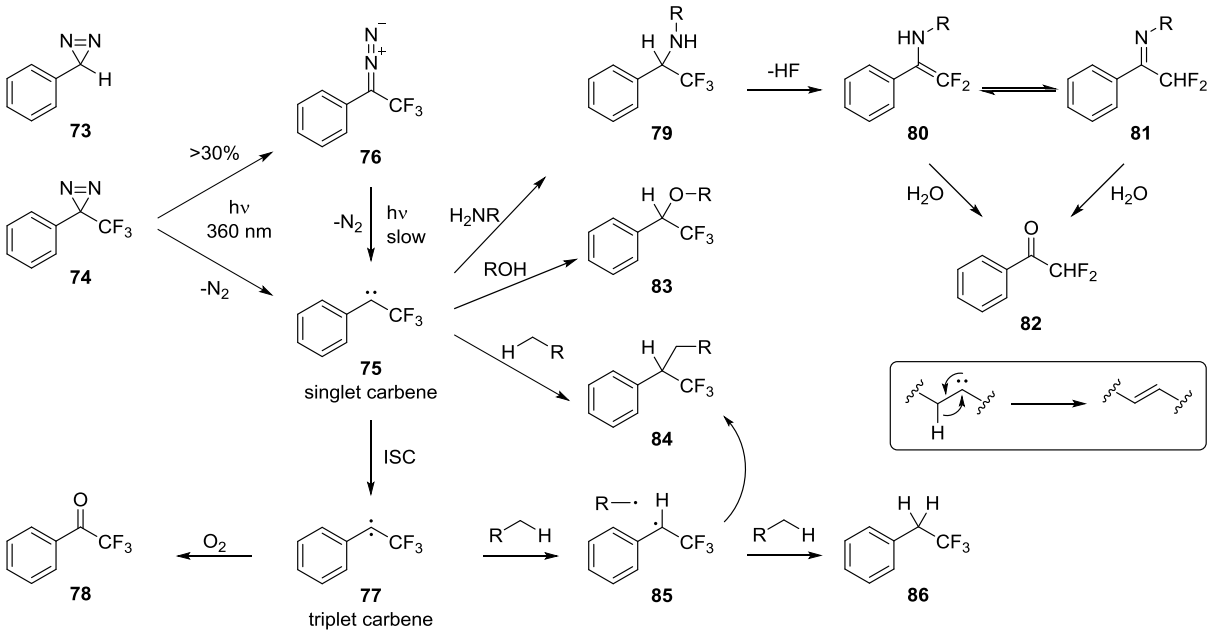
Scheme V.2: Possible reaction pathways of the reactive intermediates formed after photolysis of aryl azides. DTT, dithiothreitol; ISC, intersystem crossing; NuH, nucleophile. Scheme adapted from: *Topics in Current Chemistry* 2012, 324, 85-113.

Clearly, the main benefits of the aryl azide moiety include its relatively small size and ease of incorporation into various PAL probes. However, the biocompatibility of this photocrosslinker is hampered by its maximal absorption at a wavelength below 300 nm (i.e., 250 nm for plain phenyl azide).[1] Hence, in an attempt to improve its photochemical properties, considerable research has been focused on the development of substituted aryl azides. Most notably, it has been shown that electron-withdrawing substituents (e.g., nitro, hydroxyl, and acyl groups) both increase the molar absorptivity and red-shift the absorption maximum (up to 400 nm).[12],[13] In this respect, most substituents *ortho* to the azide should be avoided since they might give rise to undesired cyclizations after photoexcitation.[5] Additionally, (per)fluorinated aryl azides were found to rearrange more slowly from the singlet nitrene to the benzazirine (cf. Scheme V.2.), thereby enhancing the efficiency of insertion reactions.[14] Nevertheless, crosslinking yields for aryl azides are often low (i.e. < 30%) given the diverse fates, including capturing by the solvent, of their reactive photogenerated intermediates.[5]

V.1.3. Diazirine (DZ)

Irradiation of diazirines (**73**/**74**) at 350-380 nm results in molecular nitrogen loss and the generation of a singlet carbene species (**75**) (Scheme V.3.).[5] However, competitively, the starting material is considerably (> 30%) photoisomerized to the linear diazo compound (**76**), which in turn can give access to the singlet carbene (**75**) upon photoexcitation, albeit this conversion is relatively sluggish at the latter activation wavelengths.[15] Hence, the diazo isomer is reasonably long-lived and, given its generally high sensitivity towards nucleophilic attack, one of the root causes for either nonspecific

labeling or decreased crosslinking efficiency.[1] Brunner *et al.* largely circumvented this issue by introducing in the α-position the strong electron-withdrawing trifluoromethyl substituent (as in **74**), which stabilizes the diazo isomer and renders it fairly resistant towards undesired 'dark' side reactions.[16] Singlet carbenes (**75**) are highly reactive crosslinkers with typical nanosecond half-lives, giving fast insertion reactions into any adjacent C-H or heteroatom-H bond.[17] Generally, insertions into hydroxyl groups (**83**) are more efficient compared to C-H insertions (**84**).[18] Reactions with primary or secondary amines yield a covalent adduct (**79**), from which HF can be readily eliminated to afford an enamine (**80**). The latter is in equilibrium with imine **81** and both species can be hydrolyzed (e.g., in a biological context) to the corresponding ketone (**82**), entailing unwanted loss of the captured binding partner.[18] Via intersystem crossing the singlet carbene can be transformed into its triplet state congener (**77**), which reacts as a typical diradical (cf. supra) with C-H bonds to give the same insertion product (**84**). However, in this respect, two quenching side reaction can occur, i.e. aerobic oxidation (cf. supra) to the corresponding ketone (**78**); and reduction (**86**) via two sequential hydrogen radical abstractions.[13] Finally, as illustrated in the inset of Scheme V.3, unsubstituted 3-alkyl-3*H*-diazirines are less suitable as their photogenerated singlet carbenes are likely to rearrange to olefins via hydride shift.[19]



Scheme V.3: Possible reaction pathways of the reactive intermediates formed after photolysis of 3-aryl-3*H*-diazirines. The inset shows hydride shift of a singlet carbene derivative of unsubstituted 3-alkyl-3*H*-diazirines resulting in olefin formation. ISC, intersystem crossing. Scheme adapted from: *Topics in Current Chemistry* 2012, 324, 85-113.

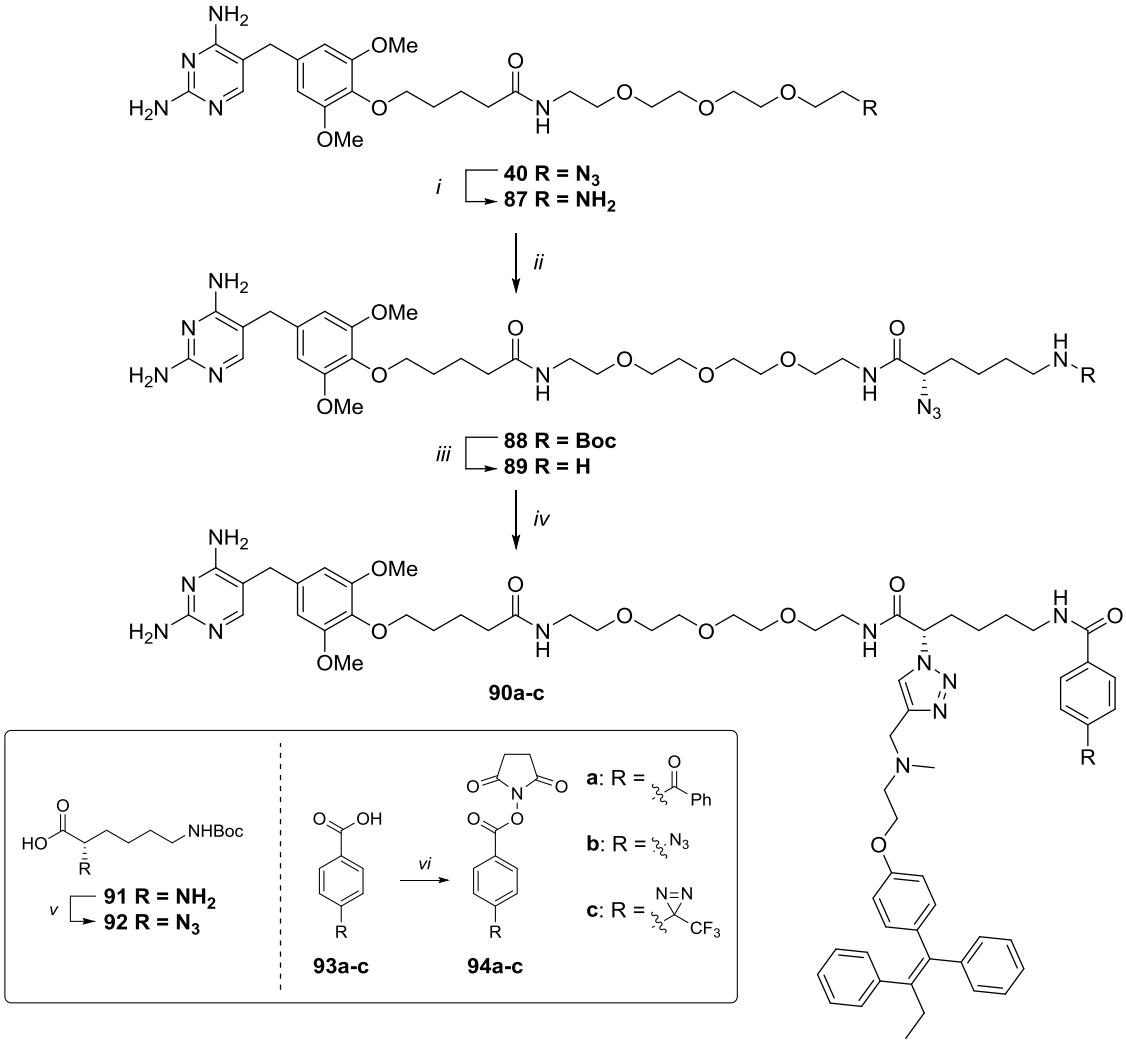
Diazirines offer a number of important advantages over aryl azides, including their stability towards strongly acidic/basic conditions, as well as various oxidating/reducing agents.[5] More importantly, their maximum absorption wavelength (350-380 nm) is well above 300 nm, thus causing no significant damage to the biological sample under analysis. On the other hand, trifluoromethyl phenyl diazirines are quite bulky, despite the relatively small size of the diazirine heterocycle itself. Two other disadvantages of the diazirine photophore, contributing to reduced crosslinking efficiency, comprise its substantial (> 30%) conversion to the diazo isomer upon photoactivation (cf. supra) and

scavenging of the corresponding singlet carbene species by water due to its intrinsic efficient reactivity towards O-H bonds.[1],[5]

In conclusion, the 'ideal photocrosslinker' does not exist. Rather, the optimal crosslinker type is likely to be application-specific due to the differences in reaction site preferences as well as the nature of the photogenerated reactive species. Additionally, it might be worthwhile to screen a panel of different photophores so as to obtain complementary insights into a particular binding event.[13],[5]

V.2. Synthesis of heterotrimeric PAL probes

As a starting point, the second-generation TMP-N_3 reagent **40** was equipped with a double ligation handle originating from (*S*)-6-amino-2-azidohexanoic acid.[20] Toward this end, azide **40** was hydrogenated and subsequently condensed with acid **92**[20] using TPTU reagent. Acidic Boc removal readily afforded the desired building block **89**, which consists of both an azide for CuAAC with alkyne-functionalized tamoxifen **7** and a terminal amine for subsequent amide coupling with NHS esterified photoaffinity labels **94a-c**.[21],[22],[23] Following a one pot sequential protocol, heterotrimeric probes **90a-c**, comprising TMP, TAM and a benzophenone/aryl azide/trifluoromethyl phenyl diazirine photocrosslinker, were successfully obtained after HPLC purification in 30%, 24% and 34% yield, respectively (Scheme V.4).



Scheme V.4: Synthesis of the heterotrimeric TMP-TAM-PAL probes. [i] H_2, Pd/C, H_2O/dioxane, 90%; [ii] **92**, TPTU, Et_3N, DMF, 69%; [iii] TFA, CH_2Cl_2, quant.; [iv] first alkyne-functionalized tamoxifen **7**, Na ascorbate, $CuSO_4$, Et_3N, TBTA, DMF; then **94a-c**, (**a** 30%; **b** 24%; **c** 34%); [v] first NaN_3, Tf_2O, CH_2Cl_2/H_2O; then K_2CO_3, $CuSO_4$, $MeOH/H_2O$, 79%; [vi] NHS, EDC.HCl, DMAP, DMF, 0°C to RT, (**a** 65%; **b** 85%; **c** 94%).

V.3. Biological evaluation of heterotrimeric PAL probes

Previously, in a standard MASPIT cell array screen against a 10,000 ORF prey collection, we identified eight novel candidate interaction partners of tamoxifen using the second-generation tamoxifen-TFC (**41**), in addition to the known estrogen receptor α (ER1) target (Figure V.3A).[24] Interestingly, at least four of these newly identified target proteins cluster in a putative complex that is functionally relevant with respect to a potentially novel mechanism of action for this 40-year-old drug. The heterotrimeric PAL probes **90a-c** were first evaluated for binding to this panel of putative tamoxifen target proteins without UV irradiation and their resulting target protein profile was compared to the one obtained for **41** (Figure V.3).

Although the small-molecule bait and photophore are separated by a flexible methylene spacer, comparison of the screening profiles revealed that the unactivated photocrosslinkers seem to sterically interfere with the binding of tamoxifen to its target proteins. Indeed, in the case of the bulky benzophenone moiety, as much as six out of nine binding partners are lost (Figure V.3B), whereas for the aryl azide and diazirine probe only four interactions are retained (Figure V.3C and 3D, respectively). Nevertheless, we initially estimated that this loss of signal output might be compensated for by successful covalent crosslinking of the target preys following photoactivation.

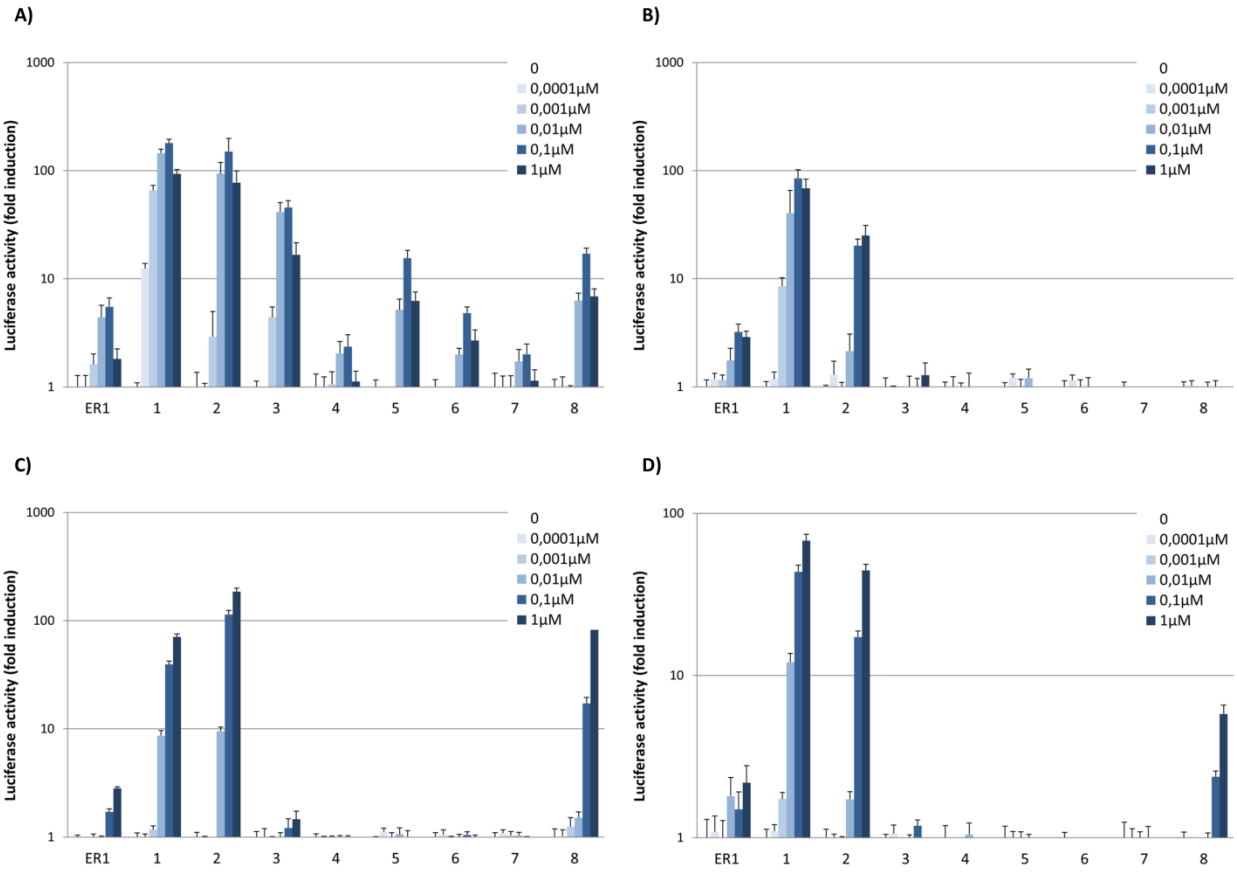


Figure V.3. Target protein profile of the second-generation tamoxifen-TFC and heterotrimeric PAL probes. A) TAM-TFC (**41**); **B)** TMP-TAM-*benzophenone* (**90a**); **C)** TMP-TAM-*aryl azide* (**90b**); **D)** TMP-TAM-*diazirine* (**90c**). Luciferase signals are expressed as fold induction relative to a control sample treated with cytokine without fusion compound. 1-8, confidential anonymized newly identified potential target preys.

However, establishing covalent immobilization of the prey chimeras with the aim of increasing the MASPIT readout turned out to be less straightforward than anticipated based on success stories from the chemical proteomics field.[5],[25] Upon extensive testing of PAL probes **90a-c** against a panel of target preys in MASPIT, no relative increase in luciferase signal was observed after photoexcitation, as exemplified below for candidate target protein '2' (Figure V.4). In this respect, variation of the irradiation time (BP probe **90a**: 4 min - 2 h; AA probe **90b**: 15 sec - 1 min; DZ probe **90c**: 5 - 40 min), cell culture medium (DMEM vs. phenol red free DMEM, i.e. to rule out potential quenching of the UV irradiation) and of the UV lamp-microtiter plate distance had no additional beneficial effects on the signal output and in fact indicated that the desired crosslinking failed.

Rather, a clear decrease in luciferase activity after photoactivation was observed, which might be attributed to degradation of either the PAL probes or of the biological system. In the latter case, the detrimental effects might originate from two different routes: first, the UV irradiation itself may directly damage the cellular environment (e.g. 254 nm); second, the photogenerated species could nonspecifically bind to essential cellular components and perturb their function (e.g. cell-cycle proteins). Another potential limiting factor includes quenching of the activating UV irradiation through intrinsic absorption by the PAL probes. To investigate this option, the UV absorption spectra of PAL probes **90a-c** were assessed from their corresponding LC-DAD-MS spectra (Figure V.5). For the benzophenone and diazirine probe no significant absorption is detected at their activation wavelength of 365 nm (Figure V.5A and V.5C, respectively). On the other hand, the aryl azide probe **90b** displays partial absorption at 254 nm, hence potentially impeding efficient photoexcitation (Figure V.5B).

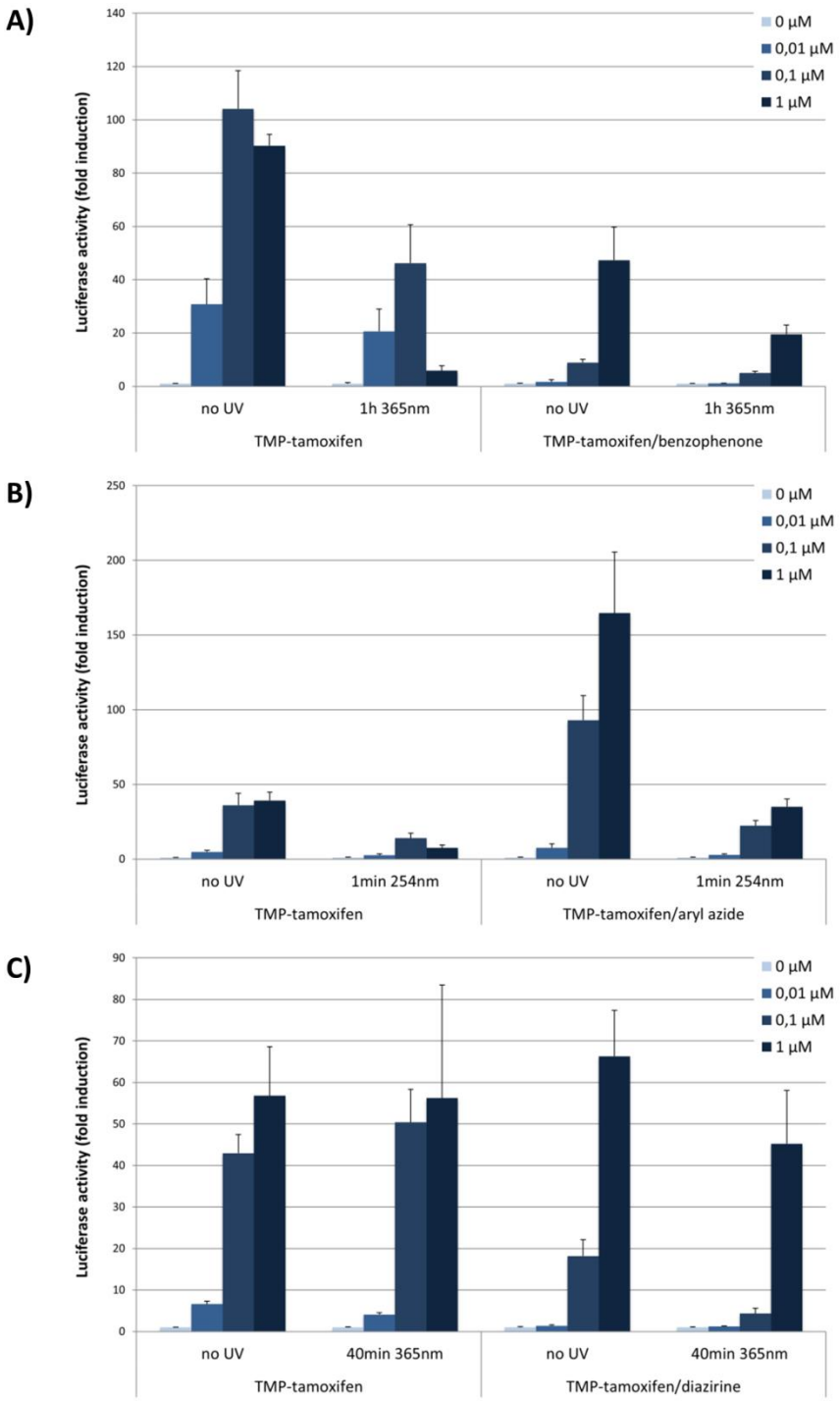


Figure V.4. Evaluation of the interaction of the second-generation tamoxifen-TFC and the PAL probes with candidate target protein '2' in MASPIT, in the absence and presence of UV irradiation. A) TMP-TAM-*benzophenone* (**90a**); **B)** TMP-TAM-*aryl azide* (**90b**); **C)** TMP-TAM-*diazirine* (**90c**). For each probe the irradiation time and wavelength is indicated. Luciferase signals are expressed as fold induction relative to a control sample treated with cytokine without fusion compound.

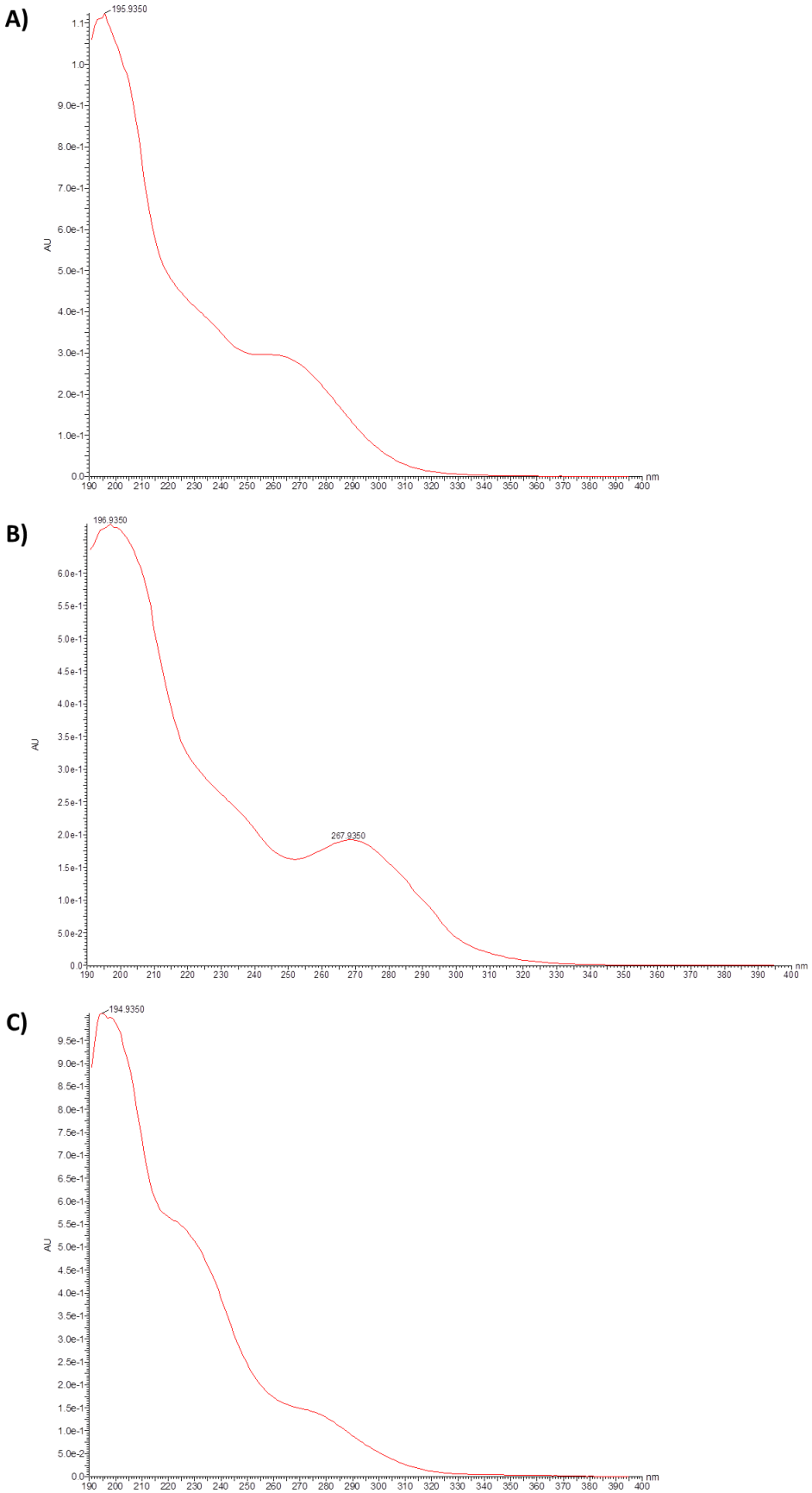


Figure V.5: UV absorption profiles (190-400 nm) of the PAL probes extracted from their corresponding LC-DAD-MS spectra. A) TMP-TAM-*benzophenone* (**90a**); **B)** TMP-TAM-*aryl azide* (**90b**); **C)** TMP-TAM-*diazirine* (**90c**).

Finally, besides functional testing in MASPIT, a more direct evaluation approach was explored using Western blot (WB) analysis. Inspired by a report of the Cornish group on the development of a covalent TMP-tag,[26] we envisioned that successful labeling of target preys with low molecular weight, such as ER1 (86 kDa) or candidate target protein '8' (37 kDa), with the photoactivated PAL probes **90a-c** (~1.3 kDa), would yield a detectable shift of the corresponding protein bands in the WB. However, also in this setup no proof of effective photocrosslinking could be provided.

V.4. Taking a short cut: a test system based on EGFR kinase inhibitors

As we might have set the bar a bit too high for the pilot PAL probes, we decided to take a step back and explore a test system that would enable the assessment of the actual contribution of a covalent linkage on the bait-end of the dimerizers with respect to the MASPIT readout, before undertaking any further empirical optimization efforts regarding design and technological aspects of the photocrosslinker approach.

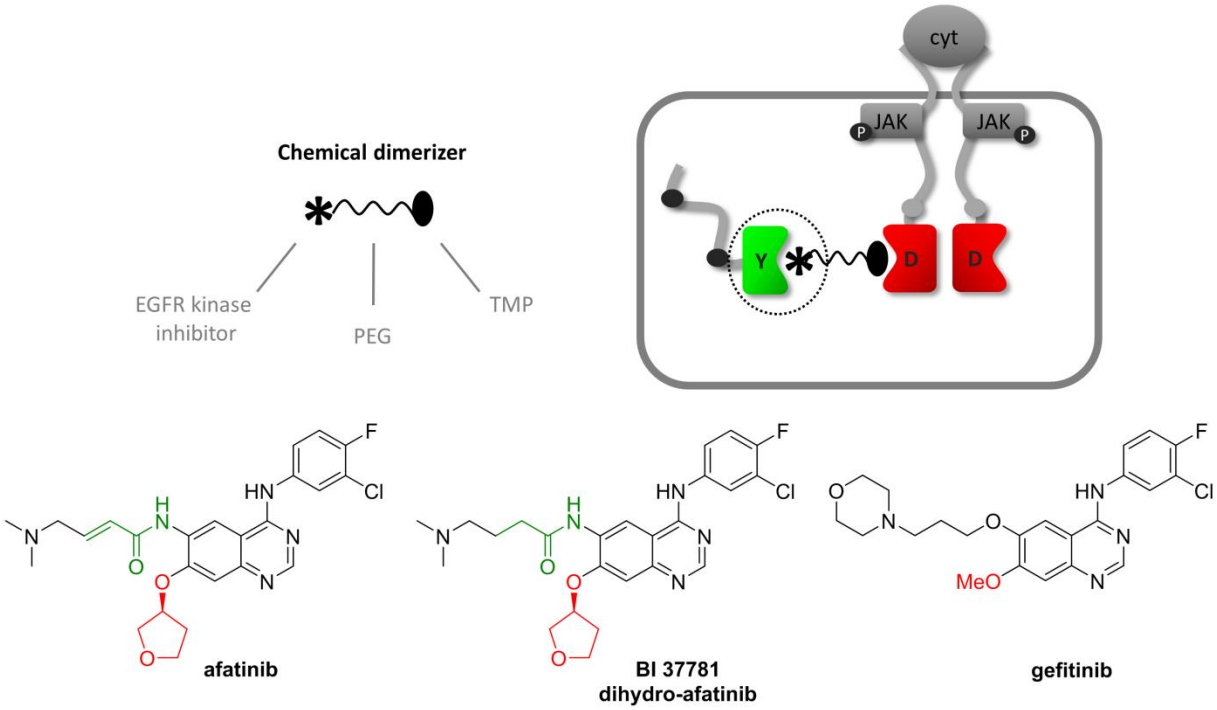


Figure V.6: Outline of the EGFR kinase inhibitor-based test system. For each inhibitor, the functional group selected as conjugation site is highlighted in red. The acrylamide/alkylamide moiety of afatinib/dihydro-afatinib is highlighted in green.

Toward this end, we were inspired by the trend observed in the epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (TKI) field, evolving over the past decades from reversible to irreversible covalent inhibitors to tackle mutant kinases, which developed resistance toward the early generation inhibitors. More specifically, we focused on a panel of three TKIs, including the recently approved afatinib; its closely related analogue BI 37781, which we coined dihydro-afatinib; and the prototypic gefitinib (Figure V.6).[27] By comparing the irreversible, covalent binding afatinib with its α,β-saturated congener, which lacks the covalent reactivity while maintaining binding affinity (K_i ($EGFR^{WT}$ kinase) = 0.8 nM vs. 0.5 nM for afatinib),[27] we estimated that this test system would allow assessment of the contribution of covalent bonding with respect to the MASPIT signal output.

However, since the dihydro-afatinib analogue represents a rather artificial reversible binder, we decided to also incorporate the natural reversible first-generation TKI gefitinib in our analysis, providing a positive control and setting at the same time a reference point for the readout attainable solely based on reversible interactions.

In the following subsections some background on the mechanism of action and binding mode of these TKIs is provided as well as our rationale for the selection of the conjugation site on these baits allowing their incorporation into appropriate TFCs.

V.4.1. MOA of gefitinib and afatinib

Gefitinib, a 4-anilinoquinazoline derivative, is the first reversible, ATP-competitive EGFR tyrosine kinase inhibitor that was approved by the FDA in 2003 for the treatment of non-small-cell lung carcinoma (NSCLC).[28] EGFR/ErbB-1/HER1 together with the closely related ErbB-2/HER2, ErbB-3/HER3 and ErbB-4/HER4 constitute the ErbB (proto-oncogene B of the avian erythroblastosis virus AEV-H strain) family of tyrosine kinase receptors.[29] Structurally, each receptor is composed of an extracellular ligand-binding domain, a transmembrane domain, and an intracellular part containing the catalytic tyrosine kinase (TK) domain, except for ErbB3 which lacks intrinsic TK activity.[30] Deregulation of the ErbB receptor network is well recognized as an oncogenic driver in a variety of epithelial malignancies, making this protein kinase family an attractive drug target.[31],[32]

Yet, the ability of gefitinib to treat NSCLC patients effectively is short-lived as it is seen that such patients typically relapse within 6-8 months of treatment.[33] This observation is partly attributed to acquired drug resistant mutations in EGFR, which affect approximately 50% of the treated population.[34],[35] One of these secondary mutations, accounting for about half of all resistance to gefitinib, involves substitution of the threonine gatekeeper residue 790 with methionine (T790M). This point mutation confers resistance by increasing the affinity for the competing physiologic substrate (ATP), rather than by interfering with specific inhibitor binding.[36] As such, T790M substitution tends to cause resistance to any ATP-competitive inhibitor, designating it as a 'generic' resistance mutation. An additional postulated mechanism of acquired resistance to gefitinib comprises increased internalization of ligand-bound EGFR, as observed in drug resistant tumor cell lines.[37] In fact, significant EGF-dependent signaling is thought to occur during the process of internalization,[38] and such alterations in receptor trafficking may be correlated with dissociation of the reversible gefitinib-EGFR complex at the low pH of intracellular vesicles.

Both resistance mechanisms may be circumvented by irreversible EGFR-TK inhibitors such as afatinib, which was approved by the FDA in 2013 for the first-line treatment of NSCLC in patients bearing activating oncogenic mutations.[39] Afatinib relies on an electrophilic acrylamide Michael acceptor grafted onto its anilinoquinazoline pharmacophore to form a covalent bond with conserved cysteine residues within the catalytic domains of EGFR (Cys797), HER2 (Cys805) and ErbB-4 (Cys803) (Figure V.7).[27] In fact, this Michael addition enables a greater occupancy of the ATP-binding site and prolonged time of exposure compared to gefitinib, thus ensuring irreversible inhibition of the enzymatic activity of these ErbB family kinases. As a result, afatinib is no longer in a competitive, reversible equilibrium with ATP and hence capable of overcoming T790M resistance. Moreover,

Kwak *et al.* assumed that irreversible inhibition of EGFR would be unaffected by alterations in receptor trafficking.[37]

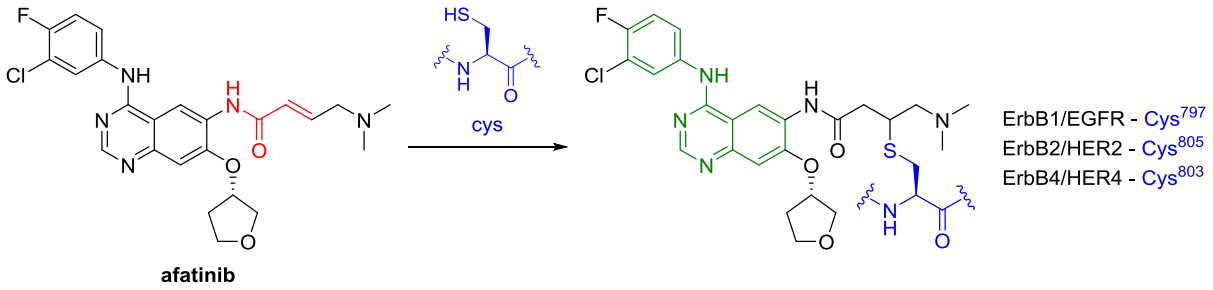


Figure V.7: Covalent binding mode of afatinib to conserved cysteine residues within the catalytic domains of ErbB receptor family members. The electrophilic acrylamide Michael acceptor and ATP-competitive 4-anilinoquinazoline moiety of afatinib are highlighted in red and green, respectively. Figure adapted from: *The Journal of Pharmacology and Experimental Therapeutics* 2012, 343 (2), 342-350.

V.4.2. SAR and conjugation site selection

Apart from the covalent bonding aspect, the binding modes of afatinib and gefitinib in the kinase domain of human wild-type EGFR are strikingly similar (PDB codes: 4G5J, 2ITY; Figures V.8,V.9, respectively).[27],[40] For either inhibitor, the quinazoline ring is positioned in the back of the ATP-binding pocket where its orientation is governed by a critical hydrogen bond between its N^1 and the amide nitrogen of Met793 at the hinge region. Further down the ATP-binding cleft, the 3-chloro-4-fluoro aniline substituent extends into a hydrophobic pocket.

In the case of afatinib (Figure V.8), the crystal structure unambiguously proves the formation of a covalent bond between Cys797 at the edge of the active site and the β-carbon of the acrylamide moiety installed at the 6 position of the heterocycle. Hence, so as not to disturb this crucial covalent bonding mechanism, we selected the 7 position of the quinazoline to introduce an alkyne ligation handle. This strategy was based on a report by Kobus *et al.*, describing an afatinib analogue bearing a ^{18}F label at the 7 position that was successfully installed following a CuAAC approach.[41] Encouragingly, profiling of this conjugate against a panel of 117 kinases revealed that it retained high binding affinities for its primary targets, EGFR and ErbB4.

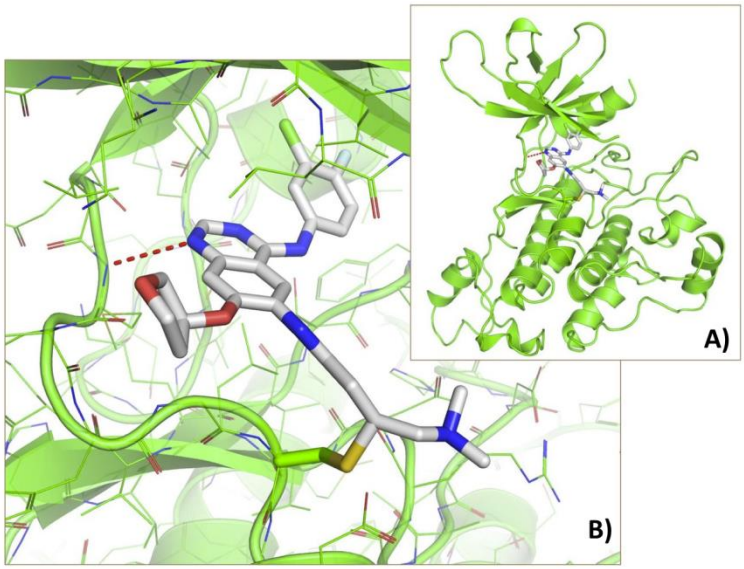


Figure V.8: Binding mode of afatinib in the active site of human wild-type EGFR (PDB code: 4G5J). **A)** The structure of the whole EGFR kinase domain. **B)** A close-up of the hinge region illustrating the covalent C-S bond (1.82 Å) formed between Cys797 at the edge of the active site and the β-carbon of the acrylamide Michael acceptor moiety of afatinib. The red dashed line represents a hydrogen bond (3.3 Å) between the amide nitrogen of Met793 at the hinge region and N^1 of the quinazoline core of the inhibitor. Figure from: *The Journal of Pharmacology and Experimental Therapeutics* 2012, 343 (2), 342-350.

Continuing along the same line, the 7 position of the quinazoline ring of gefitinib was explored as derivatization site. Although the methoxy group in this position is in van der Waals contact with Gly796, we anticipated that its substitution for an alkynoxy moiety would not significantly influence overall binding affinity. To the best of our knowledge, no literature data on the conjugation of gefitinib at this specific site exists. Rather, the 6-propylmorpholino group has received more attention, as structural data revealed that it is poorly ordered and extends into the solvent (Figure V.9).[40] These findings are in agreement with the reported SAR for gefitinib, since this substituent was exclusively implemented to improve the pharmacokinetic profile of this kinase inhibitor.[28] Hence, by exchanging the morpholine for a piperazine, a strategy previously described in Chapter 3 for the reversine bait, an appropriate ligation handle might be attached, as exemplified by Hill *et al*.[42] Nevertheless, in order to allow optimal comparison between either interaction mode (covalent vs. reversible), we decided to keep the substituent at the 6 position of the quinazoline intact (as in the case of afatinib) and to focus our ligation efforts on the 7 position, enabling a synthetic route starting from the commercially available drug.

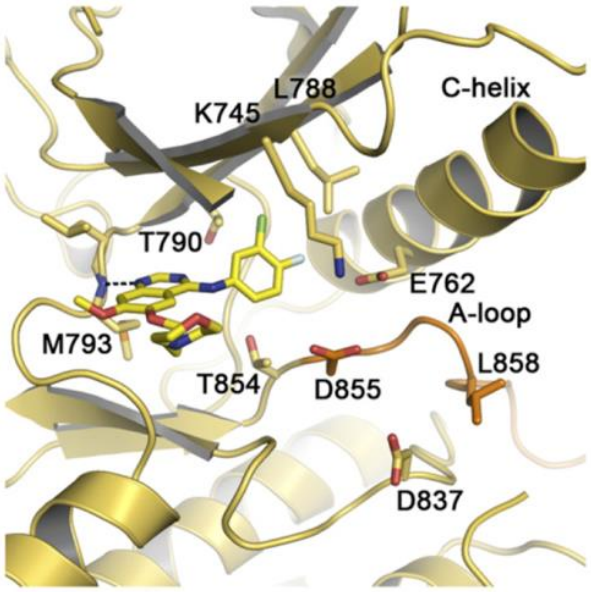
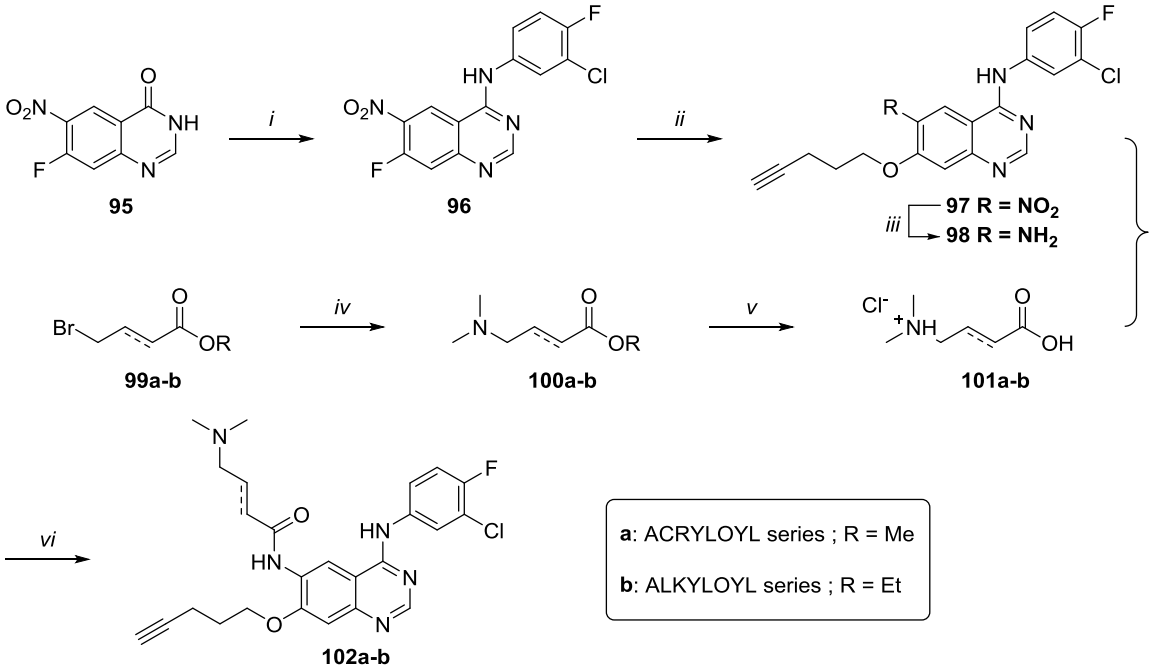


Figure V.9: Binding mode of gefitinib in the active site of human wild-type EGFR (PDB code: 2ITY). Key side chains are labeled. The black dashed line represents a hydrogen bond between the amide of Met793 at the hinge region and N^1 of the quinazoline core of the inhibitor. Figure from: *Cancer Cell* 2007, 11 (3), 217-227.

V.5. Synthesis of EGFR kinase inhibitor TFCs

The alkyne-functionalized afatinib **102a** and its α,β-saturated analogue, dihydro-afatinib **102b**, were synthesized based on a protocol by Kobus *et al.*[41] with modifications (Scheme V.5). Both alkynes were derived from a common 6-aminoquinazoline precursor **98**, which was obtained from 7-fluoro-6-nitroquinazolin-4(3*H*)-one (**95**) in 3 steps. The latter was first chlorinated and subsequently substituted with 3-chloro-4-fluoroaniline to give **96**. In a next substitution step, the alkyne ligation handle was attached using 4-pentynol, after which selective tin(II) mediated aromatic nitro reduction[43] afforded the desired aniline building block **98**. Furthermore, the divergent side chains **101a-b** were prepared from methyl 4-bromocrotonate (**99a**) and ethyl 4-bromobutyrate (**99b**), respectively. Substitution of these esters with dimethylamine, followed by saponification with concomitant acidification, provided the acid hydrochlorides **101a-b**.

Ultimately, these acids were condensed with aniline **98** in the presence of Mukaiyama's reagent (2-chloro-1-methylpyridinium iodide)[44] to yield the final alkyne-functionalized analogues **102a-b** in 30% and 53% yield, respectively. Alternative methods for this amide condensation, including HATU-mediated coupling and (in situ) acid chloride formation using oxalyl chloride[41] or Ghosez's reagent (tetramethyl-α-chloroenamine) followed by treatment with the aniline, were unsuccessful.



Scheme V.5: Synthesis of alkyne-functionalized (dihydro)-afatinib. [i] first $SOCl_2$, DMF, Δ; then 3-chloro-4-fluoroaniline, Et_3N, *i*PrOH, Δ, 92%; [ii] 4-pentynol, NaH, DMF, 82%; [iii] $SnCl_2$, EtOH, Δ, 73%; [iv] $NH(Me)_2$, THF, 0°C to RT, (**a** 91%; **b** 21%); [v] first NaOH, MeOH, 50°C; then HCl, (**a** 32%; **b** quant.); [vi] Mukaiyama's reagent, Et_3N, CH_2Cl_2/DMF, Δ, (**a** 30%; **b** 53%).

The desired afatinib TFC **103a** and dihydro-afatinib TFC **103b** (Figure V.10) were constructed by conjugating alkynes **102a-b** with the second-generation TMP-N_3 reagent (**40**) via CuAAC in 56% and 69% yield, respectively, after HPLC purification.

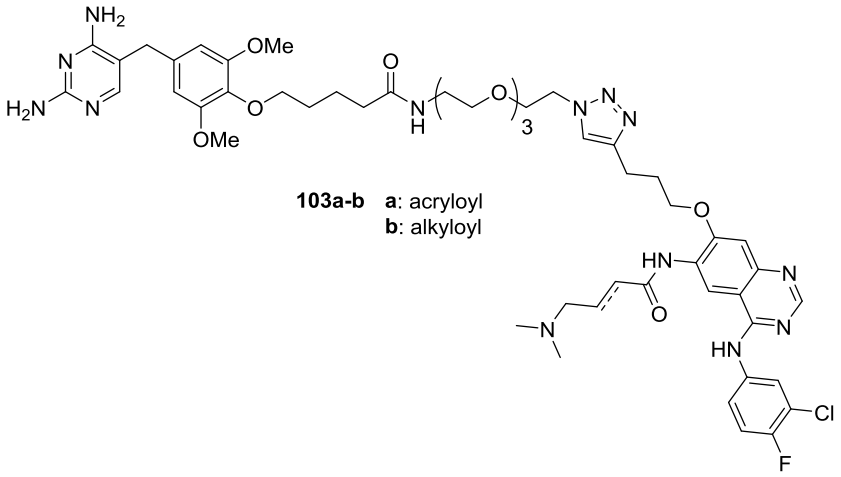
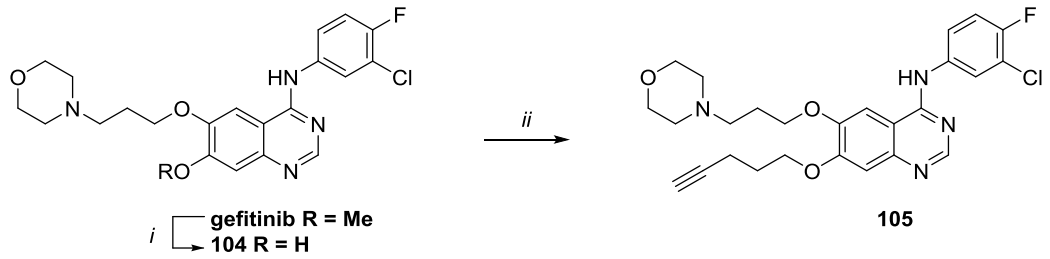


Figure V.10: Chemical structures of (dihydro)-afatinib TFCs.

To acquire an alkyne-functionalized analogue of gefitinib, the latter was first selectively demethylated using pyridinium chloride under melt conditions[45] (Scheme V.6). The same pentynol-based ligation handle as in the case of afatinib was next grafted onto the resulting 7-hydroxyquinazoline **104** via Mitsunobu reaction[45] to afford the final terminal alkyne **105** in satisfactory yield.



Scheme V.6: Synthesis of alkyne-functionalized gefitinib. [i] pyridinium chloride, 170°C, 76%; [ii] 4-pentynol, DBAD, PPh_3, THF, 0°C to 70°C, 57%.

The alkyne-functionalized gefitinib **105** was analogously click-coupled to the TMP-N_3 reagent **40** to generate the desired TFC **106** in 74% yield after HPLC purification (Figure V.11).

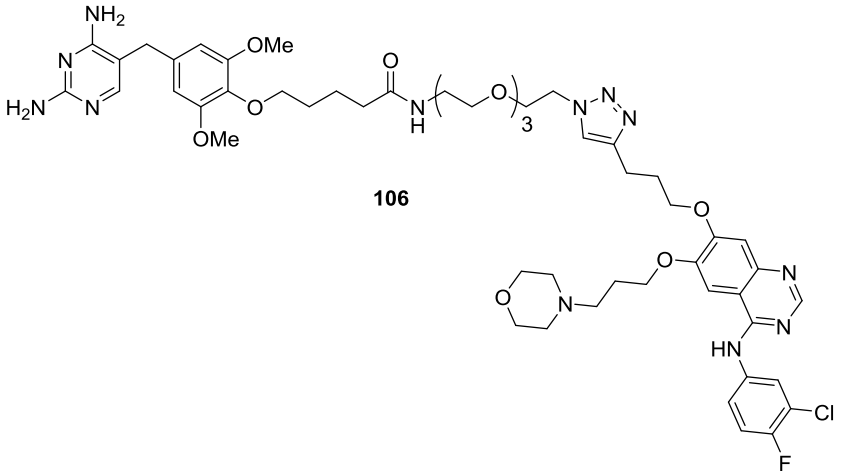


Figure V.11: Chemical structure of gefitinib TFC.

V.6. Biological evaluation of EGFR kinase inhibitor TFCs

The (dihydro)-afatinib and gefitinib TFCs (**103a-b** and **106**) were first evaluated in MASPIT for binding to their primary target protein, EGFR. Given the transmembrane nature of this ErbB family receptor, two MASPIT-compatible target prey constructs were prepared (see Experimental Section), i.e. the entire cytoplasmic tail of EGFR (cyt) and its protein kinase domain (kinase), comprising amino acid residues 696-1022 (analogous to the domain applied in the reported cocrystal structures of these TKIs, PDB codes: 4G5J, 2ITY). Unexpectedly, for both constructs only a signal for the covalently binding afatinib TFC **103a** was obtained, despite the comparable binding affinities of unfunctionalized afatinib and gefitinib toward wild-type EGFR (K_D = 0.25 nM and 1 nM, respectively)[46] (Figure V.12).

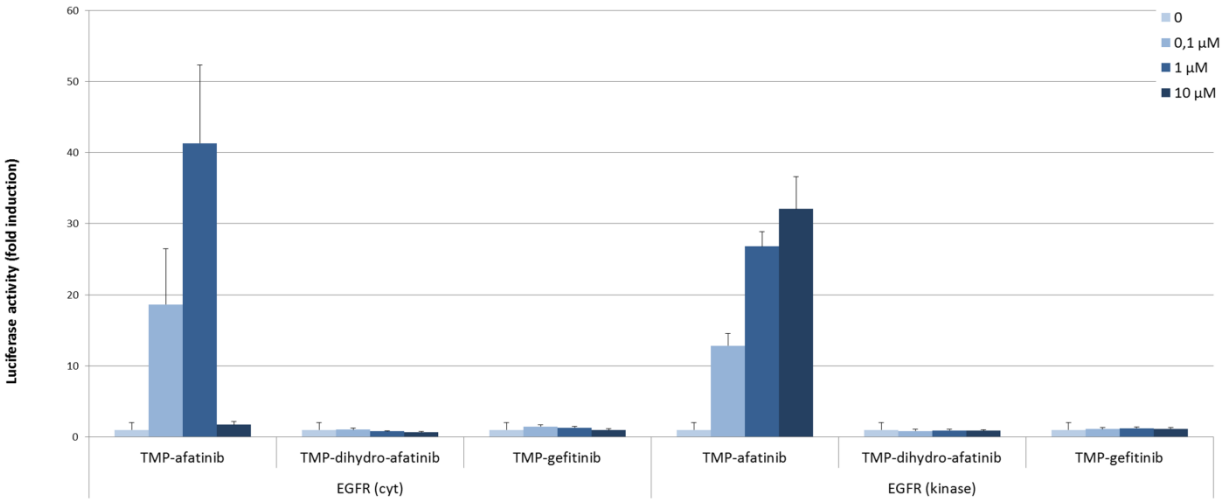


Figure V.12: Evaluation of EGFR kinase inhibitor TFCs 103a-b and 106 for binding to their primary target protein in MASPIT. Cyt and kinase denote the cytoplasmic and kinase domain of EGFR, respectively. Luciferase signals are expressed as fold induction relative to a control sample treated with cytokine without fusion compound.

Therefore, in a next step we set out to elucidate the hypothesis that modification of these kinase inhibitors might result in significant loss in binding affinity to their primary target. Toward this end, the cellular activity of the alkynylated analogues (**102a-b** and **105**) and their final TMP conjugates (**103a-b** and **106**) was assessed using a functional WB-based assay, monitoring the phosphorylation status of ERK1/2 (extracellular signal-regulated kinase), one of the main downstream signaling molecules within the EGFR signaling pathway.

This signaling cascade is initiated by extracellular binding of the EGF (epidermal growth factor) ligand, which triggers a conformational change in the extracellular domain of this receptor, thereby allowing its homodimerization and/or heterodimerization with other ErbB family members. In such dimers, one monomer allosterically activates the intrinsic intracellular catalytic tyrosine kinase activity of the partnered receptor. As a result, specific tyrosine residues within the C-terminal tail segment of the activating partner become transphosphorylated, thus providing a scaffold for the recruitment of adaptor and effector proteins. The latter are responsible for onward transmission of the signal via stimulation of their corresponding signaling cascades, including the KRAS-BRAF-MEK-ERK pathway,

PI3K-AKT-mTOR pathway, JAK-STAT pathway, and the PLCγ-PKC pathway. Ultimately, this leads to cell proliferation, angiogenesis, migration and anti-apoptosis (Figure V.13).[47],[48],[27],[49]

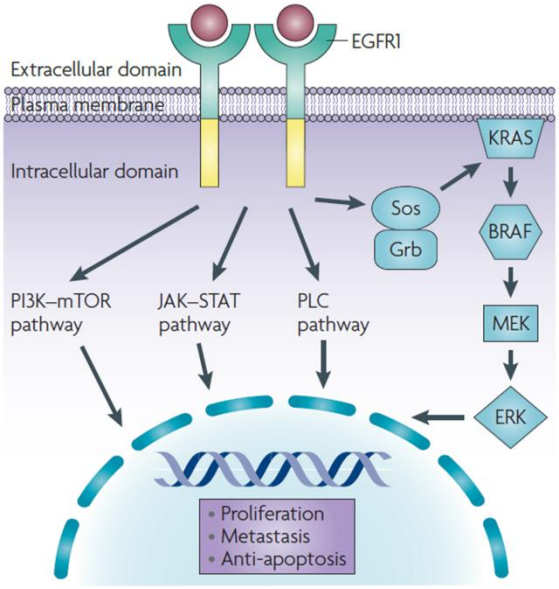


Figure V.13: Schematic outline of the epidermal growth factor receptor (EGFR) signaling pathway. Ligand-induced homodimerization of EGFR results in activation of the intracellular kinase domain and subsequent phosphorylation of tyrosine residues in the C-terminal regulatory domains, which serve as docking sites for signaling molecules that engage multiple signaling cascades including the KRAS-BRAF-MEK-ERK pathway (see text for more details). BRAF, B-rapidly accelerated fibrosarcoma; ERK, extracellular signal-regulated kinase; KRAS, Kirsten rat sarcoma; MEK, MAPK/ERK kinase. Figure from: *Nature Reviews Cancer* 2009, 9 (7), 489-499.

Thus, stimulation of intact MASPIT HEK293T cells, which endogenously express the EGF receptor, with human EGF results in phosphorylation of the downstream ERK1/2 kinases, as observed in the WB below (see Experimental Section; Figure V.14, lane 2). Accordingly, functional binding and cellular activity of an appropriate EGFR inhibitor can be read out as inhibition of ERK phosphorylation, as in the case of unfunctionalized gefitinib, serving as a positive control. Likewise, all alkyne-functionalized analogues (**102a-b** and **105**) clearly showed cellular activity, thereby validating our rationale for the selection of the conjugation site on these baits. However, as deduced from the intense pERK bands, even at concentrations up to 10 µM, no functional response for all corresponding TFCs (**103a-b** and **106**) was observed.

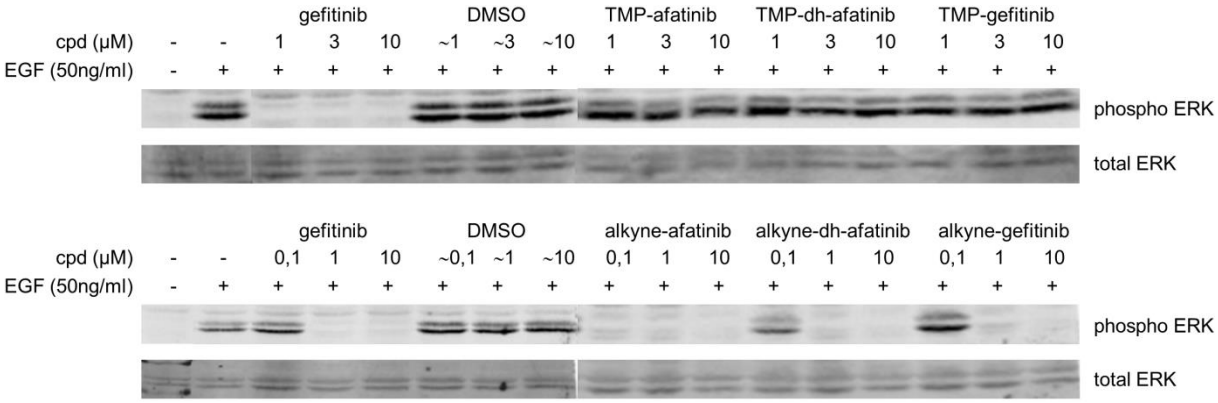


Figure V.14: Functional Western blot-based analysis determining the cellular activity of the alkyne-functionalized EGFR inhibitors (102a-b and 105) and their corresponding TMP conjugates (**103a-b and 106**). For each of the applied compound concentrations, the corresponding amount of pure DMSO was taken along as control. cpd, compound; dh, dihydro; EGF, epidermal growth factor.

The latter conflicting results lead us to the following two assumptions. First, despite the fact that implementation of the alkyne ligation handle seems to be well-tolerated, coupling to the TMP-N_3 reagent might affect overall binding affinity of the final conjugates. Second, although counterintuitive, the TMP fusion compounds might suffer from intrinsic cell permeability issues.

To explore the first postulation, we decided to determine and compare the potency of the alkynylated analogues and final TFCs using an *in vitro* (cell-free) radiometric kinase assay. The latter was performed by CEREP - Eurofins Pharma Discovery Services UK Limited (Dundee, UK), employing full-length human EGFR, γ-^{33}P-ATP, a poly(Glu, Tyr) 4:1 peptide substrate and unmodified gefitinib as control compound. Biochemical IC_{50} values resulting from these competition experiments are summarized below (Table V.1).

Table V.1: Potency of the alkyne-functionalized EGFR inhibitors (102a-b and 105) and their corresponding TMP conjugates (103a-b and 106) in a radiometric kinase assay.

Compound	EGFR inhibition IC_{50} (nM)
gefitinib	3
alkyne-gefitinib (**105**)	5
TMP-gefitinib (**106**)	7
alkyne-afatinib (**102a**)	3
TMP-afatinib (**103a**)	4
alkyne-dh-afatinib (**102b**)	4
TMP-dh-afatinib (**103b**)	6

Abbreviations used: dh, dihydro.

In agreement with the gefitinib control, all tested compounds exhibited single digit nanomolar IC_{50} values, and no significant differences among alkyne-functionalized analogues and their

corresponding TMP conjugates were observed. Hence, yet contradictory to the pERK WB analysis, these data demonstrate that coupling of the TMP anchor moiety does not hamper overall binding of the TFCs to EGFR. Moreover, these data clearly validate our rationale and pioneering work regarding the conjugation of gefitinib at the 7 position of the quinazoline ring (see Section V.4.2).

To evaluate the possibility that a lack of cellular permeability of the TMP conjugates is causing the observed absence of effect in the cellular assays, the cellular uptake of the (dihydro)-afatinib and gefitinib TFCs (**103a-b** and **106**) was indirectly determined by means of a MASPIT-based competition experiment strategy (see Experimental Section). Toward this end, we reasoned that increasing concentrations of cell-permeable competing small-molecule fusion compounds would enable the gradual disruption of an established three-hybrid interaction in MASPIT, thereby yielding a decrease in reporter readout following a sigmoidal inhibition curve. In fact, considering the interaction between the CR-eDHFR chimeric anchor protein, the second-generation tamoxifen TFC (**41**) and the ER1 target prey, two competition interfaces can be distinguished (Figure V.15).

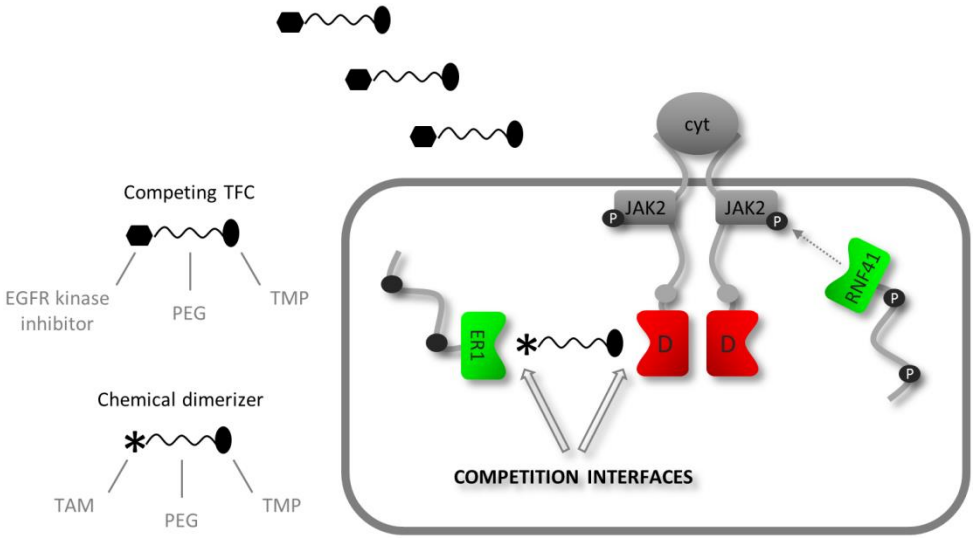


Figure V.15: Schematic outline of the competition experiment strategy to indirectly asses cell permeability of EGFR kinase inhibitor TFCs. The two competition interfaces on the target prey and anchor protein side are indicated, as well as the control interaction between RNF41 (ring finger protein 41) and JAK2 (Janus kinase 2).

First, focusing on the target prey side, the interaction between **41** and ER1 was effectively disrupted using unmodified tamoxifen as a positive control (Figure V.16A). Additionally, in order to further validate the competition approach at this specific interface, simvastatin was taken along as negative control compound. As anticipated, since the latter lacks binding affinity toward ER1, no competition effect was observed, resulting in a flat inhibition curve (Figure V.16C).

Similarly, we studied interference with the chemically-induced dimerization and proximity events, underlying successful MASPIT signaling, on the eDHFR anchor protein side starting with an unfunctionalized trimethoprim control (Figure V.16B). The inhibition curve and corresponding IC_{50} value obtained for TMP nicely complement our earlier findings reflecting the excellent cell permeability properties of this DHFR inhibitor (see Chapter 4). Moreover, the first-generation simvastatin TFC (**29c**) convincingly demonstrated that this type of competition is attainable with

appropriate TMP-based fusion compounds as well (Figure V.16D). Markedly, all EGFR kinase inhibitor TFCs (**103a-b** and **106**) displayed clear inhibition profiles and comparable IC_{50} values, which seems to indicate that these compounds are indeed effectively taken up by the HEK cells, thereby excluding our second assumption. (Figure V.16E-G).

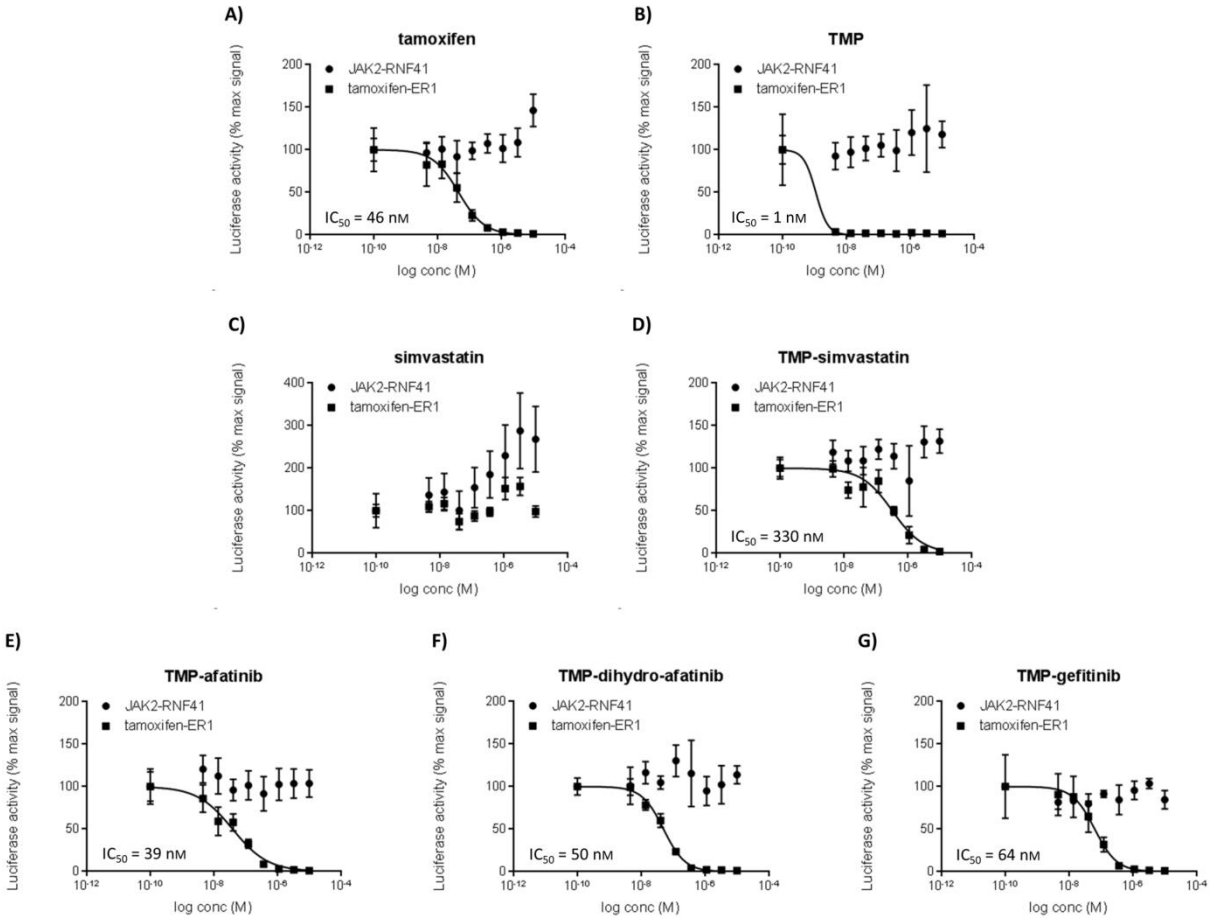


Figure V.16: Inhibition curves resulting from MASPIT competition experiments based on the CR-eDHFR/TMP-TAM/ER1 three-hybrid interaction. A) tamoxifen; **B)** trimethoprim; **C)** simvastatin; **D)** simvastatin TFC (**29c**); **E)** afatinib TFC (**103a**); **F)** dihydro-afatinib TFC (**103b**); **G)** gefitinib TFC (**106**). For each inhibition curve, the corresponding IC_{50} value of the evaluated compound is indicated. JAK2, Janus kinase 2; RNF41, ring finger protein 41.

Of note, all competition experiments described above were run in parallel with control experiments based on the RNF41-JAK2 interaction. As schematically depicted in Figure V.15, the RNF41 prey directly binds the associated JAK2 kinases of the activated CR-eDHFR chimeric receptor, thereby short-circuiting the MASPIT signaling cascade and inducing luciferase reporter gene expression. Hence, any (concentration-dependent) detrimental direct or indirect effect of the evaluated compounds on this interaction, could be attributed to either toxicity or interference with the MASPIT system, which might erroneously influence interpretation of the actual inhibition data. However, as for each competition experiment in this setup an overall flat line is observed, this does not seem to be the case.

In conclusion, currently we are facing a clear discrepancy between the results obtained from the radiometric kinase assay/competition experiments on the one hand, and the MASPIT/pERK WB

assays on the other. Whereas the first two analyses clearly demonstrate the high *in vitro* potency/effective cellular uptake of the EGFR kinase inhibitor TFCs, the latter curiously fail to show any MASPIT activity for the reversible inhibitor probes/any cellular functional effects for all TMP conjugates.

Therefore, future research will be focused on the evaluation of the interaction of these kinase inhibitor probes with alternative targets in MASPIT. As a starting point, the other two ErbB family members, HER2 and HER4, for which the same Michael addition-based covalent binding mode with afatinib is described, will be tested (see Figure V.7). Furthermore, based on a comprehensive analysis by Davis et al.,[46] we plan to analyze all additional (cytosolic) target kinases for which afatinib displays a K_D value below 250 nM, which is currently the estimated cutoff of the MASPIT system (Table V.2).

Table V.2: Highest affinity target kinases of afatinib and gefitinib.

Compound	Kinase	K_D (nM)
Afatinib	EGFR	0.25
	ErbB2	5
	ErbB4	6.3
	GAK	79
	BLK	220
	IRAK1	240
Gefitinib	EGFR	1
	GAK	13
	IRAK1	69
	YSK4	240

Table adapted from: *Nature Biotechnology* 2011, 29 (11), 1046-1051.

V.7. Experimental Section

V.7.1. Synthesis

General: All reactions were performed under nitrogen and at ambient temperature, unless stated otherwise. Reagents and solvents were purchased from Sigma-Aldrich, Acros Organics, TCI Europe, Apollo Scientific, Carbosynth, or Activate Scientific, and used as received. Reactions were monitored by thin-layer chromatography on TLC aluminum sheets (Macherey-Nagel, Alugram Sil G/UV_{254}) with detection by spraying with a solution of $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$ (25 g/L) and $(NH_4)_4Ce(SO_4)_4{\cdot}2H_2O$ (10 g/L) in H_2SO_4 (10%) or $KMnO_4$ (20 g/L) and K_2CO_3 (10 g/L) in water followed by charring. Column chromatography was performed on a Grace Reveleris X2 flash system equipped with disposable silica gel cartridges (Grace, Reveleris). LC-MS analyses were carried out on a Waters Alliance 2695 XE Separation Module by using a Phenomenex Luna reversed-phase C18 column (100×2.00 mm, 3 µm) and a gradient system of HCOOH in H_2O (0.1 %, *v/v*)/HCOOH in CH_3CN (0.1 %, *v/v*) at a flow rate of 0.4 mL/min. High-resolution spectra were recorded on a Waters LCT Premier XE Mass spectrometer. ^1H and ^{13}C NMR spectra were measured on a Varian Mercury-300BB (300/75 MHz) spectrometer. NMR solvents were purchased from Euriso-top. Chemical shifts (δ) are given in ppm relative to tetramethylsilane (^1H NMR) or $CDCl_3$, CD_3OD or $SO(CD_3)_2$ (^{13}C NMR) as internal standards. Coupling constants are given in Hz. Preparative HPLC purifications were carried out by using a Laprep preparative RP-HPLC system equipped with a Phenomenex Luna C18 column (21.20×250 mm, 5 µm) with a gradient system of HCOOH in H_2O (0.2 %, *v/v*)/CH_3CN at a flow rate of 17.5 mL/min.

TMP-NH_2 (87): The second-generation TMP-N_3 reagent **40** (386 mg, 0.67 mmol) was dissolved in H_2O/1,4-dioxane (44 mL, 10:1, v/v), and the system was purged with nitrogen for 10 min. Subsequently, the solution was charged with a catalytic amount of palladium on activated charcoal (Pd/C, 10% Pd basis, ~ 50 mg) after which hydrogen was bubbled through the reaction mixture for 1 h (1 atm, RT). Next, the system was flushed again with nitrogen for 10 min and the mixture was filtered. The filtrate was concentrated *in vacuo*, redissolved in H_2O (20 mL) and lyophilized to yield the title compound (331 mg, 0.60 mmol, 90%) as a white amorphous solid. ^1H NMR (300 MHz, CD_3OD) δ = 7.51 (s, 1H), 6.52 (s, 2H), 3.90 (t, 2H, *J* = 6.0 Hz), 3.77 (s, 6H), 3.68-3.57 (m, 12H), 3.53 (t, 2H, *J* = 5.6 Hz), 3.37 (t, 2H, *J* = 5.6 Hz), 3.07 (t, 2H, *J* = 5.0 Hz), 2.27 (t, 2H, *J* = 7.4 Hz), 1.86-1.64 ppm (m, 4H); ^{13}C NMR (75 MHz, CD_3OD) δ = 176.3, 164.5, 162.8, 155.1, 154.8, 136.6, 136.2, 108.3, 106.8, 73.9, 71.42, 71.41, 71.12, 71.11, 70.6, 68.7, 56.6, 40.9, 40.2, 36.7, 34.4, 30.5, 23.6 ppm; HRMS: calcd. for $C_{26}H_{44}N_6O_7$ $[M+2H]^{2+}$: 276.1630, found: 276.1619.

TMP-N_3-Lys(Boc) (88): TPTU (680 mg, 2.29 mmol, 1.97 eq.) and Et_3N (338 µL, 2.43 mmol, 2.09 eq.) were added to a solution of acid **92**[20] (557 mg, 2.05 mmol, 1.77 eq.) in DMF (1.9 mL). The resulting preactivation mixture was stirred for 5 min at RT. Subsequently, a solution of amine **87** (638 mg, 1.16 mmol, 1.0 eq.) and Et_3N (1.92 mL, 13.79 mmol, 11.89 eq.) in DMF (8.9 mL) was added dropwise to this mixture. The resulting reaction mixture was stirred for 29 h at RT, then concentrated *in vacuo*, and the residue was purified by silica gel chromatography (MeOH/CH_2Cl_2 0-10%, v/v with 0.1% NH_4OH) yielding the title compound (640 mg, 0.80 mmol, 69%) as a white foam. ^1H NMR (300 MHz, CD_3OD) δ = 7.45 (s, 1H), 6.52 (s, 2H), 3.91 (t, 2H, *J* = 6.2 Hz), 3.83-3.80 (m, 1H), 3.78 (s, 6H), 3.65-3.57 (m, 10H), 3.56-3.51 (m, 4H), 3.41-3.33 (m, 4H), 3.03 (t, 2H, *J* = 6.8 Hz), 2.28 (t, 2H, *J* = 7.2 Hz), 1.88-1.65 (m, 6H), 1.55-1.35 ppm (m, 13H); ^{13}C NMR (75 MHz, CD_3OD) δ = 176.2, 172.5, 164.9, 161.6,

158.5, 154.9, 152.4, 136.9, 135.7, 108.8, 106.9, 79.9, 73.9, 71.6, 71.3, 70.6, 70.4, 64.2, 56.6, 41.1, 40.4, 40.3, 36.7, 34.4, 32.5, 30.5, 28.8, 23.9, 23.7 ppm; HRMS: calcd. for $C_{37}H_{61}N_{10}O_{10}$ $[M+H]^+$: 805.4567, found: 805.4577.

TMP-N_3-Lys (89): Compound **88** (627 mg, 0.78 mmol) was taken up in TFA/CH_2Cl_2 (20 mL, 1:1, v/v), and the solution was stirred for 40 min at RT. The reaction mixture was then evaporated, coevaporated twice with toluene, and concentrated under high vacuum for 1 h. The residue was redissolved in H_2O (40 mL), after which the pH was adjusted to 10 by the addition of Amberlite IRA-400 hydroxide form. The resin was next filtered, and the filtrate lyophilized to yield the title compound (549 mg, 0.78 mmol, quant.) as a white solid. ^{1}H NMR (300 MHz, CD_3OD) δ = 7.51 (s, 1H), 6.51 (s, 2H), 3.90 (t, 2H, *J* = 6.2 Hz), 3.85 (dd, 1H, *J* = 7.5 Hz, *J* = 6.0 Hz), 3.78 (s, 6H), 3.65-3.57 (m, 10H), 3.57-3.50 (m, 4H), 3.42-3.33 (m, 4H), 2.77 (t, 2H, *J* = 7.4 Hz), 2.28 (t, 2H, *J* = 7.4 Hz), 1.87-1.65 (m, 6H), 1.65-1.36 ppm (m, 4H); ^{13}C NMR (75 MHz, CD_3OD) δ = 176.2, 172.3, 164.4, 163.2, 156.0, 154.9, 136.7, 136.4, 108.1, 106.8, 73.9, 71.6, 71.3, 71.2, 70.6, 70.3, 64.2, 56.6, 41.4, 40.4, 40.3, 36.7, 34.5, 32.4, 30.8, 30.5, 23.8, 23.7 ppm; HRMS: calcd. for $C_{32}H_{54}N_{10}O_8$ $[M+2H]^{2+}$: 353.2058, found: 353.2050.

TMP-TAM-benzophenone (90a): Azide **89** (70.5 mg, 100 µmol, 1.0 eq.) was taken up in DMF (1.13 mL), and alkyne-functionalized tamoxifen **7** (59.3 mg, 150 µmol, 1.5 eq.), $CuSO_4$ (40 µL, 0.5M, 0.2 eq.), Na ascorbate (200 µL, 0.5M, 1.0 eq.), and Et_3N (14 µL, 100 µmol, 1.0 eq.) were added. The resulting mixture was charged with a catalytic amount of TBTA[50] and vigorously stirred at RT. Upon completion of the reaction (1 h, based on HRMS analysis), NHS ester **94a**[21] (38.8 mg, 120 µmol, 1.2 eq.) was added to the crude mixture, which was stirred for another 16 h in the dark. The solution was next concentrated *in vacuo* and the residue was purified by preparative RP-HPLC (30-80% MeCN) yielding the title compound (39.6 mg, 30 µmol, 30%) as a white amorphous solid. LC-HRMS: t_R = 6.13 min (35-60% MeCN, 15 min run); HRMS: calcd. for $C_{74}H_{91}N_{11}O_{11}$ $[M+2H]^{2+}$: 654.8444, found: 654.8440.

TMP-TAM-aryl azide (90b): Azide **89** (84.6 mg, 120 µmol, 1.2 eq.) was taken up in DMF (1.13 mL), and alkyne-functionalized tamoxifen **7** (39.6 mg, 100 µmol, 1.0 eq.), $CuSO_4$ (40 µL, 0.5M, 0.2 eq.), Na ascorbate (200 µL, 0.5M, 1.0 eq.), and Et_3N (14 µL, 100 µmol, 1.0 eq.) were added. The resulting mixture was charged with a catalytic amount of TBTA[50] and vigorously stirred at RT. Upon completion of the reaction (2 h, based on HRMS analysis), NHS ester **94b**[22] (37.5 mg, 144 µmol, 1.4 eq.) was added to the crude mixture, which was stirred for another 16 h in the dark. The solution was next concentrated *in vacuo* and the residue was purified by preparative RP-HPLC (30-80% MeCN) yielding the title compound (29.9 mg, 24 µmol, 24%) as a white amorphous solid. LC-HRMS: t_R = 5.52 min (35-60% MeCN, 15 min run); HRMS: calcd. for $C_{67}H_{86}N_{14}O_{10}$ $[M+2H]^{2+}$: 623.3320, found: 623.3311.

TMP-TAM-diazirine (90c): Azide **89** (70.5 mg, 100 µmol, 1.0 eq.) was taken up in DMF (1.13 mL), and alkyne-functionalized tamoxifen **7** (42.4 mg, 107 µmol, 1.1 eq.), $CuSO_4$ (40 µL, 0.5M, 0.2 eq.), Na ascorbate (200 µL, 0.5M, 1.0 eq.), and Et_3N (14 µL, 100 µmol, 1.0 eq.) were added. The resulting mixture was charged with a catalytic amount of TBTA[50] and vigorously stirred at RT. Upon completion of the reaction (6 h, based on HRMS analysis), NHS ester **94c**[23] (49.1 mg, 150 µmol, 1.5 eq.) was added to the crude mixture, which was stirred for another 21 h in the dark. The solution was

next concentrated *in vacuo* and the residue was purified by preparative RP-HPLC (30-80% MeCN) yielding the title compound (45.1 mg, 34 µmol, 34%) as a white amorphous solid. LC-HRMS: t_R = 6.84 min (30-80% MeCN, 15 min run); HRMS: calcd. for $C_{69}H_{86}F_3N_{13}O_{10}$ $[M+2H]^{2+}$: 656.8281, found: 656.8282.

(*S*)-2-azido-6-((*tert*-butoxycarbonyl)amino)hexanoic acid (92): Preparation of triflyl azide (TfN_3) solution **(*Caution*: TfN_3 is explosive when not in solvent and should always be used as a solution!)**: To an ice-cooled solution of sodium azide (7.0 g, 107.7 mmol) in H_2O/CH_2Cl_2 (40 mL, 1:1, v/v), triflic anhydride (Tf_2O) (4.0 mL, 23.8 mmol) was added dropwise and the resulting biphasic mixture was vigorously stirred for 2 h at RT. The reaction mixture was transferred to a separation funnel and the organic layer was separated. The aqueous fraction was backwashed with CH_2Cl_2 (10 mL) and the two organic fractions were pooled.
To a solution of H-Lys(Boc)-OH (**91**) (2.50 g, 10.15 mmol, 1.0 eq.) and K_2CO_3 (2.4 g, 17.0 mmol) in H_2O/MeOH (100 mL, 1:1, v/v) was added a solution of $CuSO_4$ (150 mg) in H_2O (1 mL). To the resulting dark blue mixture, the freshly prepared TfN_3 solution and water (20 mL) were added. The reaction mixture was homogenized by the addition of a minimal amount of MeOH and vigorously stirred at RT for 24 h. After this time, the resulting dark green solution was partly concentrated *in vacuo* to remove all CH_2Cl_2, MeOH and part of the water. The aqueous residue was redissolved in water (150 mL) and the pH was adjusted to 2.5-3 by the addition of HCl (1.0M). The acidic solution was extracted with EtOAc (4 x 200 mL) and the organic fractions were pooled, dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by silica gel chromatography (MeOH/CH_2Cl_2 0-10%, v/v with 0.1% HOAc), and coevaporated with dichloroethane and chloroform to yield the title compound (2.18 g, 8.00 mmol, 79%) as a colorless oil. ^{1}H NMR (300 MHz, CD_3OD) δ = 3.94 (dd, 1H, *J* = 8.1 Hz, *J* = 4.8 Hz), 3.04 (t, 2H, *J* = 6.6 Hz), 1.91-1.78 (m, 1H), 1.78-1.62 (m, 1H), 1.55-1.39 ppm (m, 13H); ^{13}C NMR (75 MHz, CD_3OD) δ = 173.8, 158.5, 79.9, 63.1, 41.0, 32.1, 30.4, 28.8, 24.0 ppm; HRMS: calcd. for $C_{11}H_{19}N_4O_4$ $[M-H]^-$: 271.1412, found: 271.1407.

***N*-succinimidyl 4-benzoylbenzoate (94a):** The synthetic procedure was adapted from Olivo *et al.*[21] In brief, to a solution of 4-benzoylbenzoic acid (**93a**) (1.13 g, 5.0 mmol) and DMAP (31 mg, 5 mol%) in anhydrous CH_2Cl_2 (50 mL) were added *N*-hydroxysuccinimide (NHS) (0.86 g, 7.5 mmol) and EDC.HCl (1.44 g, 7.5 mmol). The resulting reaction mixture was stirred overnight at RT in the dark under nitrogen. The mixture was washed with water (10 mL) and brine (10 mL), dried over Na_2SO_4, and concentrated *in vacuo*. The thus obtained semi pure product was purified further by precipitation (CH_2Cl_2/hexane) yielding the title compound (1.05 g, 3.25 mmol, 65%) as white crystals. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 8.26 (d, 2H, *J* = 8.6 Hz), 7.90 (d, 2H, *J* = 8.6 Hz), 7.84-7.77 (m, 2H), 7.69-7.60 (m, 1H), 7.56-7.47 (m, 2H), 2.94 ppm (br s, 4H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 195.7, 169.2, 161.4, 143.2, 136.7, 133.4, 130.7, 130.3, 130.1, 128.7, 128.2, 25.8 ppm.

***N*-succinimidyl 4-azidobenzoate (94b):** The synthetic procedure was adapted from Dcona *et al*.[22] In brief, to an ice-cooled solution of 4-azidobenzoic acid (**93b**) (500 mg, 3.065 mmol) in DMF (10 mL) were added NHS (423 mg, 3.678 mmol) and EDC.HCl (705 mg, 3.678 mmol). The resulting reaction mixture was stirred overnight in the dark under nitrogen, allowing the temperature to rise to RT. The mixture was concentrated *in vacuo* and the residue was resuspended in EtOAc (50 mL), washed with water (3 x 20 mL), dried over Na_2SO_4, and concentrated *in vacuo*. The residue was purified by silica gel chromatography (EtOAc/toluene 0-25%, v/v) yielding the title compound (674 mg, 2.59 mmol,

85%) as a white solid. ^{1}H NMR (300 MHz, $SO(CD_3)_2$) δ = 8.11 (d, 2H, *J* = 8.9 Hz), 7.37 (d, 2H, *J* = 8.9 Hz), 2.90 ppm (s, 4H); ^{13}C NMR (75 MHz, $SO(CD_3)_2$) δ = 170.3, 161.1, 146.8, 132.0, 120.5, 120.1, 25.5 ppm; HRMS: calcd. for $C_{22}H_{16}N_8O_8Na$ $[2M+Na]^+$: 543.0983, found: 543.0982.

***N*-succinimidyl 4-(3-(trifluoromethyl)-3*H*-diazirinyl) benzoate (94c):** The synthetic procedure was adapted from Pourcelle *et al*.[23] In brief, to an ice-cooled solution of 4-(3-(trifluoromethyl)-3*H*-diazirinyl) benzoic acid (**93c**) (230 mg, 1.0 mmol) and NHS (138 mg, 1.2 mmol) in DMF (2.5 mL) was added dropwise a suspension of EDC.HCl (230 mg, 1.2 mmol) and DMAP (6 mg, 5mol%) in DMF (5.5 mL). The resulting reaction mixture was stirred overnight in the dark under nitrogen, allowing the temperature to rise to RT. The mixture was concentrated *in vacuo* and the residue was resuspended in EtOAc (40 mL), washed with water (3 x 15 mL), dried over Na_2SO_4, and concentrated *in vacuo*. The residue was purified by silica gel chromatography (EtOAc/toluene 0-12%, v/v) yielding the title compound (308 mg, 0.94 mmol, 94%) as a white solid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 8.17 (d, 2H, *J* = 8.6 Hz), 7.32 (d, 2H, *J* = 8.6 Hz), 2.91 ppm (br s, 4H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 169.1, 161.1, 136.0, 131.0, 126.9, 126.4, 121.8 (q, $J_{C\text{-}F}$ = 273.3 Hz), 28.5 (q, $J_{C\text{-}F}$ = 40.9 Hz), 25.8 ppm; HRMS: calcd. for $C_{15}H_{12}F_3N_4O_4$ $[M+MeCN+H]^+$: 369.0805, found: 369.0801.

4-(3-Chloro-4-fluorophenylamino)-7-fluoro-6-nitroquinazoline (96): 7-Fluoro-6-nitroquinazolin-4(3*H*)-one (**95**) (627 mg, 3.00 mmol) was suspended in thionyl chloride (15 mL), charged with a catalytic amount of DMF (50 µL), and refluxed for 4 h. After being cooled to RT, the mixture was concentrated *in vacuo*, and coevaporated with toluene (4 × 10 mL) to remove traces of thionyl chloride. The resulting crude 4-chloro substitution product and 3-chloro-4-fluoroaniline (437 mg, 3.00 mmol) were next taken up in *i*PrOH (20 mL). Triethylamine (1.5 mL, 10.8 mmol) was added and the reaction mixture was refluxed for 4 h. After being cooled to RT, the mixture was concentrated *in vacuo*. To the residue were added EtOAc (15 mL) and sat. $NaHCO_3$ (15 mL) and the biphasic mixture was agitated until the residue was completely dissolved and transferred to a separation funnel. The aqueous layer was separated and the organic fraction was washed with sat. $NaHCO_3$ (2 × 15 mL), dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by silica gel chromatography (EtOAc/toluene 0-20%, v/v) yielding the title compound (929 mg, 2.76 mmol, 92%) as a bright orange solid. ^{1}H NMR (300 MHz, $SO(CD_3)_2$) δ = 10.50 (s, 1H), 9.57 (d, 1H, $J_{H\text{-}F}$ = 8.1 Hz), 8.73 (s, 1H), 8.13 (dd, 1H, $J_{H\text{-}F}$ = 6.8 Hz, *J* = 2.6 Hz), 7.85 (d, 1H, $J_{H\text{-}F}$ = 12.6 Hz), 7.80 (ddd, 1H, *J* = 9.0 Hz, $J_{H\text{-}F}$ = 4.2 Hz, *J* = 2.6 Hz), 7.49 ppm (dd, 1H, *J* = $J_{H\text{-}F}$ = 9.0 Hz); ^{13}C NMR (75 MHz, $SO(CD_3)_2$) δ = 158.2, 158.0, 156.2 ($J_{C\text{-}F}$ = 249.6 Hz), 153.9 ($J_{C\text{-}F}$ = 13.8 Hz), 153.8 ($J_{C\text{-}F}$ = 242.7 Hz), 135.4 ($J_{C\text{-}F}$ = 3.5 Hz), 135.2 ($J_{C\text{-}F}$ = 10.4 Hz), 124.2, 123.9, 122.6 ($J_{C\text{-}F}$ = 6.9 Hz), 118.9 ($J_{C\text{-}F}$ = 18.3 Hz), 116.5 ($J_{C\text{-}F}$ = 21.8 Hz), 115.0 ($J_{C\text{-}F}$ = 19.5 Hz), 111.0 ppm; HRMS: calcd. for $C_{14}H_8ClF_2N_4O_2$ $[M+H]^+$: 337.0298, found: 337.0288.

4-(3-Chloro-4-fluorophenylamino)-6-nitro-7-(4-pentynoxy)quinazoline (97): A flask was charged with 4-pentynol (19 mL, 204 mmol), and cooled on ice under nitrogen. To the dry material sodium hydride (60% dispersion in mineral oil, 0.75 g, 18.8 mmol, 2.2 eq.) was added, and the mixture was stirred for 45 min, turning dark yellow while the temperature was allowed to rise to RT. Compound **96** (2.87 g, 8.5 mmol, 1.0 eq.) and DMF (6 mL) were added to this suspension, and the resulting reaction mixture was stirred overnight at RT, turning dark red. The mixture was subsequently poured in EtOAc (300 mL) and washed with water (2 × 50 mL) and brine (50 mL). The combined aqueous phases were next backwashed with EtOAc (750 mL), after which all EtOAc fractions were pooled, dried over Na_2SO_4 and concentrated *in vacuo*. The oily residue was further concentrated under high

vacuum for 1 h to remove excess of 4-pentynol. The crude material was purified by silica gel chromatography (EtOAc/toluene 0-20%, v/v) yielding the title compound (2.80 g, 7.0 mmol, 82%) as a yellowish-orange solid. ^{1}H NMR (300 MHz, $SO(CD_3)_2$) δ = 9.99 (s, 1H), 9.13 (s, 1H), 8.59 (s, 1H), 8.11 (dd, 1H, J_{H-F} = 6.8 Hz, J = 2.6 Hz), 7.78 (ddd, 1H, J = 9.0 Hz, J_{H-F} = 4.5 Hz, J = 2.6 Hz), 7.40 (dd, 1H, $J = J_{H-F}$ = 9.0 Hz), 7.35 (s, 1H), 4.31 (t, 2H, J = 6.0 Hz), 2.85 (t, 1H, J = 2.7 Hz), 2.40 (dt, 2H, J = 7.2 Hz, J = 2.7 Hz), 2.04-1.93 ppm (m, 2H); ^{13}C NMR (75 MHz, $SO(CD_3)_2$) δ = 157.6, 157.2, 153.6, 153.5 (J_{C-F} = 242.8 Hz), 153.2, 138.5, 135.9 (J_{C-F} = 3.5 Hz), 123.4, 122.1 (J_{C-F} = 6.8 Hz), 121.9, 118.8 (J_{C-F} = 18.3 Hz), 116.4 (J_{C-F} = 21.8 Hz), 110.0, 107.8, 83.3, 71.7, 68.1, 27.4, 14.4 ppm; HRMS: calcd. for $C_{19}H_{15}ClFN_4O_3$ $[M+H]^+$: 401.0811, found: 401.0811.

6-Amino-4-(3-chloro-4-fluorophenylamino)-7-(4-pentynoxy)quinazoline (98): Compound **97** (2.39 g, 6.0 mmol) and $SnCl_2$ (11.3 g, 60 mmol) were taken up in EtOH (90 mL). The resulting reaction mixture was refluxed for 5 h under vigorous stirring. After being cooled to RT, the pH was made slightly basic (pH 7-8) by slow addition of sat. Na_2CO_3. The alkaline solution was transferred to a separation funnel and the aqueous fraction was extracted with EtOAc (3 × 300 mL). All EtOAc fractions were pooled, dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by silica gel chromatography ($MeOH/CH_2Cl_2$ 0-4%, v/v with 0.1% NH_4OH) yielding the title compound (1.64 g, 4.4 mmol, 73%) as a light greenish-brown solid. ^{1}H NMR (300 MHz, $SO(CD_3)_2$) δ = 9.39 (s, 1H), 8.37 (s, 1H), 8.19 (dd, 1H, J_{H-F} = 6.9 Hz, J = 2.7 Hz), 7.81 (ddd, 1H, J = 9.0 Hz, J_{H-F} = 4.3 Hz, J = 2.7 Hz), 7.39 (dd, 1H, $J = J_{H-F}$ = 9.0 Hz), 7.38 (s, 1H), 7.08 (s, 1H), 5.41 (s, 2H), 4.22 (t, 2H, J = 5.9 Hz), 2.84 (t, 1H, J = 2.7 Hz), 2.47 (dt, 2H, J = 7.1 Hz, J = 2.7 Hz), 2.06-1.95 ppm (m, 2H); ^{13}C NMR (75 MHz, $SO(CD_3)_2$) δ = 155.0, 152.6 (J_{C-F} = 240.5 Hz), 151.9, 150.2, 144.7, 138.6, 137.5 (J_{C-F} = 2.3 Hz), 122.4, 121.4 (J_{C-F} = 6.9 Hz), 118.6 (J_{C-F} = 17.2 Hz), 116.4 (J_{C-F} = 20.6 Hz), 110.4, 106.3, 100.8, 84.0, 71.6, 66.7, 27.5, 14.7 ppm; HRMS: calcd. for $C_{19}H_{17}ClFN_4O$ $[M+H]^+$: 371.1069, found: 371.1071.

Methyl 4-(dimethylamino)crotonate (100a): Dimethylamine (62.8 mL, 2.0M in THF, 125.6 mmol, 3 eq.) was added dropwise to a solution of methyl 4-bromocrotonate (**99a**) (5.0 mL, 41.8 mmol) in THF (60 mL) at 0°C under nitrogen in the dark. The reaction mixture was stirred for 4 h, allowing the temperature to rise to RT, and subsequently concentrated *in vacuo*. The residue was dissolved in CH_2Cl_2 (40 mL) and washed with sat. $NaHCO_3$ (10 mL). The aqueous phase was next backwashed with CH_2Cl_2 (40 mL), after which all CH_2Cl_2 fractions were pooled, dried over Na_2SO_4, and concentrated *in vacuo* to yield the title compound (5.42 g, 37.9 mmol, 91%) as a yellowish-orange oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 6.95 (dt, 1H, J = 15.7 Hz, J = 6.2 Hz), 5.98 (dt, 1H, J = 15.7 Hz, J = 1.6 Hz), 3.74 (s, 3H), 3.07 (dd, 2H, J = 6.2 Hz, J = 1.6 Hz), 2.25 ppm (s, 6H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 166.6, 145.9, 122.6, 60.4, 51.5, 45.4 ppm; HRMS: calcd. for $C_7H_{14}NO_2$ $[M+H]^+$: 144.1019, found: 144.1016.

Ethyl 4-(dimethylamino)butyrate (100b): Dimethylamine (31.5 mL, 2.0M in THF, 63 mmol, 3 eq.) was added dropwise to a solution of ethyl 4-bromobutyrate (**99b**) (3.0 mL, 21.0 mmol) in THF (25 mL) at 0°C under nitrogen. The reaction mixture was stirred for 2 h, allowing the temperature to rise to RT, and subsequently concentrated *in vacuo*. The residue was dissolved in EtOAc (40 mL) and washed with sat. $NaHCO_3$ (10 mL). The aqueous phase was next backwashed with EtOAc (40 mL), after which all EtOAc fractions were pooled, dried over Na_2SO_4, and concentrated *in vacuo*. The residue was purified by silica gel chromatography ($EtOH/CH_2Cl_2$ 0-20%, v/v with 0.1% NH_4OH) yielding the title compound (688 mg, 4.32 mmol, 21%) as an off white oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 4.13 (q, 2H, J = 7.1 Hz), 2.52 (t, 2H, J = 7.7 Hz), 2.39 (s, 6H), 2.38 (t, 2H, J = 7.4 Hz), 1.95-1.84 (m, 2H), 1.26 ppm (t, 3H,

J = 7.1 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 173.1, 60.4, 58.3, 44.7, 31.7, 22.1, 14.2 ppm; HRMS: calcd. for $C_8H_{18}NO_2$ $[M+H]^+$: 160.1332, found: 160.1329.

4-(Dimethylamino)crotonic acid hydrochloride (101a): A solution of sodium hydroxide (856 mg, 21.4 mmol) in water (6.3 mL) was added to a solution of methyl ester **100a** (3.07 g, 21.4 mmol) in MeOH (25 mL). The resulting reaction mixture was heated to 50°C for 1 h. After being cooled to RT, the solution was acidified with hydrochloric acid (5.0M) to pH 1-2 and concentrated *in vacuo* to a thick oil. Anhydrous ethanol (25 mL) was added and the white precipitate of sodium chloride was removed by filtration. The filtrate was concentrated *in vacuo* and product precipitation was induced by the addition of *i*PrOH (13 mL) and overnight storing at -20°C. The resulting precipitate was filtered and dried under high vacuum to afford the title compound (1.12 g, 6.8 mmol, 32%) as an off white solid. ^{1}H NMR (300 MHz, D_2O) δ = 6.97-6.84 (m, 1H), 6.31 (d, 1H, *J* = 15.6 Hz), 3.99 (d, 2H, *J* = 7.2 Hz), 2.94 ppm (s, 6H); HRMS: calcd. for $C_6H_{12}NO_2$ $[M+H]^+$: 130.0863, found: 130.0857.

4-(Dimethylamino)butyric acid hydrochloride (101b): A solution of ethyl ester **100b** (627 mg, 3.94 mmol) in MeOH (5 mL) was treated with sodium hydroxide (4 mL, 4.0M). The resulting reaction mixture was heated to 50°C for 6 h. After being cooled to RT, the solution was acidified with hydrochloric acid (4.0M) to pH 1-2 and concentrated *in vacuo*. The residue was coevaporated with toluene (3 × 10 mL) and dried overnight under high vacuum to afford the title compound (3.94 mmol, considered quant.) as a white solid. ^{1}H NMR (300 MHz, D_2O) δ = 3.23 (t, 2H, *J* = 8.3 Hz), 2.94 (s, 6H), 2.55 (t, 2H, *J* = 7.2 Hz), 2.11-2.00 ppm (m, 2H); HRMS: calcd. for $C_6H_{14}NO_2$ $[M+H]^+$: 132.1019, found: 132.1023.

Alkyne-functionalized afatinib (102a): 2-Chloro-1-methylpyridinium iodide (Mukaiyama's reagent) (618 mg, 2.42 mmol) and Et_3N (642 µL, 4.61 mmol) were added to a solution of acid **101a** (381 mg, 2.30 mmol) in CH_2Cl_2 (3.5 mL). The resulting preactivation mixture was refluxed for 25 min. Subsequently, a solution of aniline **98** (142 mg, 0.383 mmol) in DMF (2 mL) was added to this mixture and stirring was continued for 50 h at 60°C. After being cooled to RT, the reaction mixture was concentrated *in vacuo*. To the residue were added EtOAc (10 mL) and sat. $NaHCO_3$ (10 mL) and the biphasic mixture was agitated until the residue was completely dissolved and transferred to a separation funnel. The aqueous layer was separated and next backwashed with EtOAc (20 mL), after which all EtOAc fractions were pooled, dried over Na_2SO_4, and concentrated *in vacuo*. The residue was purified by preparative RP-HPLC (15-30% MeCN) yielding the title compound (55.2 mg, 115 µmol, 30%) as a pale yellow amorphous solid. ^{1}H NMR (300 MHz, $SO(CD_3)_2$) δ = 9.80 (s, 1H), 9.49 (s, 1H), 8.91 (s,1H), 8.53 (s, 1H), 8.14 (dd, 1H, $J_{H\text{-}F}$ = 6.9 Hz, *J* = 2.4 Hz), 7.81 (ddd, 1H, *J* = 9.0 Hz, $J_{H\text{-}F}$ = 4.3 Hz, *J* = 2.4 Hz), 7.42 (dd, 1H, *J* = $J_{H\text{-}F}$ = 9.0 Hz), 7.28 (s, 1H), 6.81 (dt, 1H, *J* = 15.4 Hz, *J* = 5.9 Hz), 6.57 (d, 1H, *J* = 15.4 Hz), 4.28 (t, 2H, *J* = 6.2 Hz), 3.10 (d, 2H, *J* = 5.9 Hz), 2.84 (t, 1H, *J* = 2.6 Hz), 2.45 (dt, 2H, *J* = 7.1 Hz, *J* = 2.6 Hz), 2.19 (s, 6H), 2.08-1.97 ppm (m, 2H); LC-HRMS: t_R = 5.64 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{25}H_{26}ClFN_5O_2$ $[M+H]^+$: 482.1754, found: 482.1748.

Alkyne-functionalized dihydro-afatinib (102b): 2-Chloro-1-methylpyridinium iodide (Mukaiyama's reagent) (368 mg, 1.44 mmol) and Et_3N (482 µL, 3.46 mmol) were added to a solution of acid **101b** (290 mg, 1.73 mmol) in CH_2Cl_2 (3.0 mL). The resulting preactivation mixture was refluxed for 45 min. Subsequently, a solution of aniline **98** (142 mg, 0.383 mmol) in DMF (1.5 mL) was added to this mixture and stirring was continued for 19 h at 60°C. After being cooled to RT, the reaction mixture

was concentrated *in vacuo*. To the residue were added EtOAc (10 mL) and sat. $NaHCO_3$ (10 mL) and the biphasic mixture was agitated until the residue was completely dissolved and transferred to a separation funnel. The aqueous layer was separated and next backwashed with EtOAc (20 mL), after which all EtOAc fractions were pooled, dried over Na_2SO_4, and concentrated *in vacuo*. The residue was purified by silica gel chromatography (MeOH/CH_2Cl_2 0-6%, v/v with 0.1% NH_4OH) yielding the title compound (97.8 mg, 202 µmol, 53%) as a white amorphous solid. ^{1}H NMR (300 MHz, $CDCl_3$) δ = 9.02 (s, 1H), 8.63 (s, 1H), 8.47 (s, 1H), 7.89 (dd, 1H, J_{H-F} = 6.6 Hz, J = 2.7 Hz), 7.78 (s, 1H), 7.51 (ddd, 1H, J = 9.0 Hz, J_{H-F} = 4.2 Hz, J = 2.7 Hz), 7.24 (s, 1H), 7.11 (dd, 1H, J = J_{H-F} = 9.0 Hz), 4.32 (t, 2H, J = 6.2 Hz), 2.58 (t, 2H, J = 7.2 Hz), 2.51-2.42 (m, 4H), 2.30 (s, 6H), 2.21-2.11 (m, 2H), 2.11 (t, 1H, J = 2.7 Hz), 2.01-1.90 ppm (m, 2H); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 172.0, 156.8, 154.8 (J_{C-F} = 245.0 Hz), 154.5, 152.1, 148.4, 135.3 (J_{C-F} = 3.5 Hz), 128.1, 124.2, 121.8 (J_{C-F} = 6.9 Hz), 121.0 (J_{C-F} = 19.5 Hz), 116.5 (J_{C-F} = 21.8 Hz), 109.7, 109.4, 107.6, 83.0, 69.9, 68.1, 58.2, 45.3, 35.5, 27.6, 22.9, 15.6 ppm; LC-HRMS: t_R = 5.42 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{25}H_{29}ClFN_5O_2$ $[M+2H]^{2+}$: 242.5991, found: 242.5984.

Afatinib TFC (103a): The second-generation TMP-N_3 reagent **40** (21.6 mg, 37.5 µmol, 1.5 eq.) was taken up in DMF (760 µL), and alkyne **102a** (12.1 mg, 25 µmol, 1.0 eq.), $CuSO_4$ (100 µL, 0.05M, 0.2 eq.), Na ascorbate (50 µL, 0.5M, 1.0 eq.), and Et_3N (10.5 µL, 75 µmol, 3.0 eq.) were added. The resulting mixture was charged with a catalytic amount of TBTA[50] and vigorously stirred at RT for 16 h. The solution was concentrated *in vacuo* and the residue was purified by preparative RP-HPLC (10-100% MeCN) yielding the title compound (15.1 mg, 14 µmol, 56%) as an off white amorphous solid. LC-HRMS: t_R = 5.23 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{51}H_{68}ClFN_{13}O_9$ $[M+3H]^{3+}$: 353.4973, found: 353.4913.

Dihydro-afatinib TFC (103b): The second-generation TMP-N_3 reagent **40** (42.4 mg, 73.5 µmol, 1.5 eq.) was taken up in DMF (1.5 mL), and alkyne **102b** (23.5 mg, 49 µmol, 1.0 eq.), $CuSO_4$ (196 µL, 0.05M, 0.2 eq.), Na ascorbate (98 µL, 0.5M, 1.0 eq.), and Et_3N (20.5 µL, 147 µmol, 3.0 eq.) were added. The resulting mixture was charged with a catalytic amount of TBTA[50] and vigorously stirred at RT for 16 h. The solution was concentrated *in vacuo* and the residue was purified by preparative RP-HPLC (10-100% MeCN) yielding the title compound (35.8 mg, 34 µmol, 69%) as an off white amorphous solid. LC-HRMS: t_R = 5.24 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{51}H_{70}ClFN_{13}O_9$ $[M+3H]^{3+}$: 354.1692, found: 354.1663.

4-(3-Chloro-4-fluorophenylamino)-7-hydroxy-6-(3-morpholinopropoxy)quinazoline (104): Gefitinib (1.00 g, 2.24 mmol) was added batchwise to neat liquid pyridinium chloride (3.88 g, 33.6 mmol) at 170°C over 10 min and the resulting reaction mixture was gently stirred at 170°C for 3 h. After being cooled to RT, the resulting solid was taken up in water (30 mL) and the pH was increased to 7 using ammonium hydroxide. The neutral solution was concentrated *in vacuo* and the residue was purified by silica gel chromatography (MeOH/CH_2Cl_2 0-20%, v/v) yielding the title compound (742 mg, 1.71 mmol, 76%) as a brown foam. ^{1}H NMR (300 MHz, $SO(CD_3)_2$) δ = 10.42 (br s, 1H), 8.55 (s, 1H), 8.28 (s, 1H), 8.23 (d, 1H, J_{H-F} = 5.1 Hz), 7.98-7.86 (m, 1H), 7.45 (dd, 1H, J = J_{H-F} = 8.7 Hz), 7.22 (s, 1H), 4.41-4.27 (m, 2H), 3.96-3.84 (m, 4H), 3.40-3.13 (m, 6H), 2.36-2.21 ppm (m, 2H); HRMS: calcd. for $C_{21}H_{23}ClFN_4O_3$ $[M+H]^+$: 433.1437, found: 433.1423.

Alkyne-functionalized gefitinib (105): A solution of di-*tert*-butyl azodicarboxylate (DBAD) (213 mg, 924 µmol, 4.0 eq.) in THF (1.0 mL) was added dropwise to a solution of phenol **104** (100 mg, 231 µmol, 1.0 eq.), 4-pentynol (54 µL, 578 µmol, 2.5 eq.), and triphenylphosphine (242 mg, 924 µmol, 4.0 eq.) in THF (1.5 mL) at 0°C over 5 min. The resulting reaction mixture was allowed to warm to RT and subsequently heated to 70°C for 21 h. After being cooled to RT, the solution was concentrated *in vacuo* and the residue was purified by silica gel chromatography ($MeOH/CH_2Cl_2$ 0-5%, v/v with 0.1% NH_4OH) yielding the title compound (65.3 mg, 131 µmol, 57%) as a white amorphous solid. ^1H NMR (300 MHz, $CDCl_3$) δ = 8.65 (s, 1H), 7.87 (dd, 1H, J_{H-F} = 6.6 Hz, J = 2.7 Hz), 7.54 (ddd, 1H, J = 8.9 Hz, J_{H-F} = 4.1 Hz, J = 2.7 Hz), 7.40 (s, 1H), 7.25 (s, 1H), 7.15 (dd, 1H, $J = J_{H-F}$ = 8.9 Hz), 7.11 (s, 1H), 4.23 (t, 2H, J = 6.2 Hz), 4.14 (t, 2H, J = 6.5 Hz), 3.74 (t, 4H, J = 4.7 Hz), 2.57 (t, 2H, J = 7.1 Hz), 2.49 (t, 4H, 4.7 Hz), 2.46 (dt, 2H, J = 6.9 Hz, J = 2.7 Hz), 2.16-2.02 (m, 4H), 1.99 ppm (t, 1H, J = 2.7 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ = 156.2, 154.9 (J_{C-F} = 245.0 Hz), 154.8, 153.5, 149.5, 147.7, 135.4 (J_{C-F} = 3.5 Hz), 124.2, 121.8 (J_{C-F} = 6.8 Hz), 121.2 (J_{C-F} = 18.4 Hz), 116.7 (J_{C-F} = 21.8 Hz), 108.99, 108.91, 100.9, 83.1, 69.4, 67.7, 67.2, 67.1, 55.6, 53.9, 27.9, 26.3, 15.3 ppm; HRMS: calcd. for $C_{26}H_{29}ClFN_4O_3$ $[M+H]^+$: 499.1907, found: 499.1912.

Gefitinib TFC (106): The second-generation TMP-N_3 reagent **40** (43.2 mg, 75 µmol, 1.5 eq.) was taken up in DMF (1.5 mL), and alkyne **105** (24.9 mg, 50 µmol, 1.0 eq.), $CuSO_4$ (200 µL, 0.05M, 0.2 eq.), Na ascorbate (100 µL, 0.5M, 1.0 eq.), and Et_3N (20.9 µL, 150 µmol, 3.0 eq.) were added. The resulting mixture was charged with a catalytic amount of TBTA[50] and vigorously stirred at RT for 16 h. The solution was concentrated *in vacuo* and the residue was purified by preparative RP-HPLC (10-100% MeCN) yielding the title compound (40.2 mg, 37 µmol, 74%) as an off white amorphous solid. LC-HRMS: t_R = 5.29 min (10-100% MeCN, 15 min run); HRMS: calcd. for $C_{52}H_{71}ClFN_{12}O_{10}$ $[M+3H]^{3+}$: 359.1691, found: 359.1676.

V.7.2. Molecular biology

Performed by the Cytokine Receptor Laboratory, Department of Medical Protein Research, VIB, Ghent & Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University (Prof. Jan Tavernier, Dr. Sam Lievens).

Plasmid constructs: The pCLL-eDHFR and pCLG-eDHFR plasmids expressing a fusion of the partial leptin receptor fused to the *E. coli* dihydrofolate reductase enzyme have been described previously,[51] as have the pMG1C-ER1,[52] and pMG1-RNF41[53] prey plasmids. Other prey constructs encoding C-terminal fusions of a protein (domain) of interest with a gp130 receptor fragment containing functional STAT3 recruitment sites were generated by Gateway mediated transfer of the corresponding entry clones in the Gateway-compatible pMG1 prey destination vector as reported earlier.[54] Gateway entry vectors harboring partial sequences corresponding to human EGFR were produced by PCR amplifying the region C-terminal of AA669 (for the cytoplasmic domain; primers used were 5'-GGGGACAACTTTGTACAAAAAAGTTGGCACCATGCGAAGGCGCCACATCGTTCG-3' and 5'-GGGGACAACTTTGTACAAGAAAGTTGGCAATGCTCCAATAAATTCACTGC-3') or covering AA696-AA1022 (for the kinase domain; primers used were 5'-GGGGACAACTTTGTACAAAAAAGTTGGCACCATGGGAGA-AGCTCCCAACCAAGC-3' and 5'-GGGGACAACTTTGTACAAGAAAGTTGGCAAGCCCTGCTGTGGGATGAGG-TAC-3') and Gateway BP recombination into the pDONR223 vector, as reported.[55] All other prey constructs were derived from full size human ORF entry clones available in the hORFeome v8.1 collection.[55] Also the STAT3-dependent firefly luciferase reporter pXP2d2-rPAPI-luciferase has been described before.[56]

MASPIT assays: For standard MASPIT assays, HEK293T cells were cultured in 96-well microtiter plates in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum, incubated at 37°C, under 8% CO_2, and transfected with pCLL-eDHFR or pCLG-eDHFR fusion protein plasmid (10 ng per well), either of the prey plasmids (100 ng per well) and pXP2d2-rPAPI-luciferase (5 ng per well) by applying a standard calcium phosphate transfection method, as described earlier.[54] Twenty-four hours after transfection, cells were stimulated with leptin (100 ng/mL final concentration) alone or in combination with the indicated concentration of fusion compound. After another 24 h, luciferase activity was measured by using the Luciferase Assay System kit (Promega) on an Envision plate reader (PerkinElmer). Luciferase data represent the average of three technical replicates; error bars indicate standard deviation.

A similar procedure was followed in the MASPIT UV irradiation experiment, except for the following modifications: twenty-four hours after transfection, media was changed for serum-supplemented DMEM lacking phenol red and cells were treated with the fusion compound in the absence of leptin; another 24 h later, cells were irradiated with UV at 245nm or 365nm for the indicated time (Benda NU-8 KL handheld UV lamp; 2x8W), after which media was changed for serum-supplemented DMEM containing phenol red and additionally supplemented with leptin (100 ng/mL final concentration); luciferase activity was measured another 24 h later.

In MASPIT competition assays, cells were transfected according to the protocol described above, with a combination of the pCLG-eDHFR fusion protein plasmid (10 ng per well), either pMG1-RNF41 or pMG1C-ER1 (10 ng per well), and the luciferase reporter plasmid (5 ng per well). Twenty-four hours after transfection, cells were treated with a combination of either leptin alone (100 ng/mL final concentration) or leptin combined with TMP-tamoxifen (0,1 µM final concentration) and the

indicated concentration of the competing molecule (tamoxifen, trimethoprim, simvastatin, TMP-simvastatin, TMP-afatinib, TMP-dihydro-afatinib or TMP-gefitinib). Luciferase activity was measured another 24h later, as described above.

ERK phosphorylation assay: HEK293T cells were seeded at 200.000 cells/well in 6-well plates in DMEM supplemented with 10% fetal calf serum and grown as in the case of the MASPIT assays. Forty-eight hours after seeding, medium was replaced by DMEM supplemented with 1% fetal calf serum and cells were treated with the indicated small molecule concentration, all diluted from 10mM 100% DMSO stocks. For each of the applied concentrations, the corresponding amount of pure DMSO was taken along as control. Sixteen hours later, cells were stimulated with human EGF (50 ng/mL final concentration) for 5 minutes. Next, cells were rinsed with PBS and lysed in Laemmli lysis buffer at 4°C. Lysates were cleared by centrifugation, heated for 5 min at 95°C and subjected to SDS/PAGE on a 12% gel. After blotting, total and phosphorylated ERK1/2 were revealed using mouse anti-ERK1/2 (Cell Signaling Technologies #4696) and rabbit anti-phospho-ERK1/2 (Cell Signaling Technologies #9101) antibodies combined with appropriately infrared dye-labeled secondary antibodies on a LiCor Odyssey infrared imager.

References

[1] Y. Tanaka, M. R. Bond, J. J. Kohler, *Mol. Biosyst.* **2008**, *4*, 473-480.
[2] A. Singh, E. R. Thornton, F. H. Westheimer, *J. Biol. Chem.* **1962**, *237*, 3006-3008.
[3] J. R. Knowles, *Accounts Chem. Res.* **1972**, *5*, 155-160.
[4] J. Brunner, *Annu. Rev. Biochem.* **1993**, *62*, 483-514.
[5] P. P. Geurink, L. M. Prely, G. A. van der Marel, R. Bischoff, H. S. Overkleeft, *Top. Curr. Chem.* **2012**, *324*, 85-113.
[6] A. Wittelsberger, B. E. Thomas, D. F. Mierke, M. Rosenblatt, *FEBS Lett.* **2006**, *580*, 1872-1876.
[7] G. Dorman, G. D. Prestwich, *Biochemistry* **1994**, *33*, 5661-5673.
[8] E. L. Vodovozova, *Biochem.-Moscow* **2007**, *72*, 1-20.
[9] F. Kotzybahibert, I. Kapfer, M. Goeldner, *Angew. Chem.-Int. Edit.* **1995**, *34*, 1296-1312.
[10] R. J. Bergeron, J. B. Dionis, M. J. Ingeno, *J. Org. Chem.* **1987**, *52*, 144-149.
[11] J. V. Staros, H. Bayley, D. N. Standring, J. R. Knowles, *Biochem. Biophys. Res. Commun.* **1978**, *80*, 568-572.
[12] M. W. Geiger, M. M. Elliot, V. D. Karacostas, T. J. Moricone, J. B. Salmon, V. L. Sideli, M. A. Stonge, *Photochem. Photobiol.* **1984**, *40*, 545-548.
[13] S. A. Fleming, *Tetrahedron* **1995**, *51*, 12479-12520.
[14] K. G. Pinney, J. A. Katzenellenbogen, *J. Org. Chem.* **1991**, *56*, 3125-3133.
[15] R. A. G. Smith, J. R. Knowles, *J. Chem. Soc., Perkin Trans. 2* **1975**, 686-694.
[16] J. Brunner, H. Senn, F. M. Richards, *J. Biol. Chem.* **1980**, *255*, 3313-3318.
[17] H. Bayley, in *Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 12* (Eds.: W. T.S, B. R.H), Elsevier, **1983**, pp. 8-24.
[18] M. Platz, A. S. Admasu, S. Kwiatkowski, P. J. Crocker, N. Imai, D. S. Watt, *Bioconjugate Chem.* **1991**, *2*, 337-341.
[19] H. Bayley, in *Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 12* (Eds.: W. T.S, B. R.H), Elsevier, **1983**, pp. 25-65.
[20] J. T. Lundquist, J. C. Pelletier, *Org. Lett.* **2001**, *3*, 781-783.
[21] H. F. Olivo, N. Perez-Hernandez, D. M. Liu, M. Iruthayanathan, B. O'Leary, L. L. Homan, J. S. Dillon, *Bioorg. Med. Chem. Lett.* **2010**, *20*, 1153-1155.
[22] M. M. Dcona, D. Mitra, R. W. Goehe, D. A. Gewirtz, D. A. Lebman, M. C. T. Hartman, *Chem. Commun.* **2012**, *48*, 4755-4757.
[23] V. Pourcelle, C. S. Le Duff, H. Freichels, C. Jerome, J. Marchand-Brynaert, *J. Fluor. Chem.* **2012**, *140*, 62-69.
[24] S. Lievens, D. J. H. De Clercq, S. Van Calenbergh, J. Tavernier, in preparation.
[25] J. Sumranjit, S. J. Chung, *Molecules* **2013**, *18*, 10425-10451.
[26] S. S. Gallagher, J. E. Sable, M. P. Sheetz, V. W. Cornish, *ACS Chem. Biol.* **2009**, *4*, 547-556.
[27] F. Solca, G. Dahl, A. Zoephel, G. Bader, M. Sanderson, C. Klein, O. Kraemer, F. Himmelsbach, E. Haaksma, G. R. Adolf, *J. Pharmacol. Exp. Ther.* **2012**, *343*, 342-350.
[28] A. J. Barker, K. H. Gibson, W. Grundy, A. A. Godfrey, J. J. Barlow, M. P. Healy, J. R. Woodburn, S. E. Ashton, B. J. Curry, L. Scarlett, L. Henthorn, L. Richards, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1911-1914.
[29] Y. Yarden, M. X. Sliwkowski, *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 127-137.
[30] A. Citri, K. B. Skaria, Y. Yarden, *Exp. Cell Res.* **2003**, *284*, 54-65.
[31] N. E. Hynes, H. A. Lane, *Nat. Rev. Cancer* **2005**, *5*, 341-354.
[32] H. T. Zhang, A. Berezov, Q. Wang, G. Zhang, J. Drebin, R. Murali, M. I. Greene, *J. Clin. Invest.* **2007**, *117*, 2051-2058.
[33] M. G. Kris, R. B. Natale, R. S. Herbst, T. J. Lynch, D. Prager, C. P. Belani, J. H. Schiller, K. Kelly, H. Spiridonidis, A. Sandler, K. S. Albain, D. Cella, M. K. Wolf, S. D. Averbuch, J. J. Ochs, A. C. Kay, *J. Am. Med. Assoc.* **2003**, *290*, 2149-2158.
[34] S. Kobayashi, T. J. Boggon, T. Dayaram, P. A. Janne, O. Kocher, M. Meyerson, B. E. Johnson, M. J. Eck, D. G. Tenen, B. Halmos, *N. Engl. J. Med.* **2005**, *352*, 786-792.
[35] W. Pao, V. A. Miller, K. A. Politi, G. J. Riely, R. Somwar, M. F. Zakowski, M. G. Kris, H. Varmus, *PLos Med.* **2005**, *2*, 225-235.
[36] C. H. Yun, K. E. Mengwasser, A. V. Toms, M. S. Woo, H. Greulich, K. K. Wong, M. Meyerson, M. J. Eck, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 2070-2075.

[37] E. L. Kwak, R. Sordella, D. W. Bell, N. Godin-Heymann, R. A. Okimoto, B. W. Brannigan, P. L. Harris, D. R. Driscoll, P. Fidias, T. J. Lynch, S. K. Rabindran, J. P. McGinnis, A. Wissner, S. V. Sharma, K. J. Isselbacher, J. Settleman, D. A. Haber, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7665-7670.
[38] H. S. Wiley, *Exp. Cell Res.* **2003**, *284*, 78-88.
[39] R. Roskoski, *Pharmacol. Res.* **2014**, *79*, 34-74.
[40] C. H. Yun, T. J. Boggon, Y. Q. Li, M. S. Woo, H. Greulich, M. Meyerson, M. J. Eck, *Cancer Cell* **2007**, *11*, 217-227.
[41] D. Kobus, Y. Giesen, R. Ullrich, H. Backes, B. Neumaier, *Appl. Radiat. Isot.* **2009**, *67*, 1977-1984.
[42] Z. B. Hill, B. G. K. Perera, S. S. Andrews, D. J. Maly, *ACS Chem. Biol.* **2012**, *7*, 487-495.
[43] F. D. Bellamy, K. Ou, *Tetrahedron Lett.* **1984**, *25*, 839-842.
[44] H. Takami, H. Koshimura, N. Kishibayashi, A. Ishii, H. Nonaka, S. Aoyama, H. Kase, T. Kumazawa, *J. Med. Chem.* **1996**, *39*, 5047-5052.
[45] L. F. A. Hennequin, C. T. Halsall, WO 2005/026150 A1, 24.03.2005.
[46] M. I. Davis, J. P. Hunt, S. Herrgard, P. Ciceri, L. M. Wodicka, G. Pallares, M. Hocker, D. K. Treiber, P. P. Zarrinkar, *Nat. Biotechnol.* **2011**, *29*, 1046-1051.
[47] A. Walther, E. Johnstone, C. Swanton, R. Midgley, I. Tomlinson, D. Kerr, *Nat. Rev. Cancer* **2009**, *9*, 489-499.
[48] P. Seshacharyulu, M. P. Ponnusamy, D. Haridas, M. Jain, A. K. Ganti, S. K. Batra, *Expert Opin. Ther. Targets* **2012**, *16*, 15-31.
[49] X. W. Zhang, J. Gureasko, K. Shen, P. A. Cole, J. Kuriyan, *Cell* **2006**, *125*, 1137-1149.
[50] T. R. Chan, R. Hilgraf, K. B. Sharpless, V. V. Fokin, *Org. Lett.* **2004**, *6*, 2853-2855.
[51] M. D. P. Risseeuw, D. J. H. De Clercq, S. Lievens, U. Hillaert, D. Sinnaeve, F. Van den Broeck, J. C. Martins, J. Tavernier, S. Van Calenbergh, *ChemMedChem* **2013**, *8*, 521-526.
[52] D. J. H. De Clercq, M. D. P. Risseeuw, I. Karalic, A. S. De Smet, D. Defever, J. Tavernier, S. Lievens, S. Van Calenbergh, *ChemBioChem* **2015**, *16*, 834-843.
[53] J. Wauman, L. De Ceuninck, N. Vanderroost, S. Lievens, J. Tavernier, *J. Cell Sci.* **2011**, *124*, 921-932.
[54] S. Lievens, N. Vanderroost, J. Van der Heyden, V. Gesellchen, M. Vidal, J. Tavernier, *J. Proteome Res.* **2009**, *8*, 877-886.
[55] X. Yang, J. S. Boehm, X. Yang, K. Salehi-Ashtiani, T. Hao, Y. Shen, R. Lubonja, S. R. Thomas, O. Alkan, T. Bhimdi, T. M. Green, C. M. Johannessen, S. J. Silver, C. Nguyen, R. R. Murray, H. Hieronymus, D. Balcha, C. Fan, C. Lin, L. Ghamsari, M. Vidal, W. C. Hahn, D. E. Hill, D. E. Root, *Nat. Meth.* **2011**, *8*, 659-661.
[56] S. Eyckerman, A. Verhee, J. V. der Heyden, I. Lemmens, X. V. Ostade, J. Vandekerckhove, J. Tavernier, *Nat. Cell Biol.* **2001**, *3*, 1114-1119.

CHAPTER VI

GENERAL CONCLUSION

VI. General Conclusion

As a starting point, we developed a scalable synthesis of a versatile MTX reagent comprising a tetra-, hexa-, or octa(ethylene glycol) linker and an azide ligation handle (**6a-c**) that enabled rapid γ-selective generation of MFCs from alkyne-functionalized small-molecule baits (such as tamoxifen, reversine, FK506 and simvastatin) using CuAAC (Figure VI.1, panel A). This modular strategy allowed swift access to various MFCs, thereby minimizing the number of chemical manipulations for each construct. In analytical mode, MASPIT was able to give concentration-dependent reporter signals for the established target proteins of these model baits. Furthermore, we demonstrated that the sensitivity obtained with the newly developed MTX reagent was significantly stronger than that of a previously used non-regiomeric conjugate mixture.

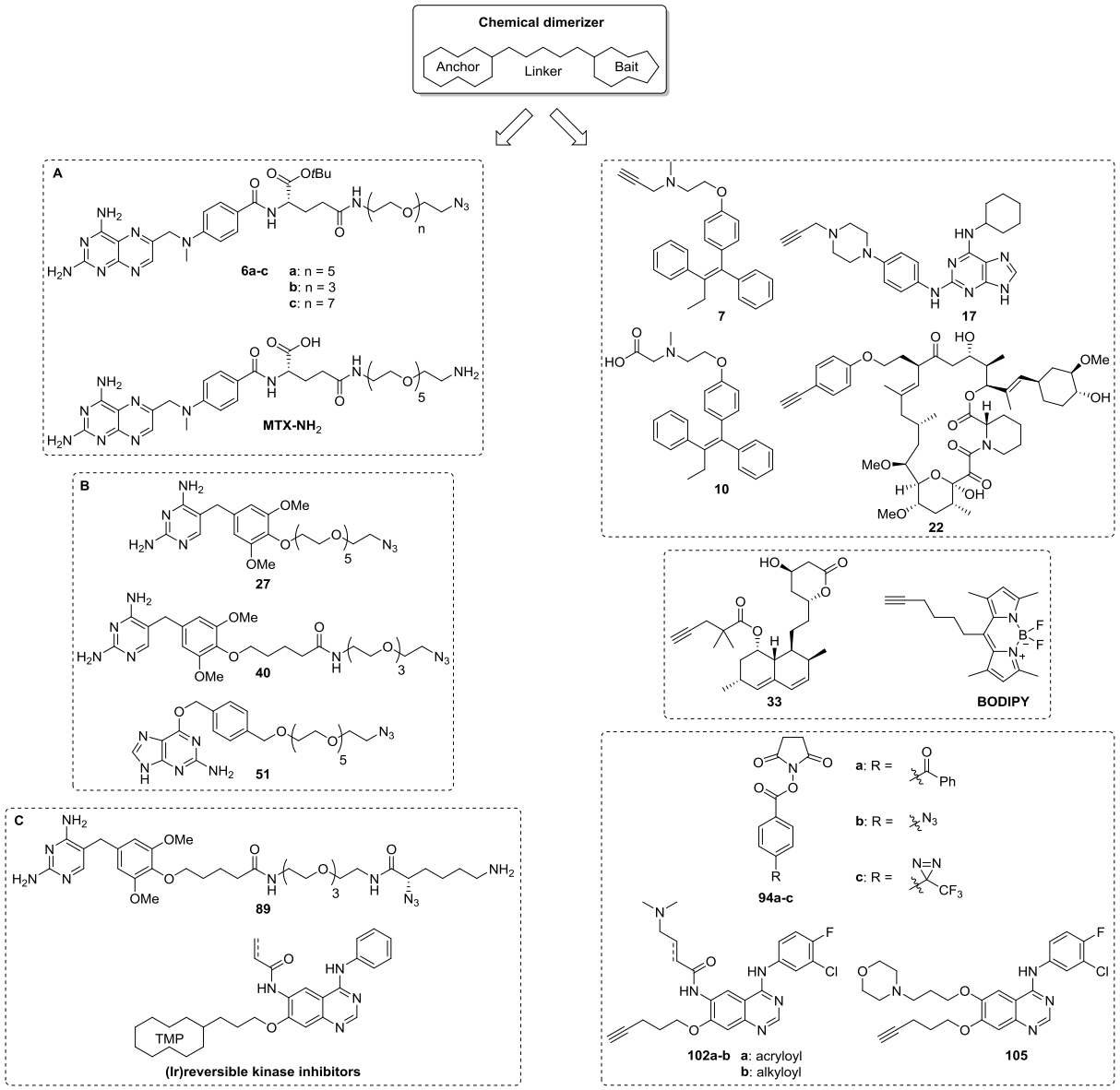


Figure VI.1: Schematic overview of strategies and building blocks applied to the synthesis of chemical dimerizers for the optimization of MASPIT.

Although the optimal spacer length may be determined by the nature of the prey and no significant differences in readout were observed upon evaluation of tamoxifen MFCs comprising the linkers mentioned above, we decided to apply the intermediate PEG_6 throughout our different optimization approaches. Likewise, a comparable reporter activity pattern was obtained for an alternative tamoxifen-based MFC synthesized via amide conjugation. Finally, in a pilot cellular array screen, FKBP12 was selectively identified as an interaction partner of FK506, thereby validating the MASPIT system.

Subsequently, evaluation of first-generation TMP fusion compounds with tamoxifen, reversine, and simvastatin as model baits, resulted in dose-response curves shifted towards lower EC_{50} values than those of their MTX congeners. Hence, these data clearly implied a successful improvement in the MASPIT system's sensitivity by introducing TMP as an alternative immobilizing anchor moiety. Furthermore, a scalable second-generation TMP-azido reagent (**40**) was synthesized that displayed a similar improvement in sensitivity, possibly owing to increased membrane permeability relative to the MTX anchor, as assessed by the differential cellular uptake of TMP- vs. MTX-linked BODIPY fluorophores (Figure VI.1, panel B). Moreover, next to its superior behavior in the MASPIT assay, **40** offered a number of important advantages over the existing MTX anchor from a chemical perspective.

Applying the SNAP-tag approach to stabilize the ternary complex via introduction of covalent bonding on the CR-hAGT chimeric anchor protein side, on the other hand, produced an inferior readout than in the MTX- or TMP-tag based assay. Additionally, the required BG-PEG_6-N_3 reagent (**51**) suffered from intrinsic thermal degradation issues, making it less attractive than the optimized TMP reagent.

Furthermore, MASPIT evaluation of the heterotrimeric PAL probes, comprising TMP, TAM and a BP/AA/DZ photophore, aimed at establishing covalent immobilization on the bait-end of the dimerizer, turned out to be less straightforward than anticipated (Figure VI.1, panel C). Upon extensive testing of these photocrosslinkers against a panel of target preys in MASPIT, no increase in luciferase signal was detected after photoexcitation, and clear-cut proof of effective photocrosslinking remained forthcoming.

Finally, we experienced some unexpected difficulties upon evaluating the EGFR kinase inhibitor-based test system that would enable the assessment of the actual contribution of a covalent linkage on the bait-end of the dimerizers with respect to the MASPIT readout. As yet, we faced a clear discrepancy between the results obtained from the radiometric kinase assay/competition experiments on the one hand, and the MASPIT/pERK WB assays on the other. Whereas the first two analyses clearly demonstrated the high *in vitro* potency/effective cellular uptake of the EGFR kinase inhibitor TFCs, the latter curiously failed to show any MASPIT activity for the reversible inhibitor probes/any cellular functional effects for all TMP conjugates. Therefore, the interaction of these kinase inhibitor probes with alternative targets in MASPIT will be explored.

CHAPTER VII

CONTEXT AND PERSPECTIVES

Broader International Context, Relevance, and Future Perspectives

The content of this chapter was partially derived from:
D. J. H. De Clercq, J. Tavernier, S. Lievens, S. Van Calenbergh, *ACS Chemical Biology* **2015**, *under review*.

VII. Broader International Context, Relevance, and Future Perspectives

VII.1. Context: Target Deconvolution and Drug Discovery

To date, merely approximately 2% of all predicted human proteins have been successfully targeted with small-molecule drugs, whereas the fraction of druggable proteins is estimated at 10-15%.[1] Likewise, it is sobering to consider that most new drugs are directed toward a small number of targets, representing only selected protein classes - the so-called 'usual suspects' - i.e. G protein-coupled receptors (GPCRs), kinases, proteases and ion channels.[2] Hence, it is believed that many potential therapeutic targets probably remain untapped. Ironically, the discovery of such novel, validated targets is currently an important bottleneck in drug discovery research, since a generally applicable methodology is still lacking.[1],[3]

Undoubtedly, there is a clear rationale for identifying the target spectrum of bioactive small molecules. For instance, target profiling technologies contribute to deciphering the mechanism of action (MOA) of leads, thereby supporting chemical optimization programs while providing a polypharmacological framework[4] for anticipating potential efficacy or toxicity-related later-stage attrition. Alternatively, these methods also allow to identify potential novel therapeutic applications of established drugs within the scope of drug repositioning projects.[5]

At present, the most frequently employed strategies to protein target identification include genomic (gene expression-based), genetic (short hairpin RNA and ORF screening), and proteomic approaches.[6],[7],[1] Classical mass spectrometry-based proteomics, analyzing relative protein abundance in a drug-treated versus control cell sample, accounts for both protein expression and degradation, hence making this approach particularly valuable in drug target discovery.[8] In this respect, the recently disclosed Functional Identification of Target by Expression Proteomics (FITExP) method might have the potential of becoming an important tool in small-molecule target deconvolution.[9]

In practice, chemical proteomics-based techniques, which mainly comprise a compound-centric approach and activity-based protein profiling (ABPP), are far more widely applied compared to elaborated genetic methods.[10],[1] An important limitation of ABPP with respect to target identification is its inherent restriction to the detection of members of a specific class of enzymes (such as proteases, glycosidases, hydrolases, phosphatases) that are active under certain conditions, for example in a particular disease state. Compound-centric chemical proteomics (CCCP), on the other hand, consists of classical affinity chromatography (pulldown) combined with high-resolution mass spectrometric (MS) analysis, followed by protein database mining and further bioinformatics for subsequent identification of binding partners.

Despite significant technological advances in quantitative MS approaches,[11],[12] affinity-based proteomics however continues to be labor-intensive, time-consuming and prone to artifacts, as it often fails to reveal physiologically relevant interactions. In fact, significant triage and additional (de)validation is required to identify the actual target(s) from the extensive list of candidates (often exceeding 500) typically resulting from such experiments.[13] In addition, sensitivity can be an important limiting factor, for instance, if the small molecule exhibits low binding affinity for its target

protein or if the target of interest is endogenously expressed at low levels. In these cases, the target protein is lost during the stringent washing steps, or its binding is obscured by the presence of highly abundant (non-specifically binding) background proteins.[14] Furthermore, the target proteins might suffer from stability issues under the classical *in vitro* conditions of chemical proteomics experiments, possibly resulting in the loss of their three-dimensional structure.[1] More than that, the protein target might be missing owing to the nature as well as the preparation conditions of the cell or tissue lysate.[13] Finally, although seemingly straightforward, the required immobilization of the small molecule of interest onto the solid support is a challenging task and represents one of the major disadvantages of pulldown approaches.

As a result, there has been an interest in the development of label-free chemical proteomics techniques, not necessitating any chemical modification of the small molecule of interest. Two such strategies, DARTS (drug affinity responsive target stability)[15] and SPROX (stability of proteins from rates of oxidation),[16] are based upon changes in thermodynamic stability resulting from a small molecule-target protein interaction. More specifically, DARTS relies on the concept that a target protein bound to a small molecule is more resistant to proteolysis compared to its unbound state.[17] SPROX, on the other hand, exploits chemical denaturant-dependent oxidation rates of methionine residues to study protein-ligand interactions. The approach is based on the finding that ligand-bound proteins are less susceptible to oxidation, hence requiring higher concentrations of denaturant to afford the same oxidation state as compared to non-binders.[18] Consequently, the application scope of the SPROX methodology is inherently restricted to proteins that contain methionine residues in their ligand-binding site. Moreover, both techniques might be limited by the binding affinity of the drug to its target, target protein abundance and sensitivity of detection by MS, as in the case of affinity chromatography.

Hence, to circumvent these limitations, several expression-cloning-based methods, such as yeast[19] and mammalian[20] three-hybrid systems, phage display,[21] and mRNA display[22] have been developed in the last decades. Collectively, these functional cloning technologies exploit the artificial overexpression of target proteins from cDNA libraries to overcome the above-mentioned potential sensitivity issues related to endogenous expression levels in proteomic methods. The last two display technologies have a number of features in common with classical pulldown approaches, including exposure of a target protein collection to the immobilized small-molecule drug, affinity-based target binding, washing of non-binders and elution of specific interaction partners. In the case of phage display, a cDNA library is fused to a phage coat protein encoding gene, so as to display candidate target proteins on the phages' surface.[21] Alternatively, in mRNA display, potential target proteins are expressed as a fusion to their corresponding mRNAs via a DNA-puromycin linker.[22]

Both phage and mRNA display benefit from the iterative amplification steps (by transfection into a bacterial host or by PCR, respectively), thus enabling the identification of low-abundance targets and of targets of low-affinity ligands.[14] Disadvantages include the lack of post-translational modifications (PTMs) and the potentially inefficient expression of the target proteins or their improper folding. Furthermore, a prerequisite for phage display is that the coding sequence of the target protein is in-frame with the one of the phage coat protein.[1] Alternatively, mRNA display might suffer from inefficient mRNA fusion or from the generation of short-length transcripts leading to the identification of protein fragments instead of full-length polypeptides.[14] In general, phage display is

more widely adopted[23],[24],[25],[26] compared to its mRNA congener, which hasn't been applied since its initial proof-of-concept study[22] more than a decade ago.

Rather than the *in vitro* pulldown and display techniques, three-hybrid systems operate in the physiologically relevant context of an intact cell. In the case of mammalian three-hybrid assays, this can be a human cell, which is the native background of the targeted protein(s) and the context in which the future drug will operate. Consequently, normal protein conformation is conserved and furthermore this might reveal potential effects of PTMs of the target or of the target's association with additional proteins or other intracellular molecules on small molecule binding. Another advantage of 3H systems includes their generality since they function regardless of the nature of both the small-molecule drug and the candidate target proteins. Moreover, 3H systems comprise unbiased screening approaches, which are not restricted to a limited panel of related proteins but cover a broad area of the human proteome.

Additionally, once a target protein for a small molecule of interest is identified, 3H systems provide an interesting binary small molecule-target assay, which allows for the evaluation of structural variations of either partner. More specifically, focusing first on the small molecule side, structure-activity relationships can be delineated, e.g. during lead optimization, based on the side-by-side comparison of analogue series via competition experiments. Hence, unfused small molecules can be ranked with respect to their *in cell* target binding affinity, as exemplified earlier for a series of antiplasmodial fumarranol analogues (see Chapter 1).[27] Alternatively, by creating and screening a systematic collection of target protein point mutants against a specific small-molecule bait, their interaction interface can be carefully mapped. This might provide additional structural information facilitating medicinal chemistry and drug design campaigns.

A final specific asset of the MASPIT screening platform includes its high throughput, which ensures rapid turnaround times (screen and hit validation in four weeks' time) and flexibility with regard to parallel screens (up to ten per week) or iterative rounds of screening compound analogues during a hit or lead optimization process.

However, as for every other target identification methodology, 3H approaches display some inherent limitations. For example, a major disadvantage of 3H methods includes the necessity to derivatize the small molecule of interest, as in the case of most chemical proteomics-based methods. Toward this end, extensive knowledge of the structure-(phenotypic) activity relationship (SAR) of the compound is needed and the process often requires elaborate chemistry efforts to identify a conjugation site on the molecule that tolerates modification without significant loss in binding affinity to the target protein(s). Additionally, the application of 3H systems is hampered in case the conjugated small-molecule drug exhibits variable or limited cellular uptake. Furthermore, most 3H methods, including MASPIT, are not compatible with the detection of (full size) transmembrane proteins.

A limitation that specifically applies to MASPIT is that compounds which exhibit affinity for any of the proteins involved in the JAK/STAT pathway might compromise the assay readout and hence are not compatible with the approach (e.g. a broad spectrum kinase inhibitor might inhibit JAK2 kinase activity). Additional disadvantages of MASPIT relate to the current target protein collection, which is

not complete (~15,000 human ORFs versus ~22,000 human genes defined) and contains only one splice variant for the majority of genes.

VII.2. Relevance

Although the number of published reports describing the application of three-hybrid systems in an actual drug discovery context is rather confined, these papers clearly demonstrate the potential of three-hybrid approaches to uncover the molecular targets of drugs and lead compounds. Major contributions to the field have been made by the Kley,[28],[20] Cornish,[29] and Johnsson[30],[31] groups. Undoubtedly, these successful *de novo* small-molecule target identification studies opened up intriguing insights into the mechanisms of action of both drug candidates and clinically approved drugs, thereby potentially providing explanations for off-target-related side effects and suggesting new or improved therapeutic applications. Moreover, 3H systems should prove useful for uncovering yet unidentified targets that contribute to the efficacy of known drugs.

For example, the discovery that sulfasalazine interferes with BH4 metabolism might explain some of its adverse effects related to changes in the concentration of neurotransmitters dependent on this cofactor.[30] Accordingly, this realization allows to set up hypotheses to alleviate these side effects, for instance by adjunct therapy with agents that would restore a correct neurotransmitter balance. Furthermore, considering sulfasalazine's specific inhibition of BH4 biosynthesis, this drug might be qualified for repurposing for other therapeutic applications, such as the suppression of chronic pain in cancer patients. Other examples highlighting the utility of 3H screening in drug MOA studies include the work by Caligiuri *et al.* and Odell *et al*.[32],[33] In the former study, the authors applied Y3H to explore the pharmacological properties and therapeutic potential of compounds derived from a trisubstituted pyrazolopyrimidinone CDK kinase inhibitor scaffold. Odell *et al.* recently used a Y3H screen to study the proteins and mechanisms involved in stage conversion in *T. gondii*, ultimately aiming at the development of new approaches to prevent the establishment of chronic infection.

Y3H screenings are also being offered as a commercial compound profiling service, e.g. by the company Hybrigenics Services[34] who acquired the three-hybrid activities of Dualsystems Biotech AG, which had formerly significantly contributed to the field.[29],[35] Also our research group has offered small molecule-target protein profiling services applying the MASPIT human cell array screening platform. The chemistry that was developed in this project has been applied to both internal drug repositioning projects with academic partners, and to fee-for-service collaborations with pharma companies.

VII.3. Recommendations and Contributions

While methotrexate-eDHFR has been the most extensively applied CID pair in the 3H field, currently two other prominent anchor moieties serve as useful starting points for compound profiling campaigns. First, based on our own experience in the design and synthesis of CIDs for the optimization of MASPIT, trimethoprim comprises a valuable alternative for MTX and in fact offers a number of important advantages over the latter, including significantly higher membrane permeability, increased solubility as well as compatibility with various reaction conditions, and lack of chirality (see Chapter 4). Second, the introduction of O^6-benzylguanine-hAGT (SNAP-tag) by the

Johnsson group, enabling screening of covalently anchored small-molecule baits, has proven a significant advance in the field.[30],[36]

Taken together, in this work we contributed to enriching the toolbox of chemical dimerizers for three-hybrid applications with two innovations. First, we introduced scalable syntheses of versatile methotrexate- and trimethoprim-azido reagents that allow for the modular and practical generation of CIDs compatible with DHFR-based 3H systems from any alkyne-functionalized small molecule of interest using click chemistry. Second, we were the first to explore the concept of photocrosslinking in the 3H field with the aim of introducing covalent bonding on the bait-end of the CID using heterotrimeric PAL probes.

VII.4. Future Perspectives

We anticipate that in the near future, the 3H field will further focus on the implementation of ligand-receptor CID pairs that enable establishing covalent immobilization of bait small molecules, following a trend that has also been seen for chemical protein labeling techniques. In the following two subsections alternative suggestions for anchor moieties and conjugation methodologies are discussed (Figure VII.1).

VII.4.1. Anchor moieties

Next to the ligand-receptor pairs used in three-hybrid systems to date (see Chapter 1), a number of alternative anchor moieties might be considered as suitable immobilization reagents (Figure VII.1). These predominantly originate from a limited number of irreversible, orthogonal covalent chemical tags, which were introduced in recent years. In order to render the TMP-tag covalent, the Cornish group designed an eDHFR variant (eDHFR:L28C) with a unique cysteine nucleophile positioned just outside the TMP-binding pocket to react with an acrylamide electrophile installed on the TMP-probe (A-TMP) via a proximity-induced Michael addition.[37],[38],[39] Furthermore, the substrate scope of the SNAP-tag labeling technology implemented in 3H systems could be extended to O^4-benzyl-2-amino-6-chloropyrimidine (CP) conjugates, which have shown higher cell permeability compared to the corresponding benzylguanine derivatives.[40] Alternatively, the specific labeling of fusion proteins based on the irreversible reaction of O^2-benzylcytosine (BC) derivatives with an engineered AGT mutant, coined CLIP-tag, might be explored.[41] Finally, yet another modular protein tagging system, i.e. HaloTag, should prove useful for covalently immobilizing small molecules of interest, originally attached to a chloroalkane linker, onto a modified haloalkane dehalogenase chimera.[42] Lastly, one could rely on biotin-based heterodimerizers and their quasi irreversible interaction with streptavidin (SA) fusions, which implies a reciprocal setup as described by Muddana *et al.*[43] or a comparable approach to the alternative Y3H system by Athavankar *et al.*[44] (see Chapter 1).

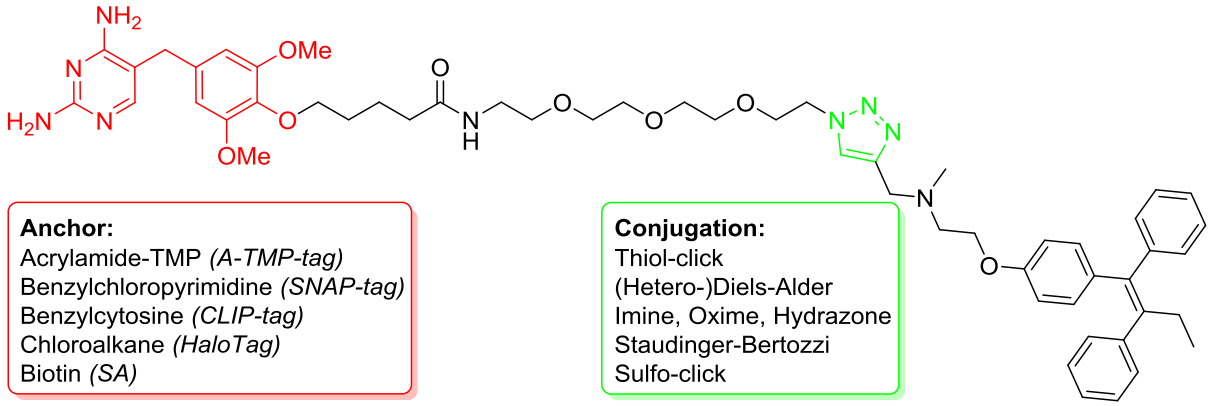


Figure VII.1: Alternative suggestions for anchor moieties and conjugation methodologies. Schematic non-comprehensive overview of anchor moieties and conjugation methods applicable to the synthesis of chemical dimerizers as exemplified by TMP-TAM **40**. All anchor moieties and conjugation methods (except for hydrazone ligation) listed, represent alternative examples that have not been applied in the context of three-hybrid studies to date. Anchor proteins are in brackets. SA, streptavidin.

VII.4.2. Conjugation methodologies

Besides the frequently employed conjugation chemistries described in Chapter 1, a range of alternative click-type modular fusion reactions could be applied (Figure VII.1). For instance, the toolbox of thiol-click reactions[45] (such as thiol-ene, thiol-yne, thiol-Michael) or (hetero-)Diels-Alder ligations,[46] encompassing [4+2] cycloadditions, might be exploited. Furthermore, one could envision the use of carbonyl-condensation reactions of the 'non-aldol' type, such as imine-, oxime- and hydrazone-bond formation.[47] An interesting recent example of the successful application of the latter conjugation reaction is provided by Moser and Johnsson,[31] describing the synthesis of BG derivatized rifampicin, among others, for target identification using an adapted SNAP-tag-based Y3H system (see Chapter 1). Finally, alternative amidation reactions, such as the Staudinger-Bertozzi ligation[48] or thio acid/sulfonyl azide amidation,[49] also denoted as sulfo-click,[50] could be appropriate for chemoselective CID construction.

It is clear that three-hybrid systems have a great potential, both for *de novo* identification of small-molecule targets and for analysis of specific compound-protein interactions. As more tools have become available that increase sensitivity and efficiency of 3H-based platforms, there is no doubt that this potential will materialize in important contributions to drug discovery.

References

[1] S. Ziegler, V. Pries, C. Hedberg, H. Waldmann, *Angew. Chem.-Int. Edit.* **2013**, *52*, 2744-2792.
[2] J. Eder, R. Sedrani, C. Wiesmann, *Nat. Rev. Drug Discov.* **2014**, *13*, 577-587.
[3] J. N. Y. Chan, C. Nislow, A. Emili, *Trends Pharmacol. Sci.* **2010**, *31*, 82-88.
[4] A. L. Hopkins, *Nat. Chem. Biol.* **2008**, *4*, 682-690.
[5] G. X. Jin, S. T. C. Wong, *Drug Discov. Today* **2014**, *19*, 637-644.
[6] G. Roti, K. Stegmaier, *Br. J. Cancer* **2012**, *106*, 254-261.
[7] M. Schenone, V. Dancik, B. K. Wagner, P. A. Clemons, *Nat. Chem. Biol.* **2013**, *9*, 232-240.
[8] M. Schirle, M. Bantscheff, B. Kuster, *Chem. Biol.* **2012**, *19*, 72-84.
[9] A. Chernobrovkin, C. Marin-Vicente, N. Visa, R. A. Zubarev, *Sci. Rep.* **2015**, *5*, 11176.
[10] U. Rix, G. Superti-Furga, *Nat. Chem. Biol.* **2009**, *5*, 616-624.
[11] M. Bantscheff, S. Lemeer, M. M. Savitski, B. Kuster, *Anal. Bioanal. Chem.* **2012**, *404*, 939-965.
[12] J. Cox, M. Mann, in *Annual Review of Biochemistry, Vol. 80* (Eds.: R. D. Kornberg, C. R. H. Raetz, J. E. Rothman, J. W. Thorner), Annual Reviews, Palo Alto, **2011**, pp. 273-299.
[13] M. Raida, *Curr. Opin. Chem. Biol.* **2011**, *15*, 570-575.
[14] G. C. Terstappen, C. Schlupen, R. Raggiaschi, G. Gaviraghi, *Nat. Rev. Drug Discov.* **2007**, *6*, 891-903.
[15] B. Lomenick, R. Hao, N. Jonai, R. M. Chin, M. Aghajan, S. Warburton, J. N. Wang, R. P. Wu, F. Gomez, J. A. Loo, J. A. Wohlschlegel, T. M. Vondriska, J. Pelletier, H. R. Herschman, J. Clardy, C. F. Clarke, J. Huang, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 21984-21989.
[16] G. M. West, C. L. Tucker, T. Xu, S. K. Park, X. M. Han, J. R. Yates, M. C. Fitzgerald, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 9078-9082.
[17] C. W. Park, S. Marqusee, *Nat. Methods* **2005**, *2*, 207-212.
[18] J. Lee, M. Bogyo, *Curr. Opin. Chem. Biol.* **2013**, *17*, 118-126.
[19] E. J. Licitra, J. O. Liu, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 12817-12821.
[20] M. Caligiuri, L. Molz, Q. Liu, F. Kaplan, J. P. Xu, J. Z. Majeti, R. Ramos-Kelsey, K. Murthi, S. Lievens, J. Tavernier, N. Kley, *Chem. Biol.* **2006**, *13*, 711-722.
[21] P. P. Sche, K. M. McKenzie, J. D. White, D. J. Austin, *Chem. Biol.* **1999**, *6*, 707-716.
[22] M. McPherson, Y. F. Yang, P. W. Hammond, B. L. Kreider, *Chem. Biol.* **2002**, *9*, 691-698.
[23] B. Van Dorst, J. Mehta, E. Rouah-Martin, V. Somers, W. De Coen, R. Blust, J. Robbens, *Toxicol. in Vitro* **2010**, *24*, 1435-1440.
[24] H. J. Jung, J. S. Shim, J. Lee, Y. M. Song, K. C. Park, S. H. Choi, N. D. Kim, J. H. Yoon, P. T. Mungai, P. T. Schumacker, H. J. Kwon, *J. Biol. Chem.* **2010**, *285*, 11584-11595.
[25] A. M. Piggott, P. Karuso, *ChemBioChem* **2008**, *9*, 524-530.
[26] J. S. Shim, J. Lee, H.-J. Park, S.-J. Park, H. J. Kwon, *Chem. Biol.* **2004**, *11*, 1455-1463.
[27] X. C. Chen, S. Xie, S. Bhat, N. Kumar, T. A. Shapiro, J. O. Liu, *Chem. Biol.* **2009**, *16*, 193-202.
[28] F. Becker, K. Murthi, C. Smith, J. Come, N. Costa-Roldan, C. Kaufmann, U. Hanke, C. Degenhart, S. Baumann, W. Wallner, A. Huber, S. Dedier, S. Dill, D. Kinsman, M. Hediger, N. Bockovich, S. Meier-Ewert, A. F. Kluge, N. Kley, *Chem. Biol.* **2004**, *11*, 211-223.
[29] A. R. Shepard, R. E. Conrow, I. H. Pang, N. Jacobson, M. Rezwan, K. Rutschmann, D. Auerbach, R. SriRamaratnam, V. W. Cornish, *ACS Chem. Biol.* **2013**, *8*, 549-558.
[30] C. Chidley, H. Haruki, M. G. Pedersen, E. Muller, K. Johnsson, *Nat. Chem. Biol.* **2011**, *7*, 375-383.
[31] S. Moser, K. Johnsson, *ChemBioChem* **2013**, *14*, 2239-2242.
[32] M. Caligiuri, F. Becker, K. Murthi, F. Kaplan, S. Dedier, C. Kaufmann, A. Machl, G. Zybarth, J. Richard, N. Bockovich, A. Kluge, N. Kley, *Chem. Biol.* **2005**, *12*, 1103-1115.
[33] A. V. Odell, F. Tran, J. E. Foderaro, S. Poupart, R. Pathak, N. J. Westwood, G. E. Ward, *Plos One* **2015**, *10*.
[34] http://www.hybrigenics-services.com/contents/our-services/discover/ultimate-ychemh
[35] M. Rezwan, D. Auerbach, *Methods* **2012**, *57*, 423-429.
[36] S. Gendreizig, M. Kindermann, K. Johnsson, *J. Am. Chem. Soc.* **2003**, *125*, 14970-14971.
[37] S. S. Gallagher, J. E. Sable, M. P. Sheetz, V. W. Cornish, *ACS Chem. Biol.* **2009**, *4*, 547-556.
[38] Z. Chen, C. Jing, S. S. Gallagher, M. P. Sheetz, V. W. Cornish, *J. Am. Chem. Soc.* **2012**, *134*, 13692-13699.
[39] T. Y. Wang, L. J. Friedman, J. Gelles, W. Min, A. A. Hoskins, V. W. Cornish, *Biophys. J.* **2014**, *106*, 272-278.
[40] I. R. Correa, B. Baker, A. H. Zhang, L. Sun, C. R. Provost, G. Lukinavicius, L. Reymond, K. Johnsson, M. Q. Xu, *Curr. Pharm. Design* **2013**, *19*, 5414-5420.

[41] A. Gautier, A. Juillerat, C. Heinis, I. R. Correa, M. Kindermann, F. Beaufils, K. Johnsson, *Chem. Biol.* **2008**, *15*, 128-136.
[42] G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohane, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit, K. V. Wood, *ACS Chem. Biol.* **2008**, *3*, 373-382.
[43] S. S. Muddana, B. R. Peterson, *Org. Lett.* **2004**, *6*, 1409-1412.
[44] S. Athavankar, B. R. Peterson, *Chem. Biol.* **2003**, *10*, 1245-1253.
[45] C. E. Hoyle, A. B. Lowe, C. N. Bowman, *Chem. Soc. Rev.* **2010**, *39*, 1355-1387.
[46] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem.-Int. Edit.* **2001**, *40*, 2004-2021.
[47] S. Ulrich, D. Boturyn, A. Marra, O. Renaudet, P. Dumy, *Chem. - Eur. J.* **2014**, *20*, 34-41.
[48] M. Köhn, R. Breinbauer, *Angew. Chem.-Int. Edit.* **2004**, *43*, 3106-3116.
[49] N. Shangguan, S. Katukojvala, R. Greenburg, L. J. Williams, *J. Am. Chem. Soc.* **2003**, *125*, 7754-7755.
[50] D. T. S. Rijkers, R. Merkx, C.-B. Yim, A. J. Brouwer, R. M. J. Liskamp, *J. Pept. Sci.* **2010**, *16*, 1-5.

APPENDIX

CURRICULUM VITAE

Curriculum Vitae

PERSONAL DETAILS

Name: Dries J. H. De Clercq
Date of Birth: May 10, 1988
Nationality: Belgian
E-mail: declercq.dries@gmail.com

EDUCATION

2011 - Present **PhD Pharmaceutical Sciences** – Ghent University, Belgium

- *Title:* "Synthesis of Chemical Dimerizers for the Optimization of MASPIT"
- Laboratory for Medicinal Chemistry, UGent
- *Promoters:* Prof. Serge Van Calenbergh, Laboratory for Medicinal Chemistry, UGent; Prof. Jan Tavernier, Cytokine Receptor Lab, VIB & UGent; Dr. Sam Lievens, Cytokine Receptor Lab, VIB & UGent

2006 - 2011 **Master of Science in Drug Development** – Ghent University, Belgium

- Graduated 'Summa cum laude'
- *Thesis:* "Immediate Release Filmcoating of Tablets: Process Development on the Laboratory Coating System LHC-25 via Design of Experiments (DoE)" (Johnson & Johnson PRD, Janssen Pharmaceutica, Beerse, Belgium)

2000 - 2006 **Latin-Mathematics** – Broederschool Humaniora Sint-Niklaas, Belgium

- Graduated 'Summa cum laude'

ADDITIONAL EDUCATION

2013 - 2014 Organometallics and Catalysis (Department of Organic Chemistry, UGent)

2012 - 2013 Advanced NMR Spectroscopy: Application to Structure Analysis (Department of Organic Chemistry, UGent)

Advanced Organic Chemistry - Medicinal Chemistry Course (Johnson & Johnson PRD, Janssen Pharmaceutica & Department of Chemistry, KU Leuven)

2011 - 2012 Synthetic Methods in Organic Chemistry (Department of Organic Chemistry, UGent)

RESEARCH FELLOWSHIPS

2011 - 2015 PhD Grant from the Special Research Fund of Ghent University (BOF)

TEACHING EXPERIENCE

2011 - 2015 **Lab Instructor** – Integrated Bachelor Course (J000407)

- Ghent University, Prof. B. De Spiegeleer
- 3rd Bachelor Pharmaceutical Sciences: Practical coaching of Bachelor Projects on the synthesis and structural characterization of small-molecule drugs

2011 - 2013 **Teaching Assistant** – Medicinal Chemistry (J000405)

- Ghent University, Prof. S. Van Calenbergh
- 3rd Bachelor Pharmaceutical Sciences: Seminars and coached exercises on the synthesis of prototypical small-molecule drugs

SUPERVISION OF MASTER STUDENTS

2013 - 2014 Lynn Verbestel

- *Thesis:* "Synthese van een Fotoactiveerbare Tamoxifen Fusiecomponent ter Optimalisatie van MASPIT"

2012 - 2013 Brecht Vanbillemont

- *Thesis:* "Synthese van Tamoxifen Fusiecomponenten ter Optimalisatie van MASPIT: Variatie van de Immobiliserende Groep"

PUBLICATIONS

- Towards Synthetic Tools to Induce Covalent Bait-Prey Interactions in MASPIT.
  D. J. H. De Clercq, M. D. P. Risseeuw, I. Karalic, A.-S. De Smet, D. Defever, J. Tavernier, S. Lievens, S. Van Calenbergh
  Manuscript in preparation

- A Cell Microarray Platform for High-Throughput Mapping of Protein Interactions in Mammalian Cells.
  S. Lievens, I. Petta, D. Masschaele, L. De Ceuninck, V. Vauthier, J. Van der Heyden, I. Lemmens, **D. J. H. De Clercq**, M. D. P. Risseeuw, S. Gupta, D. Defever, N. Vanderroost, A.-S. De Smet, L. Martens, S. Van Calenbergh, C. Libert, K. De Bosscher, D. Hill, M. Vidal, J. Tavernier
  Manuscript in preparation

- Chemical Dimerizers in Three-Hybrid Systems for Small Molecule-Target Protein Profiling.
  D. J. H. De Clercq, J. Tavernier, S. Lievens, S. Van Calenbergh
  ACS Chemical Biology **2015**
  Under review

- Alternative Reagents for Methotrexate as Immobilizing Anchor Moieties in the Optimization of MASPIT: Synthesis and Biological Evaluation.
  D. J. H. De Clercq, M. D. P. Risseeuw, I. Karalic, A.-S. De Smet, D. Defever, J. Tavernier, S. Lievens, S. Van Calenbergh
  ChemBioChem **2015**, *16*, 834-843.

- Kinase Substrate Sensor (KISS), a Mammalian In Situ Protein Interaction Sensor.
  S. Lievens, S. Gerlo, I. Lemmens, **D. J. H. De Clercq**, M. D. P. Risseeuw, N. Vanderroost, A.-S. De Smet, E. Ruyssinck, E. Chevet, S. Van Calenbergh, J. Tavernier
  Molecular & Cellular Proteomics **2014**, *13*, 3332-3342.

- A "Clickable" MTX Reagent as a Practical Tool for Profiling Small-Molecule–Intracellular Target Interactions via MASPIT.
  M. D. P. Risseeuw,* **D. J. H. De Clercq**,* S. Lievens, U. Hillaert, D. Sinnaeve, F. Van den Broeck, J. C. Martins, J. Tavernier, S. Van Calenbergh
  *Co-first author
  ChemMedChem **2013**, *8*, 521-526.

CONFERENCES

- Gordon Research Seminar & Conference: High Throughput Chemistry & Chemical Biology
  New London, NH, USA, June 13-19, 2015
  Poster presentation: Small-Molecule Fusion Compounds for the Optimization of MASPIT, a High-Throughput Three-Hybrid Target Deconvolution Assay (**D. De Clercq**, S. Lievens, M. Risseeuw, J. Tavernier, S. Van Calenbergh)

- Annual One-Day Meeting on Medicinal Chemistry of SRC & KVCV (MedChem 2014)
  Braine-L'Alleud, Belgium, November 21, 2014

- Chemical Biology 2014 (EMBO Conference Series)
  Heidelberg, Germany, August 20-23, 2014
  Poster presentation: Synthesis and Biological Evaluation of Small Molecule Fusion Compounds for the Optimization of MASPIT, a Three-Hybrid Target Deconvolution Assay (**D. De Clercq**, S. Lievens, M. Risseeuw, J. Tavernier, S. Van Calenbergh)

- Chemistry Conference for Young Scientists 2014 (ChemCYS)
  Blankenberge, Belgium, February 27-28, 2014
  Oral presentation: Design, Synthesis and Biological Evaluation of Tamoxifen-Fusion Compounds for the Optimization of MASPIT, a Three-Hybrid Target Deconvolution Assay (**D. De Clercq**, S. Lievens, M. Risseeuw, J. Tavernier, S. Van Calenbergh)

- Knowledge for Growth 2013 (FlandersBio's Annual Life Sciences Convention)
  Ghent, Belgium, May 30, 2013
  Poster presentation: Syntheses of Methotrexate-Fusion Compounds for Target Profiling of Small Molecules with MASPIT (**D. De Clercq**, M. Risseeuw, S. Lievens, J. Tavernier, S. Van Calenbergh)

- 2nd Meeting of the Paul Ehrlich MedChem Euro-PhD Network
  Ljubljana, Slovenia, September 9-11, 2012
  Oral presentation: Syntheses of Methotrexate-Hybrid Compounds for Target Profiling of Small Molecules with MASPIT (**D. De Clercq**, M. Risseeuw, S. Lievens, J. Tavernier, S. Van Calenbergh)

- Chemistry Conference for Young Scientists 2012 (ChemCYS)
  Blankenberge, Belgium, March 1-2, 2012
  Poster presentation: Syntheses of Methotrexate-Hybrid Compounds for Target Profiling of Small Molecules with MASPIT (**D. De Clercq**, M. Risseeuw, S. Lievens, J. Tavernier, S. Van Calenbergh)

- Annual One-Day Meeting on Medicinal Chemistry of SRC & KVCV (MedChem 2011)
  Ghent, Belgium, November 25, 2011

Dankwoord – Acknowledgements

Dankwoord – Acknowledgements

Na een intense onderzoeksperiode van vier jaar is het tijd om de vele mensen te bedanken die hebben bijgedragen tot het realiseren van deze thesis.

Eerst en vooral wens ik mijn promotoren, Prof. Serge Van Calenbergh, Prof. Jan Tavernier en Dr. Sam Lievens, uitvoerig te bedanken. Serge, hartelijk dank voor de kans die je me gegeven hebt om een doctoraat in jouw labo aan te vatten. Ik heb je steeds openstaande deur, opvolging en ondersteuning tijdens dit onderzoek enorm geapprecieerd. Bedankt ook voor je vertrouwen en de mogelijkheden die ik kreeg om verschillende cursussen, workshops en congressen in binnen- en buitenland te volgen. Speciale dank ook voor je nauwgezette verbeteringen bij het uitwerken van mijn manuscripten en dit proefschrift.
Jan, je enthousiasme, heldere inzichten en constructieve suggesties betekenden zowel voor dit onderzoek als voor mezelf een enorme stimulans. Ik wens je hartelijk te bedanken voor de kans die je me aanreikte om enkele weken van dichtbij mee te volgen in je labo. Bovendien heb ik de vrijheid die je me in de afgelopen vier jaar gaf sterk gewaardeerd. Bedankt ook voor het kritisch nalezen van mijn papers en dit werk.
Sam, speciale dank voor de efficiënte en nauwe samenwerking op vele vlakken die aan dit werk verbonden zijn. Ik wil je in het bijzonder bedanken voor het ontwikkelen en opstellen van de moleculaire biologie protocollen ter evaluatie van de gesynthetiseerde verbindingen. Hartelijk dank ook om mij hierin zo geduldig wegwijs te maken. Ook voor je waardevolle bijdragen tot mijn artikels en deze thesis wil ik je erg bedanken.

Verder wens ik iedereen te bedanken die betrokken was bij de chemie van dit onderzoek, in de eerste plaats Dr. Martijn Risseeuw. Martijn, ongelooflijk bedankt om me vier jaar geleden deskundig te introduceren tot de fascinerende wereld van de synthetische chemie. Ik zal mijn eerste final compound, 2-methylamfetamine, die we samen bereidden in opdracht van een toxicologisch labo niet snel vergeten. In de jaren die daarop volgden heb je me enorm veel kennis en vaardigheden bijgebracht, waarvoor ik je bijzonder dankbaar ben. Ik ben ervan overtuigd dat je fijnzinnige zinsneden ook in de toekomst nog een bron van 'incubatiehumor' zullen vormen.
Uiteraard mag ook Izet Karalic hier niet ontbreken. Izet, hartelijk dank voor je hulp bij de vaak uitdagende prep HPLC opzuiveringen van mijn final compounds die dikwijls tot avontuurlijke projecten leidden. Verder ook erg bedankt voor het vertrouwen m.b.t. de LCMS, de 'Karalic-modification' van de rotavaps, hulp bij de NMR en de dagelijkse technische/veiligheids-ondersteuning in het algemeen. Speciale dank ook voor je 'helicopter view' die me veel relativeringszin heeft bijgebracht en me op een strategische manier naar wetenschappelijke probleemstellingen deed kijken.
Hierbij wil ik eveneens Dr. Ulrik Hillaert en Jolien Janssens bedanken om het synthese-pad al wat te effenen voordat ik het labo vergezelde. Ook Brecht en Lynn, de thesisstudenten die ik begeleid heb, wil ik bedanken voor hun bijdrage tot dit onderzoek.

Daarnaast wens ik alle leden van het Cytokine Receptor Lab (VIB-UGent) die instonden voor de biologische evaluatie van de gesynthetiseerde verbindingen uitvoerig te bedanken, in het bijzonder Anne-Sophie De Smet en Dieter Defever. Hartelijk dank ook dat ik jullie tijdens mijn stage in het CRL

letterlijk op de voet mocht volgen en voor het delen van jullie kennis en het gezamenlijk redeneren over mijn meest uiteenlopende vragen.

Prof. José Martins, Dr. Davy Sinnaeve en Dr. Freya Van den Broeck (NMR and Structure Analysis Unit, UGent) wil ik bedanken voor de nauwgezette NMR analyse van enkele MTX- en FK506-analogen.

Het Bijzonder Onderzoeksfonds van de Universiteit Gent wens ik te bedanken voor vier jaar financiële steun (doctoraatsmandaat).

I would like to express special thanks to all my colleagues. Your presence, humour and enthusiasm created an enjoyable working atmosphere and made the days in the lab fly by. I highly appreciated the inspiring discussions and I am very grateful for the sharing and transfer of knowledge, skills, specific experience, protocols, creative ideas... Of course, many thanks for all the amusing activities outside the lab!
René, I spent practically my entire training period next to you in the lab and I would like to thank you for the great moments we shared. Thanks for furnishing my fume hood with your squawking lizard, I am confident that Mingcheng will further take care of him. Arno, erg bedankt voor je aangenaam gezelschap in de bureau en het labo, de boeiende chemie- en NMR-uitwisselingen en alles daarbuiten. Joren en Fabian, speciale dank voor de fijne tijden tijdens de lessen op de Sterre en verschillende congressen, alsook jullie vele tips en tricks (2D NMR, µW, …) en waardevolle suggesties. Lijun and Mingcheng, a big thanks for helping me to find virtually any paper that I couldn't access. Also many thanks for sharing your office during my final writing period. Charlotte, bedankt voor de toffe momenten en gesprekken in het labo en de leuke samenwerking tijdens de FaBaP. Jonas, bedankt voor het delen van je inzicht tijdens de werkcolleges op S4, en het spontane binnenvallen om interessante reactiemechanismen of synthesestrategieën uit de doeken te doen. Jakob, ook bij jou kon ik steeds terecht voor een aangename babbel tijdens mijn schrijffase bij jullie in de bureau.
I wish you all the best in your present and future endeavors!

Many thanks to all my former colleagues! Kiran, Nora, Annelies en Thomas, thanks for your company, nice chats, helpful advice and guidance during my first steps in the lab.

Natuurlijk wil ik ook graag Annelies Van Hoeck hartelijk bedanken. Annelies, erg bedankt om alle administratie en bestellingen steeds zo vlot en efficiënt te regelen. Bedankt ook voor je hulp bij het maken van reisplannen voor congressen, en de goede gesprekken tussendoor. Speciale dank om ons labo en de hele faculteit op sleeptouw te nemen voor de UGent zwemkampioenschappen!

Uiteraard wil ik ook iedereen buiten het labo die mij de voorbije jaren gesteund heeft speciaal bedanken.

De wekelijkse turntrainingen waren voor mij een ideale ontspanningsbron en uitlaatklep. Ik wil iedereen van 'De Turners' hartelijk bedanken voor de fantastische sfeer elke vrijdagavond en om mijn gedachten even volledig van mijn onderzoek weg te halen. Speciale dank hierbij ook voor het KG-team: Pierre, Peter, Jelle en Pieter, ik hoop dat we onze T-shirts nog veel kunnen laten schitteren de komende jaren!

Ook alle koersvrienden wil ik bedanken voor de ontspannende tochten de afgelopen jaren. Bijzondere dank ook aan kopman Sam en wegkapitein Bert: het deed telkens deugd om eens goed uit te waaien langs de Schelde tijdens mijn schrijfperiode.
Ook HVM Thomas en Charlotte mogen hier niet ontbreken. Hartelijk bedankt voor jullie steun, raadgevingen, de ontspannende uitstapjes samen en zoveel meer!

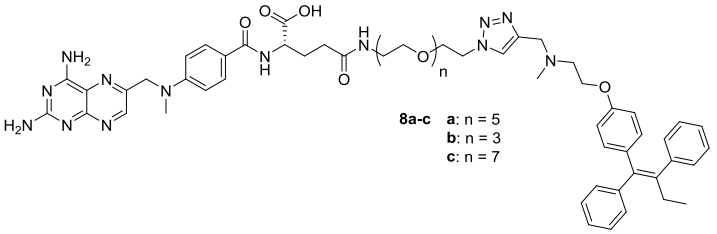
Uiteraard wil ik ook mijn hele familie uitdrukkelijk bedanken voor alle aanmoedigingen en interesse.

Annemie, Bart en Gert, ik wil jullie ontzettend bedanken voor jullie begrip, meeleven en steun tijdens mijn doctoraat en daarbuiten.

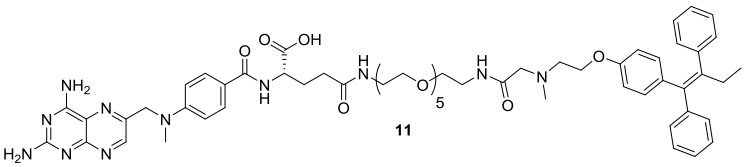
Een speciale plaats hier is weggelegd voor mijn ouders en broer: Mams, Paps, Bart, ik ben jullie ongelooflijk dankbaar voor alles wat jullie tijdens mijn hele leven al voor mij gedaan en betekend hebben. Bedankt voor alle kansen en mogelijkheden die jullie mij keer op keer aanbieden. Het deed steeds bijzonder veel deugd om tijdens de afgelopen negen jaar in Gent terug thuis te komen. Jullie zorgden steeds voor een plaats waar ik volledig tot rust kon komen en alles kwijt kon. Bedankt om zo'n fantastische ouders en broer te zijn!

Mijn allergrootste dank gaat uit naar de belangrijkste persoon in mijn leven: Marieke. Ik besef dat mijn doctoraat ook op jou een hele impact heeft gehad en ik wil je enorm bedanken voor alles wat je voor mij betekend hebt. Bedankt om dit alles met mij te delen. Zonder jou had ik dit nooit kunnen realiseren. Bedankt om er altijd zo onvoorwaardelijk te zijn!

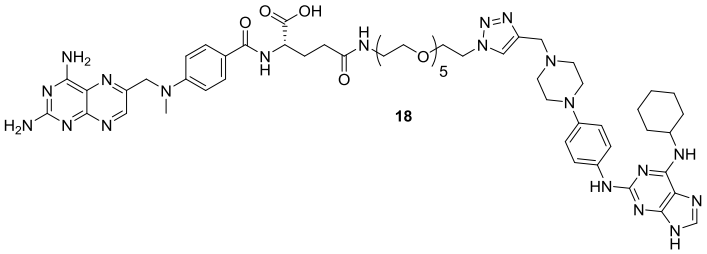
Overview of Synthesized and Evaluated Compounds



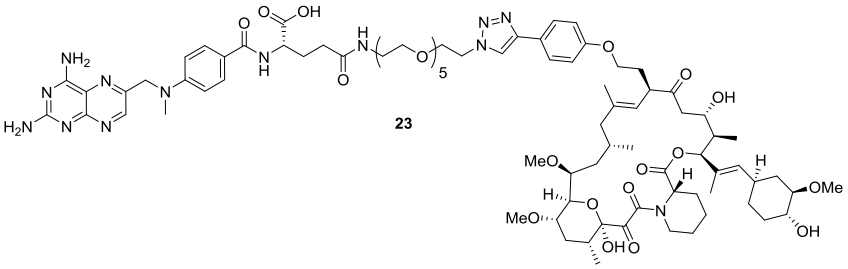
8a-c **a**: n = 5
b: n = 3
c: n = 7



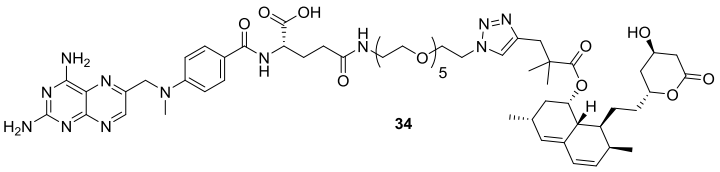
11



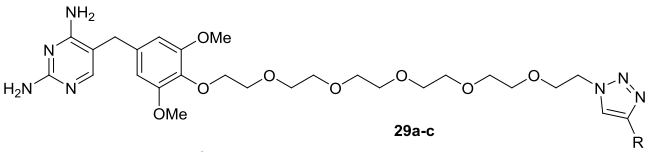
18



23

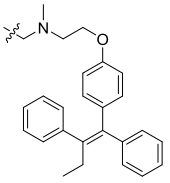


34

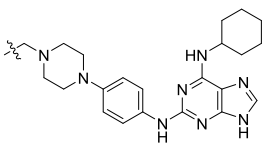


29a-c

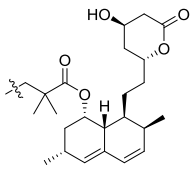
a: R =

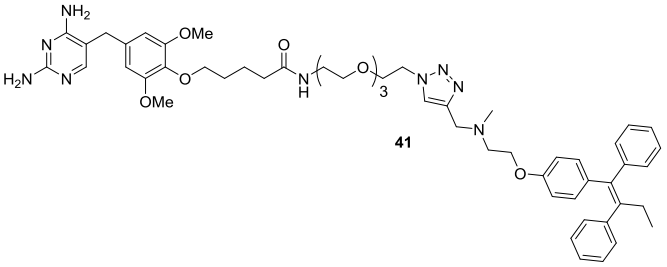


b: R =

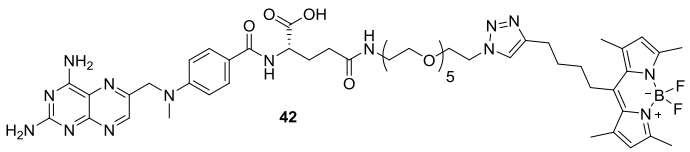


c: R =

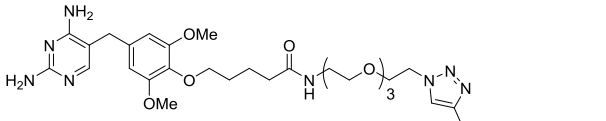




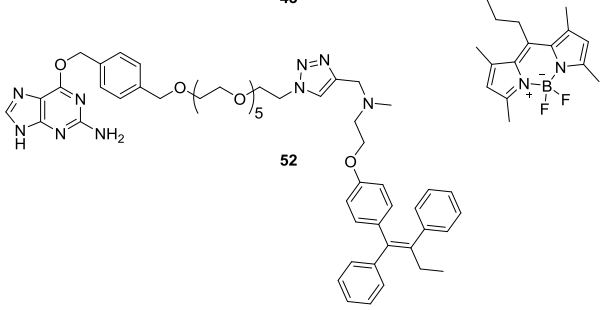
41



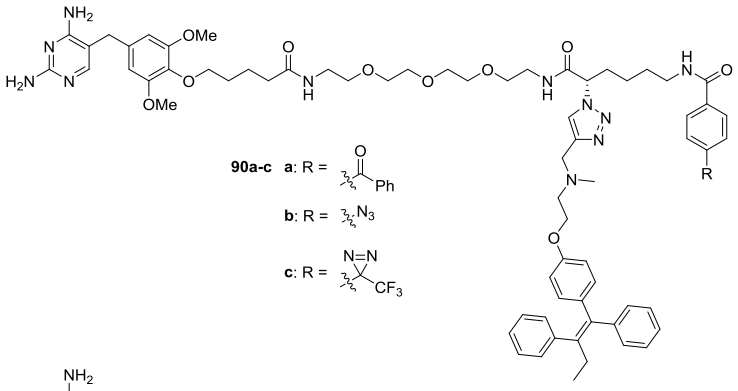
42



43



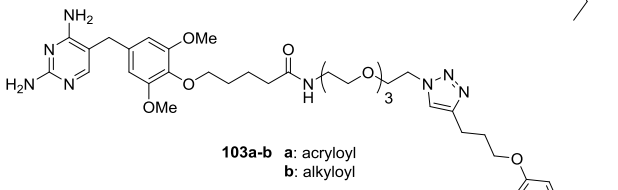
52



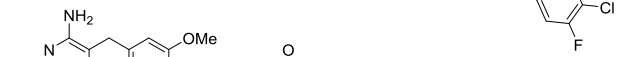
90a-c **a**: R =

b: R =

c: R =



103a-b **a**: acryloyl
b: alkyloyl



106