



UNIVERSITAT DE
BARCELONA

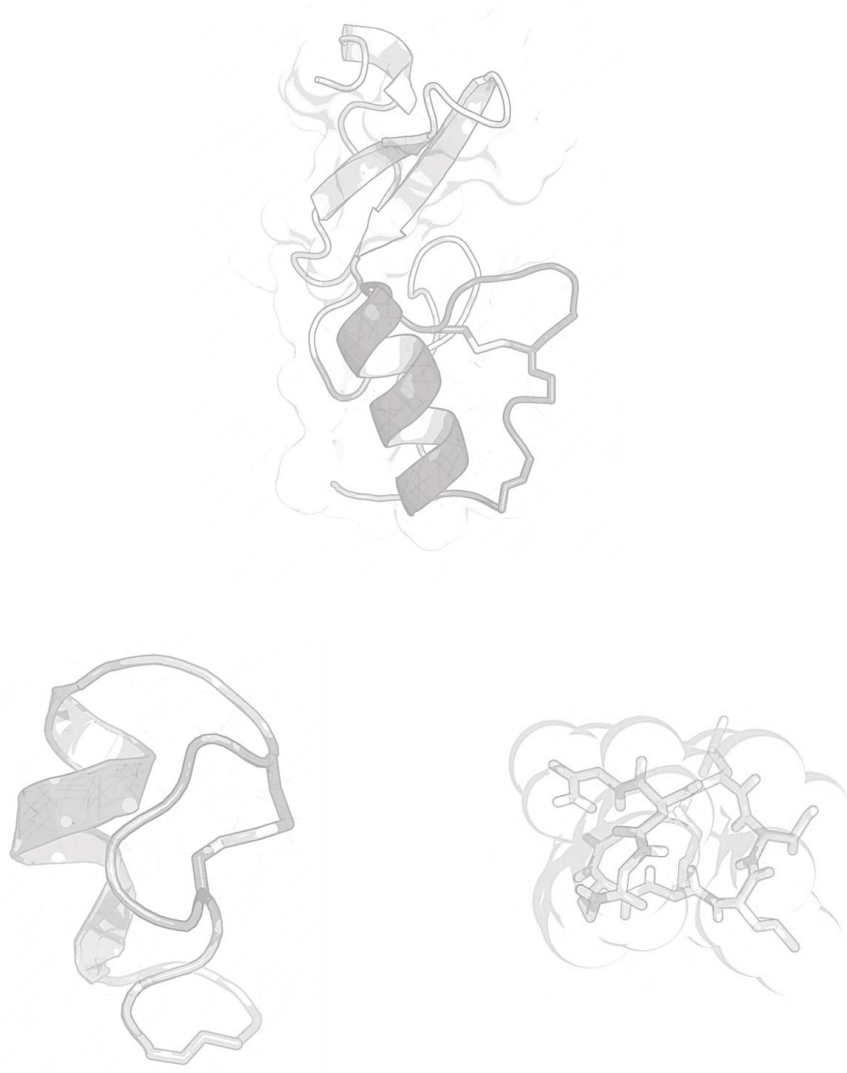
Design, synthesis and biophysical evaluation of peptides targeting pharmacologically relevant proteins

Salvador Guardiola Bagán

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

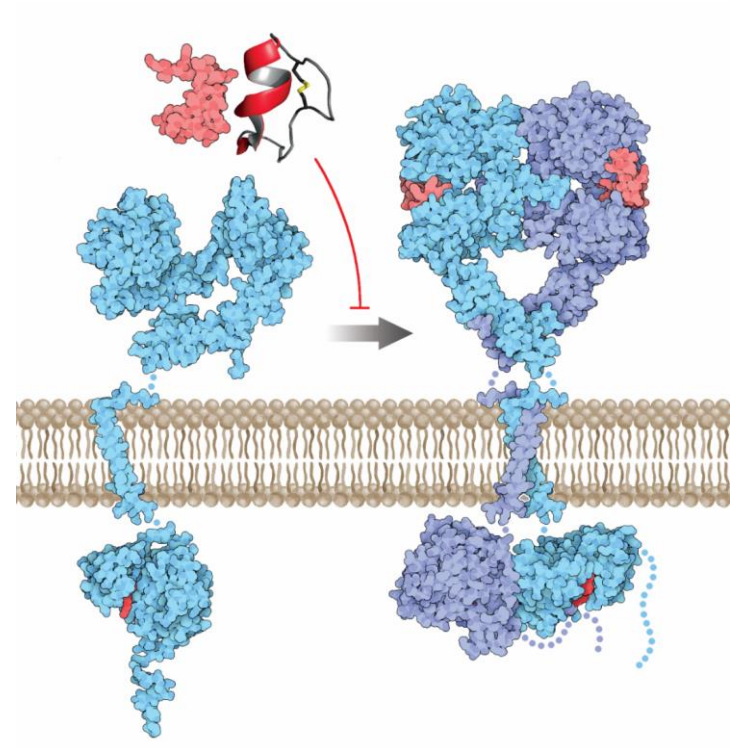


2017

Tesi doctoral

Salvador Guardiola Bagán

DESIGN, SYNTHESIS and BIOPHYSICAL evaluation of PEPTIDES targeting pharmacologically relevant **PROTEINS**



SALVADOR GUARDIOLA BAGÁN



UNIVERSITAT DE
BARCELONA



Programa de doctorat de química orgànica

INSTITUTE
FOR RESEARCH
IN BIOMEDICINE

Design, synthesis and biophysical evaluation of peptides targeting pharmacologically relevant proteins

Salvador Guardiola Bagán

Tesi doctoral dirigida per:

Prof. Ernest Giralt Lledó

Dra. Laura Nevola

Universitat de Barcelona
Facultat de Química
Departament de Química Orgànica

Co-fundadora i CSO
IDP Pharma

Barcelona, 2017.

Agraïments

En primer lloc, vull agrair als meus directores de tesi Ernest i Laura. A l'**Ernest**, vas ser la primera persona que vaig conèixer del lab, i em vas transmetre la confiança i el suport per fer la tesi en el teu grup. Ets un gran científic i millor persona, a més l'excel·lent visió científica i pedagògica que dones als projectes, transmetes uns valors molt importants de respecte i integritat. A la **Laura**, perquè tot i la teva marxa, has estat una persona propera, atenta i d'un talent que sembla no esgotar-se mai. Els teus consells i la teva creativitat m'han solucionat més d'un maldecap! Et desitjo molt d'èxit a IDP Pharma (i amb la càmera!).

Als companys del lab, per haver-me acompanyat, ajudat i aguantat aquests anys (i el que encara me queda). Els que estaven de bon principi: el **Xavi** (a Magdalena encara se recorden del xic de l'Hospitalet), la **Mendi** (l'animadora salsera #1 i de les porres periques que acaben en golejada), la **turbo-Prinyó** (la xica del multi-tasking magistral, ara casta empresarial), el **Berni** (capaç de fer miscible la cuina molecular i el bon futbol), l'**Abraham** (pur sentiment BTTero a peu de poiata), el **Benji** (*lord* de la química, la botànica i la poesia), la **Núria** (per la teva experiència i interès en la nostra ciència, i per donar suport a aquella raquetada èpica), l'Albert **Puigpinós** (per tenir solucions per tot i estar disposat a compartir-les) i la Marta **Mena** (per la teva dedicació i bon rotllo). Segurament que em deixo algú, deu ser el doctorat que mata moltes neurones.

Als companys que han estat aquí fins el final, i que tot just acaben de marxar: al **Pol** (per la teva energia, companyerisme i feina incansable—han agut d'apujar el preu del cafè de màquina 0.05€ per mantindre marges, i per les fitxes de pòquer que ens quedem en herència), a l'**Aurelio** (per convertir molts emails en obres d'art i donar una mica de cultura a les nostres vides), a la **Montserrat** (ets una gran scientist, i no t'ho negaré, pels postres italians que portaves), a la **Sonia** (pel teu carisma, per estar sempre llesta per donar un cop de mà i amb un somriure al capdavant), al **Martí** (potser no trobes la cura a l'Alzheimer, però algun dia m'explicaràs com fas que els dies duren >24h) i a la **Natàlia** (pel teu entusiasme, que és contagiós, i la teva dedicació a la ciència, je te souhaite le meilleur chez les voisins).

I clar, als supervivents que seguïu ací a peu del canó, així com els que heu arribat aportant sang fresca al lab: al tio **Pep** (pare dels entrenaments aquàtics i de tants altres moments memorables, si te fiques cabut t'arree quatre carajillos de la terreta), a la **Cris "füsta"** Belmonte (pel teu bon rotllo

de cada dia, però ojo, que si me guanyes a la piscina li tiro malefici al SAW), a la **Macarena** (nuestra mejor líder en el team-building y en el running... cuándo la revancha de los Paketones?), a la **Cris Díaz** (la reina dels D-peptides, sempre ahí ajudant a domesticar els UPLCs que es rebel·len), a la **Júlia** (amb qui comparteixo afició incondicional pels bicycles), a la **Monica** (por hacer fácil lo complicado... si lo confieso, el de tu queso fui yo, perdón!), a l'**Edu** (pel bon rotllo i per les sortides a cremar roda per Collserola), al **Jesús Seco** (per l'ajuda amb els càlculs i la teva paciència ensenyant un *dummie* de Linux), al **Daniele** (por tu magnetismo hacia las bandejas de canapés, te echaremos en falta), a la **Cris Garcia** (que be se't veu , cada dia més gran i feliç), a l'**Adam** (el fondista polifacètic del lab, ets un màquina), al **Mark** (en qué mal día se te ocurrió hablar de ese bicho Mycoplasma, con lo felices que éramos?), al **Toni** (tu botella de rakija espera impaciente el gran momento de la lectura de esta tesis) i a la **Maria** i la **Silvia** (per portar aire fresc al lab, ànims que sou the future of science!). I als empresaris d'Iproteos: la **Tere**, el **Roger**, la **Sandra**, el **Pep** i la **Núria**.

Als ex-companys del 300 i 100, ara traslladats a les profunditats del subsòl, no us oblidem: l'**Alex** (el quimic amb més *touch*), el cocodrilo **Ivan** (por los memes, videos yonkis y las risas que nos hemos echado), la **Judith** (ja t'has infiltrat aquí com una més), i l'**Helena**. A tu sobre tot per haver-te convertit en una persona tan especial en aquests darrers temps. I als que heu estat els puntals de l'equip en el passat: al **Kamil**, que no ens oblidem i ens portes de tant en tant l'essència de la Pachamama; i al **Juan**, el jabalí de Collserola, gracias por tu experiencia y consejos, y por las tertulias de verano del Tour.

A totes les integrants de la facility d'espectrometria de masses de l'IRB, on he fet una part important de la meva recerca, feu entre totes un gran equip! En especial a la **Marta**, per la teva defensa del SAW (fins i tot quan allò era una poc més que ferralla a Rüsselsheim) i a la **Mireia**, per totes les hores d'espectres, fittings i deconvolucions (al final vas guanyar al Lumos per golejada).

L'IRB està ple de gent increïble, tot i no compartir la recerca amb ells, he tingut la sort de conèixer uns quants. Als companys amb qui vaig aterrar aquí fa 4 anys: en especial a l'**Àlbert**, la **Busra**, la **Rosa**, la **Sandra** i la **Francisca**. A la **Laura Orellana**, per la teva dedicació durant la lab rotation. Als seguidors de Dixit i enamorats dels unicorns: al **Francisco**, la **Georgia**, l'**Antonija**, la **Joana**, l'**Erik** i l'**Adrian**. Als companys de patejades mil per les muntanyes de Tarragona i Castelló: a la **Gemma**, la

Laura, l'**Helena** Roura, el **Juergen** i el **Ricardo**. Al **Craig**, el guiri més mediterrani. I com no, als bojos de la bici, que deixant de banda al *rider* #1 (dr. López), són el **Carlos** (t'agraeixo sobre tot els sobres de guijuelo, que ens han salvat de més d'una) i el **Miquel** (a pesar dels paranys que ens preparaves a peus del Rocacorba). No m'oblido del *boss* **Guinovart**, la teua feina al capdavant de l'IRB ha estat encomiable, no més que la teua memòria per recordar batalletes i anècdotes.

A la gent que, d'una forma o d'una altra, heu posat les pedres per construir aquest camí que m'ha endut fins aquí. A Vicente **Cotanda**, extraordinari mestre del Ribalta. A **Pelayo** i a **Santi**, grans profes de la facultat, pel seu suport i dedicació; i als que vau estar al seu grup quan vaig fer el meu torpe debut a la poiata: a l'**Eva**, a la **Marta**, al **Matías**, a l'**Elena**, a l'**Eli** i al **Carles**. Aquest últim és l'únic que s'atreveix a compartir algun dinar amb mi darrerament.

Als amics de tota la vida de Castelló, per tots els grans moments que hem passat junts. Tot i que escampats (que no estampats) per la geografia, seguim estant junts: **Pau**, **Aldo**, **Jaume**, **Yurra**, **Jose**, **Pepe**, **Deivid**, **Manuel**, **Carlos**, **Nico** i **Fran**. Als companys de pis: a **Ramón**, a **Sol**, a **Lluís** i al cavaller de Castella, el **Raúl**. Tots m'heu fet el dia a dia molt més fàcil i feliç.

A **pare** i **mare** (aquí valorats sota el prisma de la genètica que ens uneix), m'ho heu donat tot i m'ho seguiu donant, sense esperar res a canvi. Heu sobreviscut estoicament a 4 anys de distanciament científic del vostre fill únic (*mater dixit*).

I, finalment, a tu, **lector anònim**, que amb esperit intrèpid agafes aquest treball per endinsar-te en el mon dels pèptids i les interaccions biomoleculars. Estic segur que t'agradarà, i si no trobes resposta a allò que busques, espero que almenys t'emportes una part de la il·lusió i la entrega amb que aquesta feina s'ha fet.

*No deixarem mai d'explorar
i al final de la nostra exploració
arribarem allà on vam començar
i coneixerem el lloc per primera vegada*

T.S. Eliot

CONTENTS

ABBREVIATIONS	I
INTRODUCTION	3
Modulation of protein-protein interactions (PPIs) using designed peptides	3
PPIs: typical features	4
The potential of designed peptides for targeting PPIs	4
<i>Cyclic peptides to mimic protein structural motifs</i>	6
Cyclic peptides that modulate PPIs with anti-cancer activity	8
<i>A dual MDM2 and MDMX stapled peptide for p53-dependent cancer therapy</i>	8
<i>Cyclic α/β-peptides that target the VEGF-VEGFR interaction</i>	11
The EGF-EGFR pathway: from oncogene discovery to targeted cancer therapeutics	14
The EGFR signaling pathway in cancer	14
<i>The ErbB family of cell-surface receptors and their ligands</i>	15
<i>Basic insights into EGFR structure and dynamics</i>	17
<i>Downstream signaling and cellular responses</i>	19
Development of EGFR antagonists for cancer treatment	20
Strategies to overcome resistance: next-generation EGFR inhibitors and combination therapies	23
Peptides as covalent inhibitors of specific targets. The case of prolyl oligopeptidase (POP)	27
The interest in covalent peptide inhibitors	27
Proteases as peptide targets	28
POP, a protease with pleiotropic effects	29
The medicinal chemistry of POP inhibitors	31
OBJECTIVES	35
RESULTS AND DISCUSSION	39
CHAPTER 1: DOCKING-BASED DESIGN OF EGF PEPTIDE LIGANDS	41
Structural features of EGF	43
<i>Identification of potential binding sites on EGF</i>	44
Docking studies	46
Biophysical tools to study the interaction of peptides with EGF.....	47

<i>A fluorescence method to assess ligand binding</i>	47
<i>Isothermal titration calorimetry (ITC)</i>	49
<i>NMR methods to probe ligand-EGF interactions</i>	50
<i>SAW for the quantitative analysis of ligand-EGF interactions</i>	53
Discussion	56
CHAPTER 2: DESIGN OF PEPTIDES MIMICKING INTERACTING DOMAINS OF EGFR	61
Structural analysis of the EGF-EGFR interaction.....	63
Mimicry of a 7-mer loop with constrained cyclic peptides.....	65
Mimicry of a 28-mer cyclic domain using designed mini-proteins.....	69
Mimicry of a β -sheet epitope of domain III of EGFR.....	74
Bicyclic peptides as improved analogues of cp28.....	79
<i>First generation bicyclic peptides</i>	79
Computational methods for lead optimization	85
<i>Second generation bicyclic peptides</i>	85
<i>Third generation bicyclic peptides</i>	91
Discussion	95
CHAPTER 3: TOWARDS AN ANTI-EGF PEPTIDE DRUG: BIOLOGICAL ACTIVITY AND ADME PROPERTIES	101
PPI inhibition assay	103
Binding of EGF to cells	106
Cell viability assays.....	110
Stability in human serum	111
Stability in human hepatic microsomes	114
Structural characterization of cp23G	116
Discussion	119
CHAPTER 4: A NEW CLASS OF COVALENT PEPTIDES AS POP INHIBITORS	123
Design based on the structure-activity relationships of POP inhibitors.....	126
Synthesis of the potential inhibitors	128
Determination of POP inhibition.....	129
Kinetic studies on the mode of action.....	130
Docking insights on the binding mode	133

Selectivity versus related peptidases	134
Permeability through brain lipids	135
Discussion	137
GENERAL DISCUSSION.....	139
CONCLUSIONS.....	145
EXPERIMENTAL SECTION.....	151
MATERIALS AND METHODS.....	153
Solvents and reagents	155
Peptide synthesis and characterization	155
Solid-phase peptide synthesis.....	155
Peptide synthesis in solution	160
Purity assessment	161
Identification	161
Quantification by amino acid analysis	162
Structural analysis	163
Computational experiments.....	163
Recombinant human EGF expression.....	166
Fluorescence emission titrations	167
SAW experiments	167
ITC experiments	168
NMR spectroscopy	168
PPI inhibition assay	169
Binding of EGF to cells	170
Cell viability assay	170
Stability in human serum	171
Stability in human hepatic microsomes	171
POP activity assay.....	171
FAP activity assay.....	172
DPP-IV activity assay	173
Parallel Artificial Membrane Permeability Assay (PAMPA).....	173

PRODUCT CHARACTERIZATION	175
Peptides targeting EGF.....	177
Peptidomimetics targeting POP	187
Recombinant EGF protein characterization	191
SAW sensorgrams.....	192
REFERENCES.....	195
SUMMARY IN CATALAN	213

ABBREVIATIONS

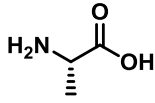
AA	amino acid
ACN	acetonitrile
AcOEt	ethyl acetate
ADME	absorption, distribution, metabolism and excretion
AF488	Alexa Fluor® 488
AMC	7-amino-4-methylcoumarin
ATP	adenosine triphosphate
BBB	blood-brain barrier
BdFB	4-(benzyloxy)-3,5-difluorobenzoyl
Boc	<i>tert</i> -butyloxycarbonyl
CD	circular dichroism
CNS	central nervous system
COMU	(1-cyano-2-ethoxy-2-oxoethylideneaminoxy)dimethylamino-morpholino-carbenium-hexafluorophosphate
COSY	correlation spectroscopy
CRC	colorectal cancer
CSP	chemical shift perturbation
Dbz	3,4-diaminobenzoyl
DCM	dichloromethane
DIC	diisopropylcarbodiimide
DIPEA	N,N-diisopropylethylamine
DMAP	4-dimethylaminophthalimide
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DODT	3,6-dioxa-1,8-octane-dithiol
DPP-IV	dipeptidyl peptidase-IV
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
eq	equivalent
ESI	electrospray ionization
FACS	fluorescence activating cell sorting
FAP	fibroblast activation protein
FDA	Food and Drug Administration
Fmoc	9-fluorenylmethyloxycarbonyl

FPLC	fast protein liquid chromatography
HATU	1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HBSS	Hank's balanced salt solution
HBTU	N,N,N',N'-tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate
HOAt	1-hydroxy-7-azabenzotriazole
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence spectroscopy
IC ₅₀	half maximal inhibitory concentration
IP3	inositol triphosphate
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
LB	lysogeny broth
m/z	mass over charge
mAb	monoclonal antibody
MD	molecular dynamics
MeOH	methanol
MMP	matrix metalloproteinase
MS	mass spectrometry
MW	molecular weight
NCL	native chemical ligation
NHS	N-hydroxysuccinimide
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser spectroscopy
NSCLC	non-small cell lung cancer
OD	optical density
PAMPA	parallel artificial membrane permeability assay
PDA	photodiode array
PDB	Protein Data Bank
PDGF	platelet derived growth factor
PDGFR	platelet derived growth factor receptor
<i>Pe</i>	effective permeability
POP	prolyl oligopeptidase

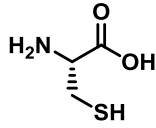
PPI	protein-protein interaction
PyBOP	(benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
REMD	replica exchange molecular dynamics
RMSD	root mean square deviation
rt	room temperature
RTK	receptor tyrosine kinase
SAM	self-assembled monolayer
SAR	structure-activity relationship
SAW	surface acoustic wave
SAXS	small-angle X-ray scattering
SD	standard deviation
SPPS	solid phase peptide synthesis
SUMO	small ubiquitin-like modifier
TBME	<i>tert</i> -butyl methyl ether
TBTU	N,N,N',N'-tetramethyl-O-(benzotriazol-1-yl)uronium tetrafluoroborate
<i>t</i> Bu	<i>tert</i> -butyl
TCEP	tris(2-carboxyethyl)phosphine hydrochloride
TFA	trifluoroacetic acid
TFE	2,2,2-trifluoroethanol
TGF- α	transforming growth factor α
TIS	triisopropylsilane
TOCSY	total correlation spectroscopy
TOF	time-of-flight
Trt	trityl
Trx	thioredoxin
UPLC	ultra-high performance liquid chromatography
UV	ultraviolet
VEGF	vascular endothelial growth factor
VEGFR	vascular endothelial growth factor receptor
ZPP	benzyloxycarbonyl prolyl proline

Proteinogenic amino acids

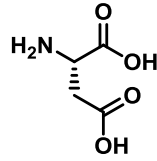
A
Ala
L-Alanine



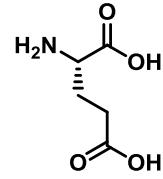
C
Cys
L-Cysteine



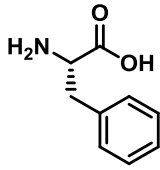
D
Asp
L-Aspartate



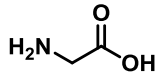
E
Glu
L-Glutamate



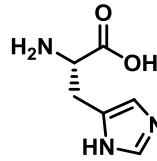
F
Phe
L-Phenylalanine



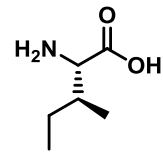
G
Gly
L-Glycine



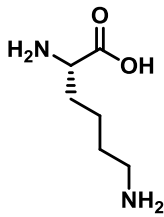
H
His
L-Histidine



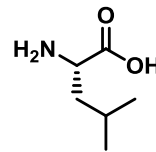
I
Ile
L-Isoleucine



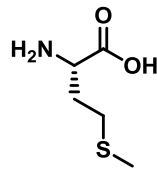
K
Lys
L-Lysine



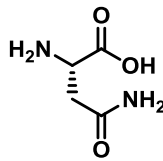
L
Leu
L-Leucine



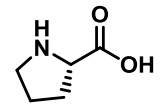
M
Met
L-Methionine



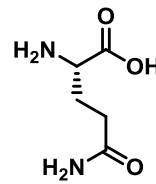
N
Asn
L-Asparagine



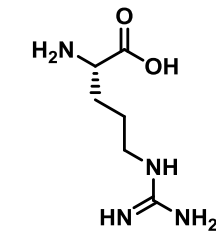
P
Pro
L-Proline



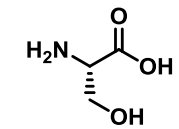
Q
Gln
L-Glutamine



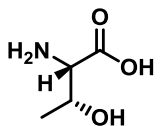
R
Arg
L-Arginine



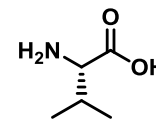
S
Ser
L-Serine



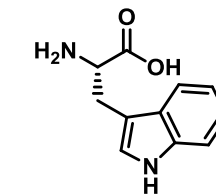
T
Thr
L-Threonine



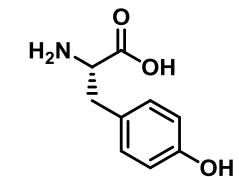
V
Val
L-Valine



W
Trp
L-Tryptophan

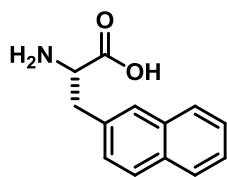


Y
Tyr
L-Tyrosine

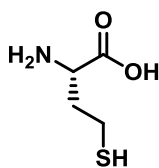


Non-proteinogenic amino acids

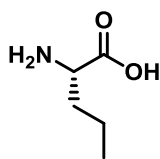
2NaI
L-2-naphthylalanine



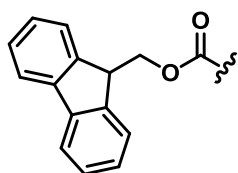
HCys
L-homocysteine



Nva
L-norvaline



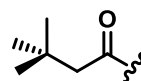
Protecting groups



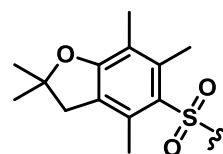
9-fluorenylmethoxycarbonyl
(Fmoc)



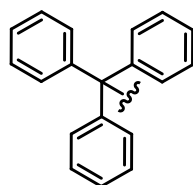
tert-butyl
(tBu)



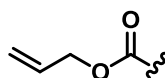
tert-butoxycarbonyl
(Boc)



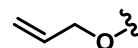
2,2,4,6,7-pentamethyl-
dihydrobenzofuran-5-sulfonyl
(Pbf)



trityl
(Trt)

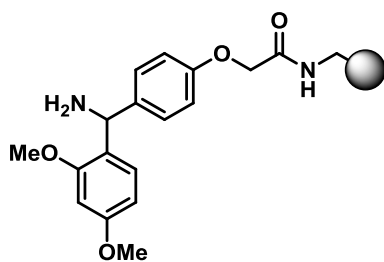


allyloxycarbonyl
(Alloc)

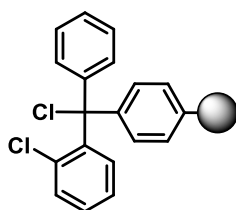


allyl
(OAll)

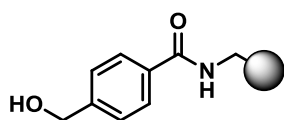
Resins



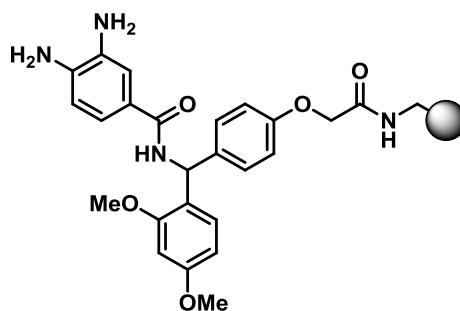
RinkAmide Chemmatrix resin



2-chlorotrityl resin

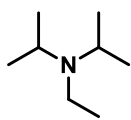


4-(Hydroxymethyl)benzoyl-
aminomethyl (HMBA) resin

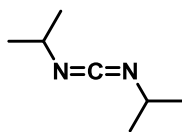


3,4-diaminobenzoyl AM
(Dawson DBZ) resin

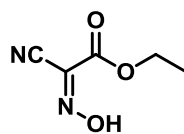
Coupling reagents and additives



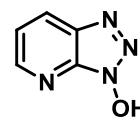
N,N-diisopropyl-
ethylamine
(DIPEA)



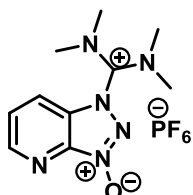
diisopropylcarbodiimide
(DIC)



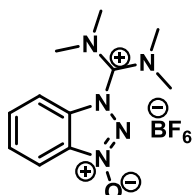
ethyl cyano-
glyoxylate-2-oxime
(Oxima)



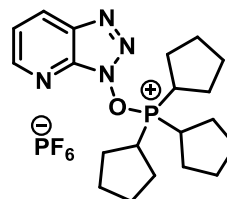
1-hydroxy-7-
azabenzotriazole
(HOAt)



1-[bis(dimethylamino)methylene]-
1H-1,2,3-triazolo[4,5-b]pyridinium 3-
oxid hexafluorophosphate
(HATU)



N,N,N',N'-tetramethyl-O-
(benzotriazol-1-yl)uronium
tetrafluoroborate
(TBTU)



(benzotriazol-1-yloxy)-
tripyrrolidinophosphonium
hexafluorophosphate
(PyBOP)

INTRODUCTION

Modulation of protein-protein interactions (PPIs) using designed peptides

The pharmaceutical industry, together with most research groups in academia, has traditionally focused the development of new drugs on the inhibition of particular receptors and enzymes. These proteins typically feature deep cavities where endogenous ligands bind (i.e. neurotransmitters), and during the second half of the 20th century they became the most popular drug targets. Low molecular weight molecules were designed to compete for the binding to the active pocket, thus acting as inhibitors, and convenient and sensitive screening assays allowed for the rapid optimization of hit compounds.

Although major progress has been made with this strategy (effective drugs having been marketed for dozens of diseases), the genomics and proteomics revolutions that started ~20 years ago have uncovered tens of thousands of new proteins, which constitute the key elements of the regulatory networks in cells and open a new scenario of therapeutic targets.

In human cells, these networks are highly complex and are estimated to have up to 650,000 protein-protein interactions (PPIs).¹ Despite its huge size, only ~15% of the human interactome is currently known,² and there are yet few examples of drugs that are able to successfully modulate PPIs. One of the main reasons for this is that, unlike most conventional targets such as enzymes and receptors, the binding surface (interface) between proteins is relatively large (1,000-3,000 Å²), flat and featureless³. Another reason is that, in most cases, there are no small natural substrates or PPI ligands that can be used as a starting point to design new inhibitors. On top of this, the chemical space of the traditionally explored small molecule inhibitors differs considerably from the larger and more diverse one that PPIs require.⁴

Despite these issues, extensive medicinal-chemistry and SAR (structure-activity relationship) efforts have recently been channeled into the discovery of PPI modulators and, as we will see, some challenging targets considered *undruggable* not long ago have been successfully modulated with a variety of drugs, ranging from small molecules⁵ to peptides⁶ and peptide-like scaffolds (peptidomimetics).⁷

PPIs: typical features

The buried surface area for PPI interfaces for which inhibitors have been discovered varies between 1,000-6,000 Å².⁸ Despite these large sizes, some specific regions of the interface greatly contribute to the interaction energy of a pair of proteins, or a protein and a ligand. These regions are called hotspots.⁹ Alanine scanning analysis of thousands of PPIs has revealed a preference for Trp, Tyr and Arg as hotspot residues. To a lesser extent, the polar residues Asp and His are also enriched.¹⁰ From a structural perspective, we can distinguish between domain-mediated and peptide-mediated PPIs.⁸ In the former, proteins (typically globular-shaped) interact through a discontinuous epitope, in which various regions of the protein sequence, brought into spatial proximity by the protein folding, interact with the partner protein. In contrast, peptide-mediated PPIs are characterized by a linear sequence in one of the proteins, which often adopts a precise secondary structure. A more detailed classification of PPIs, depending on the shape and flexibility of the interacting partners, is shown in *Table 1*. All in all, the range of affinities for PPIs spans about six orders of magnitude, from picomolar (strong and long-lived) to high micromolar (weak and transient).¹¹

The potential of designed peptides for targeting PPIs

As we have mentioned, protein-protein interfaces involve large contact patches, high solvent exposure and a variable degree of flexibility in the side chains and backbone of the interacting partners. These are the main reasons why many of the classical drug discovery tools (i.e. small molecule screening) fail to provide leads. However, peptides are ideal candidates to tackle these limitations, as they are able to closely mimic the structural features of proteins, and at the same time they can be readily obtained in the laboratory by means of modern synthetic techniques that allow access to a diversity of topological modifications.

Compared to small organic molecules, peptides have higher degree of flexibility, which allows them to better adapt to irregular protein surfaces. As peptides are modular structures, great diversity can be readily obtained by changing the amino acid building blocks. The diversity of functional groups that side chains can accommodate translates into enhanced potency and selectivity. Last but not least, especially for therapeutic purposes, peptides are biocompatible: they are degraded to amino acids in the body and they accumulate only to a minor extent in tissues, which means they

show low toxicity.⁸ Nevertheless, not all that glitters is gold, and peptides have several drawbacks. In this regard, they tend to show low stability to proteolytic degradation by digestive and plasmatic enzymes, high hepatic/renal clearance and poor ability to cross physiological barriers.

As a result of these limitations, researchers have envisaged a wide range of artificial modifications through which to achieve better serum stability and cell membrane permeability. An obvious place to start the search for drug candidates, particularly in peptide-mediated interactions, is by identifying the interacting peptide of one of the binding partners at the interface.


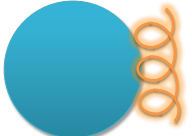

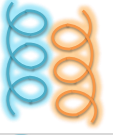

PPI class	Description	Sketch	Examples
Protein-peptide (continuous epitope)	A continuous epitope (β -strands and loops) binding to a protein surface		<ul style="list-style-type: none"> • HIV integrase-LEDGF • Integrins • RAD51-BRCA2 • PDZ domains
Protein-peptide (discontinuous epitope)	A helical domain (discontinuous epitope) binding into a groove		<ul style="list-style-type: none"> • MDM2-p53 • BCL-XL-BAK • ZipA-FtsZ • MCL1-BH3
Protein-peptide (anchor residue)	A single residue on the peptide acts as an anchor group dominating the interaction		<ul style="list-style-type: none"> • Bromodomains • PDEδ-KRAS • SH2 domains • VHL-HIF1α
Peptide-peptide	A pair of helical motifs form an elongated binding interface		<ul style="list-style-type: none"> • MYC-MAX • NEMO-IKK • Annexin II-P11
Globular protein-globular protein	Two proteins often presenting discontinuous epitopes		<ul style="list-style-type: none"> • IL-2-IL-2R • TNF-TNF • EGF-EGFR

Table 1. Structural classification of protein-protein interactions and examples. Adapted from Scott, D. et al.¹²

In this context, the use of linear peptide segments as PPI modulators has a major disadvantage, namely that their high degree of flexibility results in negligible or random structure, thereby severely impairing recognition by a well-structured protein target.¹³ This gains relevance in domain-mediated interactions, where hotspots are more scattered over the interface and only a constrained peptide that is already pre-organized into its bioactive conformation will efficiently bind to the interface.¹⁴

Cyclic peptides to mimic protein structural motifs

The introduction of non-natural amino acids and cyclic constraints into the peptide structures is probably the strategy most widely exploited to improve the structural and functional mimicry of protein surfaces.¹³ In most cases, ring structures are formed by disulfide bridges between cysteine residues. Other types of cyclization are head-to-tail (the ring is formed via an amide bond between the N and C termini), side-chain-to-side-chain (between two side-chain functional groups), and side-chain-to-head or to-tail (*Figure 1*).

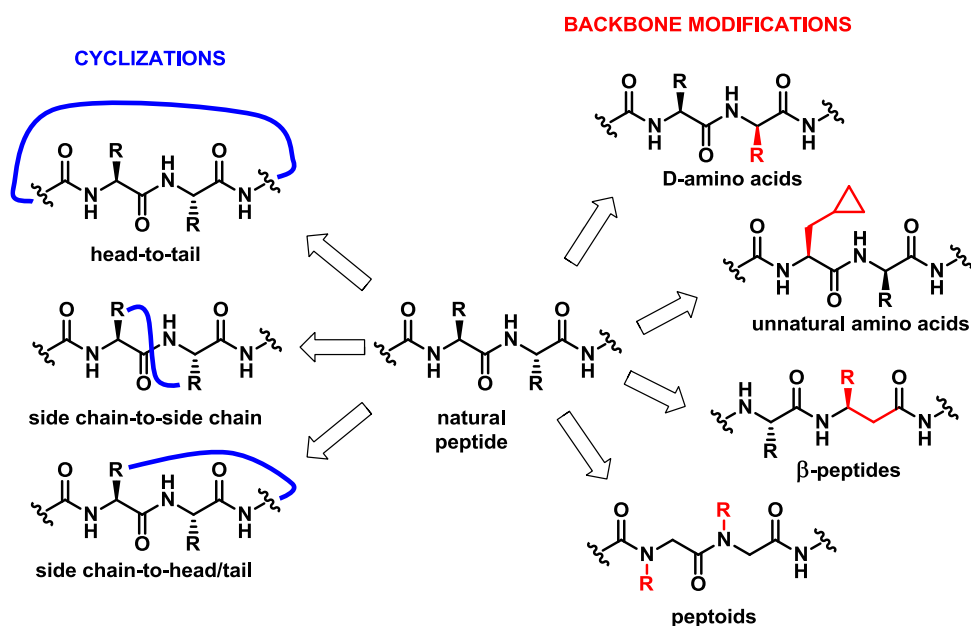


Figure 1. Main structural modification types performed on peptides for targeting protein-protein interactions. Adapted from Wójcik, P. et al.¹⁵

In PPIs, a loop in one of the interacting partners often makes the greatest contribution to the molecular recognition event. In these cases, loops with a ring-like structure cannot be targeted by small molecules or by linear peptides, which have to pay a high entropic price because of their flexibility. Indeed, a recent study demonstrated that some PPIs can be targeted only by cyclic molecules, which favor strong binding to protein targets that otherwise may be considered *undruggable*.¹⁶

In most documented PPIs, the interacting peptide in one of the partners adopts a specific secondary structure, most usually an α -helix.¹⁷ In this situation, the stability of the secondary structure is closely related to its ability to interact with the target protein. Thus, rigidifying the peptide structure by cyclization is one of the main strategies to narrow the flexibility of these molecules and constrain their folding into well-defined topologies.

Amongst secondary structures, α -helices play a major role in PPIs. A systematic analysis of PPIs in the PDB found that 62% of protein complexes feature an α -helix at their interface.¹⁸ In these PPIs, the binding energy is largely derived from the hydrophobic interactions of the side chains of residues on one face of the helix with the partner protein. α -helix-mediated PPIs can be classified into the following three types: (a) proteins with a well-defined cleft for helix binding, such as p53/HDM2 and Bcl-xL/Bak; (b) extended interfaces mediated by a number of helical motifs, which often involve more than one side of the helix, such as hexameric gp41; and (c) low-affinity transient interactions lacking specific hotspots.

In such a context, a designed peptide ligand that successfully reproduces the key features of the helix is likely to act as a competitive inhibitor of the interaction. This has proven particularly successful for the first type of helix-mediated PPIs, as reflected by the increasing number of publications of cyclic helical peptides that target therapeutically relevant PPIs.¹⁹ A wide range of stapling techniques (usually through the side chains of two residues) has been developed to promote the helical folding of a peptide. However, no single technique is universally superior to the others, thus the choice depends on each particular case and a number of factors (such as ease of synthesis, permeability and solubility of the final peptide, and target affinity).¹³

Although not as common as α -helices, β -sheets are also involved in PPIs and protein-DNA interactions. For instance, many proteases (e.g. HIV-1 protease) work as dimers that are held together by the formation of intermolecular β -sheets. Likewise, β -sheet formation is involved in some oncogenic PPIs, like the Raf-Rap1 interaction that leads to tumor cell proliferation. Despite presenting a greater challenge than helical structures, peptides mimicking β -strands show a great potential as therapeutic agents. Cyclic peptides, in particular, can efficiently orientate the side chains in a β -strand pattern. In these structures, artificial turns and template units are introduced to nucleate and stabilize the β -strands in either a parallel or anti-parallel fashion, depending on the target protein of interest.

Over the past 20 years, there has been an increase in research efforts seeking to develop PPI inhibitors, and, for the reasons previously highlighted, peptides have emerged as privileged structures through which to modulate PPIs. Peptide cyclization has been successful in providing chemical tools for targets that were previously considered *undruggable*. In this regard, the first peptide PPI inhibitors have entered clinical trials (see

Table 2).¹² However, there are challenges that need to be addressed, such as the poor pharmacokinetics and ADME properties inherent to all types of peptide ligands when used *in vivo*. With PPIs becoming the focus of mainstream drug discovery projects, cyclic constrained peptides have the properties required to deliver major advances in terms of new drugs for the treatment of disease.

Cyclic peptides that modulate PPIs with anti-cancer activity

A dual MDM2 and MDMX stapled peptide for p53-dependent cancer therapy

The p53 tumor suppressor is a potent inducer of cell cycle arrest, apoptosis, senescence, and innate immunity.²⁰ It is activated in response to oncogenic transformation, extrinsic stress, and viral infection to protect higher organisms from cancer. In healthy cells, p53 activity is kept at minimal levels by interactions with MDM2 and MDMX, both of them increasing ubiquitin ligase activity, thus leading to proteasomal degradation of p53. Genetic or functional inactivation of p53 is an obligatory step during cancer development.

In human tumors that retain wild-type p53, overexpression of MDM2 or MDMX serves as an alternative mechanism of p53 inactivation.²⁰ As shown by the first MDM2-p53 disruptors (the Nutlin family of compounds), the inhibition of MDM2 is compensated by the activation of MDMX. Therefore, simultaneous targeting of MDM2 and MDMX is needed to achieve optimal activation of p53.²¹ Although both MDM2 and MDMX bind to an α -helix motif in p53, there is sufficient diversity in their binding regions to make the development of small-molecule dual antagonists challenging.

With this information in hand, Chen *et al.* used phage display to identify a linear peptide (pDI) that inhibited p53 interactions with MDM2 (IC₅₀ = 44 nM) and MDMX (IC₅₀ = 550 nM). This peptide was 300-fold more potent *in vitro* than the equivalent p53 peptide.²² The pDI peptide (LTFEHYWAQLTS) is distinct from the p53 peptide (16-QETFSDLWKLLP-28), but retains three key p53 hydrophobic residues (Phe19, Trp23, and Leu26, underlined) that are important hotspots for the interaction.

Name	Therapeutic area	Clinical phase	Developer
Aplidin®	Oncology	III	PharmaMar
Debio025	Infectious diseases	III	Novartis
MK4261	Infectious diseases	III	Merck
PT141	Hematology	III	Palatin Technologies
ALRN6924	Oncology	II	Aileron Therapeutics
APL1	Ophthalmology	II	University of Pennsylvania & Alcon Laboratories
APL2	Hematology	II	University of Pennsylvania & Potentia
ASP3291	Gastroenterology	II	Astellas Pharma
AT1001	Metabolic disorders	II	Alba Therapeutics
AZP531	Cardiovascular	II	Eli Lilly
MEN11420	Gastroenterology	II	Menarini
PL3994	Cardiovascular	II	Palatin Technologies
POL6326	Immunology	II	Polyphor
RG7929	Infectious diseases	II	Roche
SCY635	Infectious diseases	II	Waterstone Pharmaceuticals
Voclera	Dermatology	II	Roche
Vosoritide	Metabolic disorders	II	BioMarin
ALRN5281	Metabolic disorders	I	Aileron Therapeutics
OBP801	Oncology	I	Astellas Pharma
RA101495	Hematology	I	Ra Pharmaceuticals

Table 2. List of cyclic peptides ongoing phase I to phase III clinical development (retrieved from Medtrack, February 23, 2016).

In the continuation of this work, they applied a peptide stapling approach to create a macrocyclic analogue of the pDI peptide with the main goal of restraining its α -helical bioactive conformation.²³ The resulting $i, i+7$ stapled peptide displayed a 10-fold increased inhibitory potency over the linear parent peptide against both MDM2 and MDMX. However, the introduction of the hydrophobic staple (in the place of Glu and Thr residues) resulted in low aqueous solubility, thereby limiting its biochemical applications. A rational design process, based on the crystal structures of MDM2 and MDMX in complex with the peptide, allowed the modification of Leu26 by cyclobutylalanine (Cba), the incorporation of Glu21 and of two Ala residues at the C terminus, resulting in peptide ATSP-7041 (Ac-LTF-cyclo[R8-EYWAQ(Cba)-S5]-SAA-NH₂, see *Figure 2*).

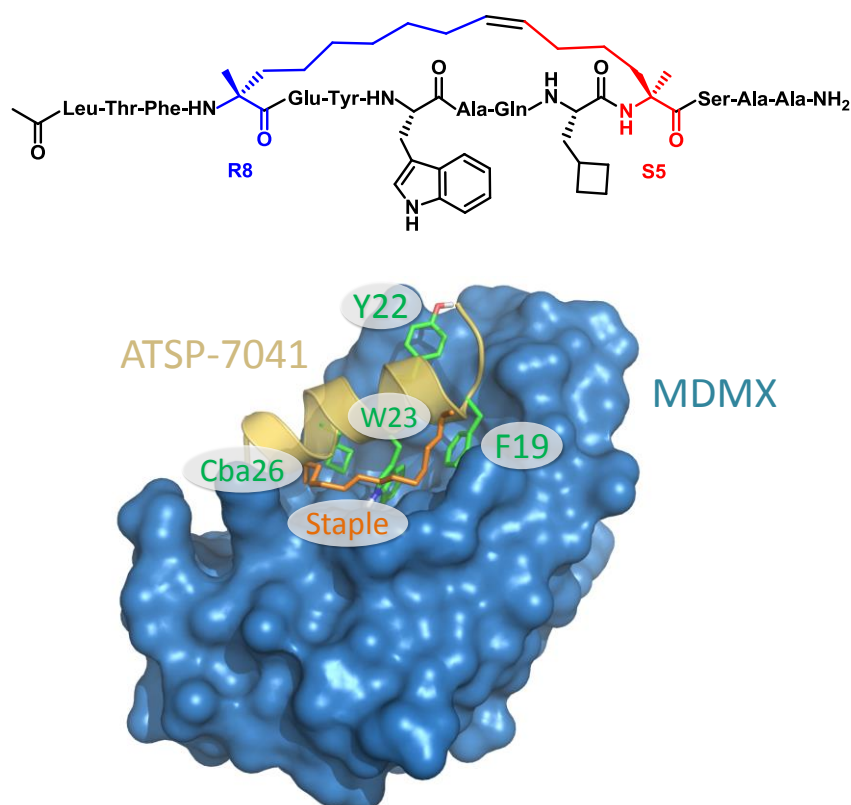


Figure 2. The chemical structure of ATSP-7041 is shown, highlighting the R8 (blue) and S5 (red) amino acids that form the 11-carbon staple (top). A high-resolution X-ray structure of ATSP-7041 bound to MDMX delineates its intermolecular contacts with the target protein (PDB code: 4N5T). The side chain is shown in green: Phe19, Trp23, Cba26, and Tyr22 are the main interacting residues responsible for the high binding affinity, whereas the staple olefin, also contributing to the binding affinity, is shown in orange.

This stapled peptide binds both MDM2 ($K_d = 0.9$ nM) and MDMX ($K_d = 6.8$ nM) with low nanomolar affinities, featuring a conformation in which the well-known triad (F19, W23, and Cba26) perfectly matches the binding mode of p53; whereas the staple and the additional Y22 establish further interactions with MDMX (*Figure 2*). ATSP-7041 shows efficient cell penetration and strong efficacy in multiple human cancer cell lines, as well as in *in vivo* xenograft models. In MCF7 cells, it led to an elevation of p53 transcriptional targets (MDM2 and p21), an effect that persisted significantly longer following drug removal than that of RG7112, a phase I clinical candidate of the Nutlin family. Upon intravenous administration, ATSP-7041 exhibited low clearance and long plasma half-life in models of mouse, rat and monkey, as well as broad organ distribution. Currently, this stapled peptide, licensed as ALRN-6924 by Aileron Therapeutics, is undergoing phase II clinical trials in p53-wild type advanced hematologic and solid malignancies (*Table II*).²¹

All in all, this α -helical stapled peptide is a highly potent dual inhibitor of MDM2 and MDMX that is able to reactivate the p53 pathway in a mechanism-dependent manner. Its development can be seen as a highly relevant proof-of-concept of stapled peptides targeting intracellular PPIs, thus offering viable new modalities for cancer therapy.

Cyclic α/β -peptides that target the VEGF-VEGFR interaction

There are many examples of PPIs in which, as opposed to the example above, there is no specific well-structured peptide on the interface. Instead, this type of domain-mediated interactions, typically involving globular proteins, feature large and topologically irregular surfaces with hotspot residues scattered over the interface. These types of PPIs are highly demanding to target, as they cannot be effectively engaged by a single α -helix (or β -strand) mimic, which would be too narrow to provide sufficient contact area for strong binding. In this example,²⁴ an extracellular domain-mediated interaction is inhibited by using a helix-loop-helix cyclic-constrained mini-protein.²⁵

The vascular endothelial growth factor (VEGF) is a crucial regulator of vascular development in inflammation, tissue regeneration and angiogenesis, sharing regulatory mechanisms with other well-characterized receptor tyrosine kinases (RTKs), such as PDGFR and EGFR.²⁶ Pathological angiogenesis is one of the hallmarks of cancer, and several anti-VEGF therapies are currently used to treat various forms of cancer. VEGF is an extracellular homodimeric protein that binds to cell-

surface receptors (VEGFR), triggering downstream signaling that promotes the proliferation and migration of endothelial cells. The receptor binding on VEGF is large ($> 800 \text{ \AA}^2$) and lacks *druggable* cavities or pockets. Only anti-VEGF mAbs (such as bevacizumab) have been able to effectively disrupt this PPI; all small molecules designed so far have yielded poor results.²⁶

In 2011, Deshayes *et al.* reported a 59-residue Z-domain (Z-VEGF) that was able to bind to the receptor-recognition site of VEGF with low micromolar affinity ($K_i = 0.41 \text{ \mu M}$) and high selectivity.²⁷ The Z-domain scaffold is a type of antibody mimetic that features a three-helix-bundle tertiary structure, which has been used for developing ligands for specific targets via screening of phage-display libraries, in which residues on helix 1 and 2 are varied. Despite the potent binding affinity, Z-VEGF displayed poor drug-like properties for further therapeutic development, such as high susceptibility to proteolytic degradation.

Based on the crystal structure of Z-VEGF in complex with VEGF, Checco & Gellman shortened the overall length of this mini-protein by removing the non-interacting helix 3 and introducing instead an interhelical disulfide cross-link in order to stabilize the helix-loop-helix structure.²⁸ This disulfide bridge is crucial for maintaining the bioactive conformation, its removal abolishing the interaction with VEGF. Next, they performed an iterative design strategy to introduce several β -amino acid residues in the helical domains, without compromising the affinity of the resulting peptide for VEGF (*Figure 3*).²⁴ It is known that peptides containing β -amino acid residues interspersed among α -residues (α/β -peptides) are good mimics of the recognition surface projected by an α -helix, and at the same time the resulting unnatural backbone hinders proteolytic degradation. Their extensive design effort culminated in the development of α/β -VEGF-1, a 39-mer α/β -peptide that displayed a K_i value of 0.11 \mu M for VEGF and, more importantly, conferred a >400 -fold increase to proteolytic degradation, relative to Z-VEGF.

Furthermore, their design strategy of Z-domain-mimetic α/β -peptides was validated in two other PPI systems, namely IgG and TNF- α , thus confirming its efficacy to generate bio-stable cyclic α/β -peptides that bind tightly and specifically to targets of biomedical interest.²⁸ Such scaffolds are very useful for targeting large and flat PPIs and may find application as diagnostic and therapeutic tools.

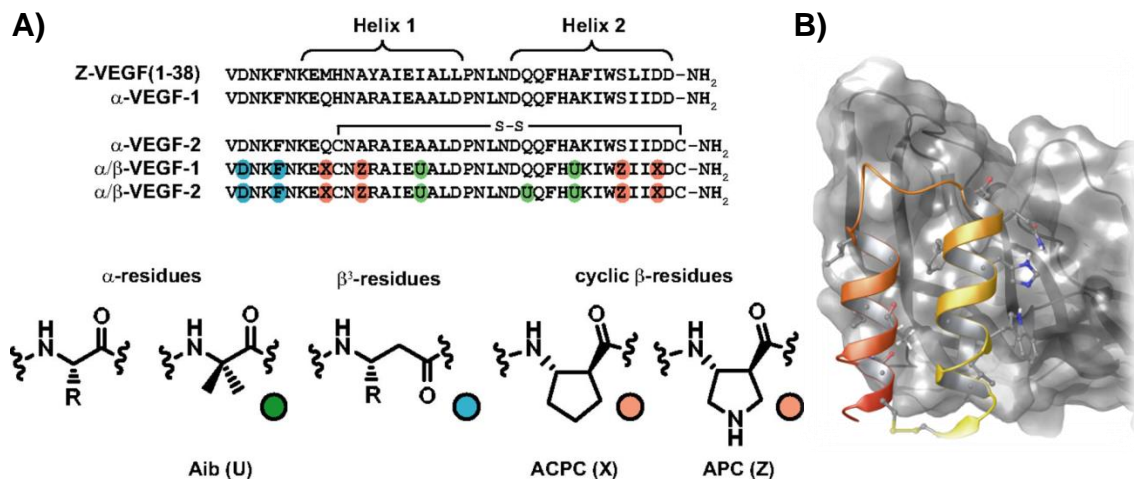


Figure 3. A) Sequences of α/β -VEGF-1 and other analogues derived from the three-helix Z-VEGF mini-protein. Nonnatural residues are indicated by colored circles, and its structure is depicted below. Extracted from Checco et al.²⁴ B) Crystal structure of α/β -VEGF-1 in complex with VEGF (right). The side chains of the peptide that appear to make contact with VEGF, as well as the disulfide bond, are shown as sticks (PDB code: 4WPB).

The EGF-EGFR pathway: from oncogene discovery to targeted cancer therapeutics

The EGFR signaling pathway in cancer

By studying neuronal differentiation in mice, Cohen and co-workers first identified, in 1962,²⁹ a growth factor that they termed epidermal growth factor (EGF) because it stimulated the proliferation of epithelial cells in skin and cornea. In the following years, Cohen purified and characterized the sequence of human EGF,³⁰ and in 1986 he was awarded the Nobel Prize in Medicine for his contributions to the discoveries of growth factors. Later, it was demonstrated that EGF exerts its action through binding to a cell receptor (EGFR) and activation of its tyrosine kinase domain.³¹

Cancer cells can acquire the capacity for self-proliferation through the uncontrolled production of specific molecules that promote cell growth (i.e. growth factors) or through increased expression of certain receptors (i.e. growth factor receptors). These molecular events eventually generate intracellular signals that cause tumor cells to proliferate and penetrate blood or lymphatic vessels, thereby allowing their circulation through the intravascular stream and colonization of new tissues, a process known as metastasis. This ability of cancer cells to spread to distant organs and impair their normal function is what makes cancer a life-threatening disease.

EGFR was first considered to be an oncogene because of its homology to v-ErbB, a retroviral protein that enables the avian erythroblastosis virus to transform chicken cells.³² This similarity attracted the attention of researchers, who demonstrated the cell-transforming capacity of EGFR in animal models. Certainly, EGFR gene amplification was the first recognized oncogenic driver in human epithelial cancer,³³ and EGFR overactivation is found in many malignancies, such as colorectal cancer (CRC), non-small-cell lung cancer (NSCLC) and head and neck cancers.³⁴ Nonetheless, under physiological conditions, EGFR function is essential for early neuronal, mammary gland and myocardium development, amongst others. Not surprisingly, EGFR knockout mice suffer from skin, lung and brain defects, dying at early stages of embryonic development.³⁵

In a tumor setting, however, the EGFR pathway becomes hyperactivated by various mechanisms, such as the overexpression of the receptor itself,

the overproduction of EGFR ligands, and/or the appearance of oncogenic mutations that constitutively activate the receptor (*Table 3*).³⁶ No matter the cause, EGFR activation is related to cell cycle progression and activation of anti-apoptotic mechanisms, which are responsible for cancer cell proliferation, survival, invasion, metastasis, and angiogenesis.³⁷ In certain types of cancer, EGFR represents the paradigm of “oncogene addiction”, in which cancer cells become dependent on a specific oncogenic signaling pathway, and drugs that are able to turn off this pathway lead to tumor cell death through apoptotic mechanisms.³⁸

Tumor type	EGFR overexpression	EGFR mutations
Breast	14-91	78
Colorectal	25-77	50-90
Glioblastoma	40-60	NA
NSCLC	40-80	85-100
Head and neck	80-100	NA
Ovarian	35-70	55-100
Pancreatic	30-50	95
Prostate	41-100	NA

Table 3. Incidence of EGFR overexpression and/or dysregulation in human tumors. Adapted from Arteaga, C. et al.³⁹

The ErbB family of cell-surface receptors and their ligands

EGFR is the prototypical member of the ErbB family of cell receptors, which consists of four members in mammals that appear to have diverged from a common ancestral receptor, namely EGFR (or ErbB1), ErbB2/HER2, ErbB3/HER3 and ErbB4/HER4. Binding of peptide ligands to the extracellular domain of EGFR, ErbB3 and ErbB4 induces the homo- or hetero-dimerization of the receptor and its activation.⁴⁰ ErbB2, in contrast, does not bind any extracellular ligand and exists in an extended conformation that resembles a ligand-activated state, thus favoring dimerization with other partners of the ErbB family.⁴¹ All four members of the ErbB family share an analogous structure, which consists of an extracellular ligand-binding domain, a single hydrophobic transmembrane region, and an intracellular segment that holds a highly conserved tyrosine kinase domain (*Figure 4*).⁴²

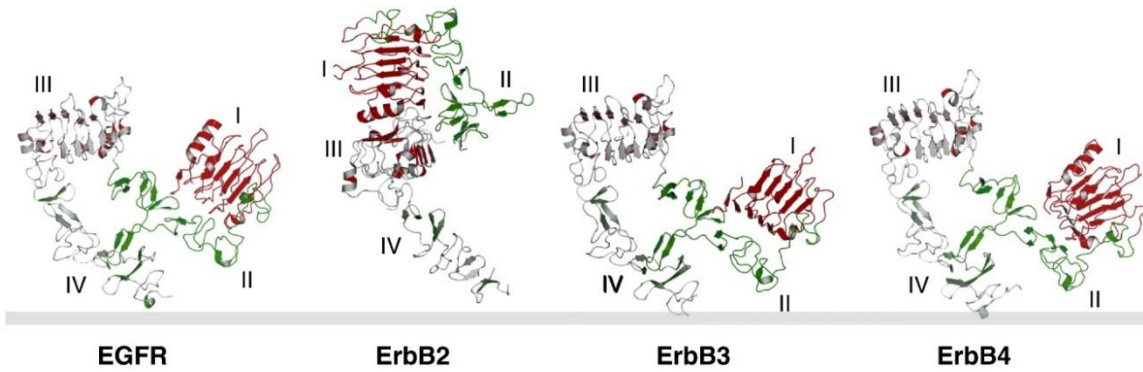


Figure 4. Structures of all four ErbB extracellular regions without bound ligand. EGFR, ErbB3 and ErbB4 all adopt tethered (inactive) conformations in the absence of ligand, whereas the “ligandless” ErbB2 adopts an extended active conformation. Extracted from Lemmon, M.⁴³

Up to 13 ErbB receptor ligands have been described to date, including EGF, TGF- α and amphiregulin, among others, and they show specific binding affinities to each receptor type.⁴⁴ These peptide growth factors are expressed as membrane proteins and are cleaved by metalloproteinases to release the soluble mature ligand into the extracellular space, a process that is upregulated in some tumors. All family members contain one or more repeats of the conserved amino acid sequence CX₇CX₄₋₅CX₁₀₋₁₃CXCX₈GXRC, where X represents any amino acid.⁴⁵ This sequence contains six Cys residues that form three intramolecular disulfide bonds. Disulfide bond formation generates a globular structure with three structural loops that are essential for high-affinity for ErbB receptors (*Figure 5*).⁴⁵ Among all the growth factor family, EGF is the one with the highest affinity for EGFR⁴⁶ (*Table 4*) and has been repeatedly identified to contribute to the progression of several types of cancer.³⁶

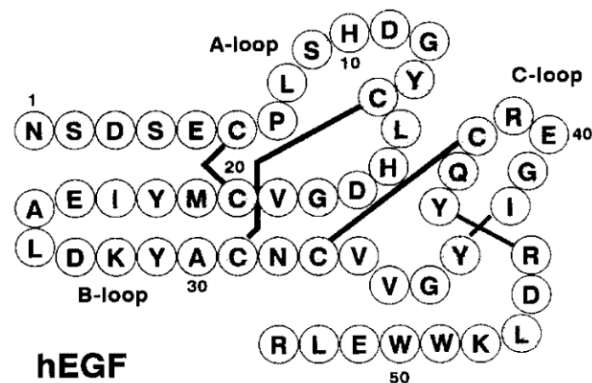


Figure 5. Primary sequence of human EGF. Note that the six Cys residues, together with adjacent Gly that form the A, B and C loops are generally conserved between all EGF-like proteins. Adapted from Van Zoelen, E. et al.⁴⁷

Ligands	ErbB1	ErbB2	ErbB3	ErbB4
EGF	1.9 nM			
TGF- α	9.2 nM			
Amphiregulin	NA			
HB-EGF	7.1 nM			
Betacellulin	1.4 nM			3.6 nM
Epigen				NA
Epiregulin	2800 nM			110 nM
Neuregulin-1			190 nM	174 nM
Neuregulin-2			NA	56 nM
Neuregulin-3				2400 nM
Neuregulin-4				NA
Tomoregulin				NA
Neuroglycan C			NA	

Table 4. Summary of relative IC_{50} for reported ErbB ligands. High affinity ligands are colored green and low affinity ligands yellow. Note that the binding affinity values can strongly depend on specific homo- or hetero-dimeric ErbB combinations. NA = not available. Adapted from Jones, J. et al.⁴⁸

Basic insights into EGFR structure and dynamics

Like all tyrosine-kinase receptors, the human EGFR comprises a large extracellular region, a single transmembrane helical domain, an intracellular juxtamembrane region, a tyrosine kinase domain, and a C-terminal regulatory region. Binding of EGF and other ligands occurs on the extracellular region of EGFR, which is divided into four domains: I, II, III and IV. Domains I and III share about 37% sequence identity, while domains II and IV are cysteine-rich.⁴⁹ Domains I and III participate in receptor-ligand interactions; however, most of the binding energy is contributed by domain III alone, having an affinity of approximately 400 nM for EGF.⁵⁰

In domain II, a β -hairpin motif, commonly referred to as the dimerization arm, has been found to play a key role in the stabilization of the dimeric conformation of the receptor. In the tethered (inactive) state, this motif is buried by intramolecular interactions with domain IV,⁵¹ but ligand binding induces the exposure of this dimerization arm, which makes extensive contacts with domain II of its binding partner, as observed in the X-ray structures of the dimeric complex.⁵² In addition to this motif, the trans- and juxta-membrane domains also stabilize receptor dimerization (*Figure 6*).

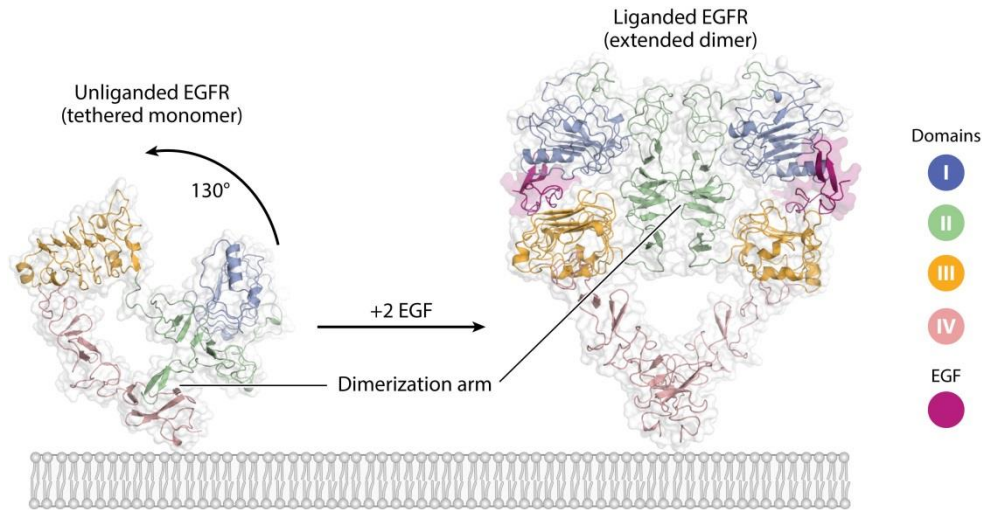


Figure 6. Ligand (EGF) binding induces a $\sim 130^\circ$ rotation of domains I and II with respect to domains III and IV, thereby converting the extracellular module from a tethered conformation to an extended one and generating a “back-to-back” dimer configuration in which the ligand is trapped between domains I and III of each subunit of the dimer. Adapted from Kovacs, E. et al.⁵³

As a whole, EGFR is a highly dynamic protein that explores a range of conformations that coexist in equilibrium. Contrary to other receptor-activating mechanisms, EGF (or other EGFR ligands) makes no direct contribution to the dimerization interface, but instead promotes a massive conformational change in the receptor structure. The dynamics of EGFR have been extensively studied using crystallography,⁵² NMR,⁵⁴ SAXS,⁵⁵ electron microscopy,⁵⁶ and molecular dynamics simulations,⁵⁷ but the detailed mechanism by which the EGF activating signal (on the extracellular domain) is transmitted across the cell membrane to the kinase domain is not fully understood. This is mainly due to the technical difficulty of crystallizing single-pass transmembrane proteins, of which few, if any, structures are currently available.

According to the model of growth factor activation (induced fit), ligand binding would promote receptor dimerization and activation of intracellular kinase activity;⁵⁸ other models (conformational selection) suggest that receptors are pre-dimerized on the cell surface and that ligand binding alters the equilibrium between inactive and active dimeric states.⁵⁹ Another aspect of the binding of EGF to EGFR that has puzzled researchers for years was the identification of two receptor populations in cells, as shown by non-linear Scatchard plots (as would be expected for independent binding sites on the receptor). This reveals a small population of EGFR ($\sim 2\text{-}5\%$) with high affinity for EGF ($K_d \sim 50$ pM) and a predominant ($\sim 95\%$) lower affinity population ($K_d \sim 3$ nM).⁶⁰ As an explanation of this observation, it has been

proposed that EGF binds with high affinity to preformed dimers, but with low affinity to the monomeric inactive forms. In any case, receptor dimerization brings the intracellular domains into close proximity and promotes the association of the kinase domains. In the resulting dimer, the kinase domains do not transphosphorylate each other, in contrast to what was first hypothesized. Instead, the two domains interact in an asymmetric fashion so that one, termed the activator, switches on the other, termed the receiver, in an allosteric mechanism analogous to CDK activation by cyclins.⁶¹

Downstream signaling and cellular responses

Once the receptor is activated, the intracellular phosphotyrosine residues act as docking sites for effector proteins containing SH2 and/or PTB domains, triggering intracellular signaling pathways that include the Ras/Raf mitogen- activated protein kinase (MAPK), the phosphatidylinositol 3-kinase (PI3K/AKT), the signal transducer and activator of transcription (STAT), and the phospholipase C/protein kinase C (PKC).³⁷ The activated EGFR is then internalized through clathrin-mediated endocytosis and sorted at the endosome. Here, the fate of the receptor has important consequences: the receptor recycling pathway favors cell proliferation, whereas the degradation pathway correlates with normal cellular homeostasis.⁶²

Early research on the EGFR pathway gave rise to the concept of linear signaling, which is initiated by ligand-bound active receptors and propagated vertically downstream through several effectors. However, a paradigm shift occurred when the siblings of EGFR (ErbB2, ErbB3 and ErbB4), as well as other tyrosine kinase receptors, were discovered. From a systems biology point of view, the EGFR signaling pathway is described to have a “bow-tie” or “hourglass” structure (see *Figure 7*),⁶³ in which diverse inputs (growth factors) and outputs (cell proliferation, differentiation, etc.) are linked through a conserved processing core (the four ErbB receptors).

In addition to EGFR, other types of receptor tyrosine kinases (such as VEGFR or PDGFR) can also activate common nodes on the signaling cascade, giving rise to a highly interconnected and complex network.⁶⁴ Thus, there is substantial redundancy and cross talk between different signaling pathways, a distinctive feature, shaped by evolution, which confers robustness to the system.⁶⁵ As we will see in the coming sections, these related pathways are one of the main sources of drug resistance, in the

sense that prosurvival stimuli are diverted to complementary pathways when a specific node of the network is blocked.

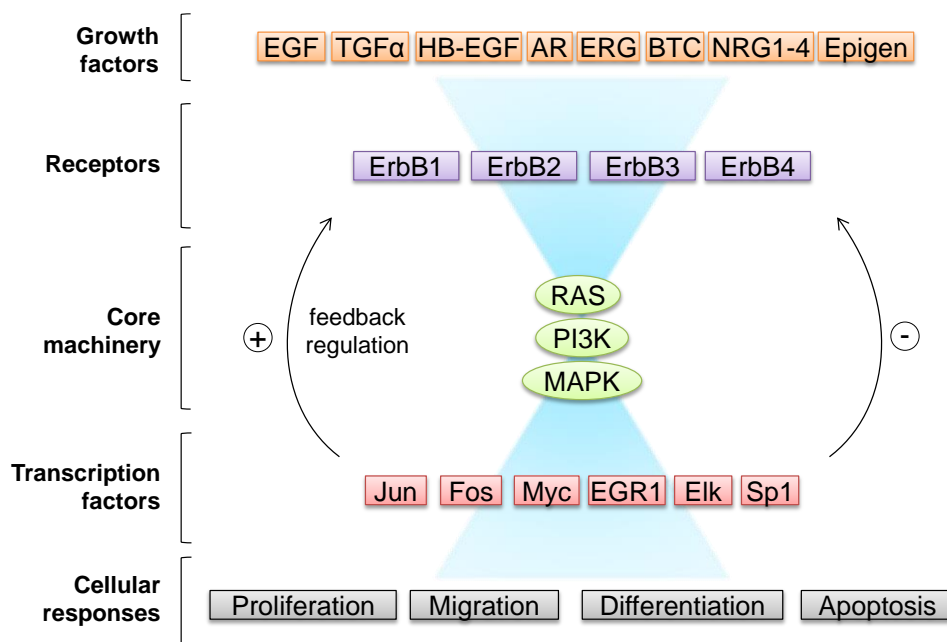


Figure 7. The elements of the EGFR signaling pathway are intertwined in a complex “bow tie”-shaped network, in which wide input layer includes multiple growth factors and RTKs, all of which influence a relatively small number of “core processes”. The behavior of the core processes is read out by a wide output layer that consists of transcription factors that dictate cellular responses. Feedback processes within the core elements are crucial for the underlying mechanisms that generate drug resistance. Adapted from Citri, A. & Yarden, Y.⁶³

Development of EGFR antagonists for cancer treatment

In 1980, even before Cohen discovered the kinase activity of EGFR, Mendelsohn and Sato postulated that antibodies against EGFR could prevent ligand binding and inhibit activation and cancer cell proliferation. Based on this hypothesis, interest in anti-EGFR treatments for specific tumors has led to the development of monoclonal antibodies (mAbs). A humanized version of one of such murine antibodies, named M225, showed anti-proliferative activity *in vivo* and progressed into clinical trials.³⁴ This chimeric antibody was approved by the FDA in 2004 (under the name of cetuximab) for the treatment of metastatic CRC and head/neck cancer, and its clinical use has been expanded to NSCLC.⁴⁹ Cetuximab targets the extracellular domain of EGFR (see *Figure 8*) with a higher affinity ($K_d = 0.39$ nM)⁶⁶ than its natural ligands ($K_d = 1.9$ nM for EGF), thus preventing their binding and the receptor activation. In addition, it promotes receptor internalization and antibody-mediated cytotoxicity.³⁸ Cetuximab has shown efficacy in CRC patients, especially when used in combined chemotherapy,

although in a small subset of patients, an acquired mutation in the extracellular domain of EGFR (S492R) abrogates its binding, leading to clinical resistance.

In 2006, a fully humanized, high affinity ($K_d = 50$ pM), EGFR-targeted antibody (panitumumab) was approved by the FDA, showing activity in a variety of advanced cancer patients, including renal carcinoma and metastatic CRC.⁴⁹ Moreover, tumor cells with the S492R mutation, which confers resistance to cetuximab, still respond to panitumumab treatment, a difference in susceptibility that is possibly due to the different binding epitopes on EGFR.⁶⁷ Both antibodies were the first to demonstrate KRAS, a protein downstream the EGFR pathway, as an effective predictive biomarker of resistance. As a result, KRAS screening is currently used to plan appropriate treatment regimen, as only patients with wild-type KRAS benefit from either cetuximab or panitumumab treatment.⁴⁹

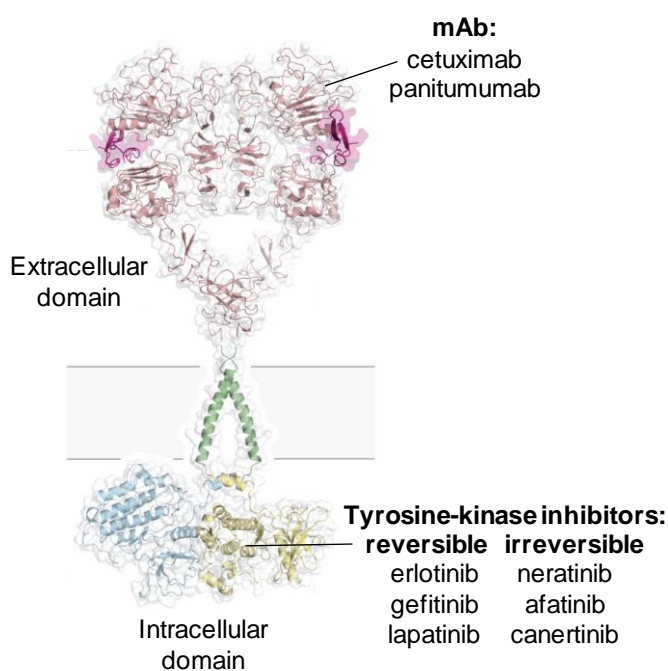


Figure 8. Overview of currently approved anti-EGFR drugs and their site of action on EGFR.

A second class of agents targeting EGFR are tyrosine-kinase inhibitors (TKIs). These molecules reversibly compete for the intracellular ATP binding site and thus block EGFR-induced downstream signaling activation (Figure 8). In 1996, researchers at Astra-Zeneca first discovered a family of 4-anilinoquinazolines that caused a decrease in EGF-stimulated human tumor cell growth *in vitro*.⁶⁸ Their lead compound, gefitinib, received FDA approval in 2003 for the treatment of patients with advanced NSCLC.

Subsequently, more potent TKIs, such as erlotinib, were developed, as were others having an irreversible mode of action (neratinib, afatinib) or a dual mechanism, such as lapatinib, which inhibits both ErbB1 and ErbB2.

After years of clinical experience, it is widely accepted that EGFR mutation status determines the tumor response to these TKIs, a significant discovery meaning that frequent EGFR-activating mutations (i.e. exon 19 deletions and L858R, which account for 85% of all EGFR mutations) predict sensitivity to TKIs such as gefitinib, erlotinib and afatinib. Most of these therapeutic agents are typically not used alone but in combination with cytotoxic chemotherapy or with other targeted therapies that enhance the clinical efficacy and prevent or aim to overcome drug resistance.

Another possible strategy to inhibit EGFR activation is by directly targeting the other partner of this specific PPI, namely EGF. However, like other cases of success with therapeutics that target extracellular growth factors (such as bevacizumab against VEGF), there are limited precedents in the literature on anti-EGF drugs. This is probably due to the fact that EGF is a small globular-shaped protein with poor druggability properties and also to the complexity and diversity of mechanisms that trigger EGFR activation. Suramin, a polysulfonated naphthylurea originally developed in the 1920s for the treatment of parasitiasis, is the only EGF chemical ligand that has been reported to date, albeit showing low affinity (K_D in the millimolar range) and selectivity.⁶⁹ Indeed, suramin has been reported to interfere with the function of a number of growth factors, such as platelet-derived growth factor (PDGF), VEGF and hepatocyte growth factor. Its tumor anti-proliferative effects are thus due to a range of mechanisms of action, which also account for its toxicity at therapeutic doses. Its use has in fact been discontinued by the FDA due to safety concerns and lack of long-term efficacy.⁷⁰

In the same line, the first results achieved with the vaccine CimaVax provide an indirect validation of the therapeutic effects of targeting EGF. CimaVax was developed by Cuban physicians and is the first therapeutic vaccine for lung cancer, causing an increase in anti-EGF antibody titers that is correlated with a decrease in serum levels of EGF. So far, CimaVax has gone through three clinical phases and has demonstrated a significant effect, especially in NSCLC patients with high pretreatment concentrations of EGF.⁷¹

Strategies to overcome resistance: next-generation EGFR inhibitors and combination therapies

Despite the demonstrated benefits of EGFR-targeting agents, not all patients with cancer respond to treatment, and any gain in the median progression-free survival with these therapies compared to chemotherapy is, rather disappointingly, less than 1 year.³⁸ Generally, resistance to drugs can be classified as primary or secondary. Intrinsic or primary resistance is the failure to respond to cancer-targeted therapies—not to be confused with the lack of drug exposure on the tumor site, which has other causes. Primary resistance may be associated with the lack of activating mutations in EGFR, which make cancer cells highly dependent on this pathway. EGFR-activating mutations cluster in the catalytic kinase domain and are a distinctive feature of lung adenocarcinomas. In these cells, the overactivated kinase activity (caused by abrogated autoinhibition mechanisms) is responsible for the oncogenic transformation of fibroblast and lung epithelial cells.⁷²

Acquired or secondary resistance, on the other hand, occurs in patients who initially benefit from anti-EGFR drugs but after some time of treatment show disease progression. The genomic instability of cancer cells, magnified under the selective pressure imposed by targeted therapies, favors the onset of acquired resistance mutations (*Figure 9*).

From a mechanistic point of view, resistance mechanisms to FDA-approved EGFR TKIs or antibodies can be grouped into four categories, the first and second being the most common types. First, we find EGFR mutations that lead to a drug-resistant state, for example, the T790M or S492R mutations, which impair drug efficacy but do not diminish the activity of the receptor. The T790M “gatekeeper” mutation occurs on the ATP binding pocket of the kinase domain and is responsible for over 50% of cases of acquired resistance to first-generation TKIs. To tackle this issue, second and third generation TKIs, such as afatinib and osimertinib, have been developed against T790M-positive tumors.⁷³ Similarly, S492R is an acquired mutation found in 2 out of each 10 patients treated with cetuximab. This mutation reduces the affinity of cetuximab, but does not affect the binding of panitumumab to EGFR.⁶⁷ Finally, deletion of exons 2 to 7 of the EGFR gene results in a smaller EGFRvIII that exhibits ligand-independent constitutive phosphorylation. This mutant protein is found in various types of cancer, including glioblastomas, breast cancer, and NSCLC, and it typically presents poor responses to current anti-EGFR drugs.⁷⁴

	Lung	Colon	Glioblastoma	Head and neck
Primary resistance	EGFR-T790M BIM deletion	Activating mutations in KRAS, PI3K, BRAF, PTEN loss	R84K A265V/D/T Extracellular domain deletions (EGFRvIII)	
Acquired resistance				
EGFR mutation	T790M	S492R	EGFRvIII	EGFRvIII
Alternative pathway activation	ErbB2, ErbB3, VEGF, IGF, FGF, BRAF, PTEN loss	ErbB2, IGF, KRAS	PDGF, PI3K, STAT3, MET, PTEN loss	ErbB2, ErbB3, Aurora
Histologic transformation	Acquisition of stem cell properties Epithelial-to-mesenchymal transition SCLC transformation			Epithelial-to-mesenchymal transition

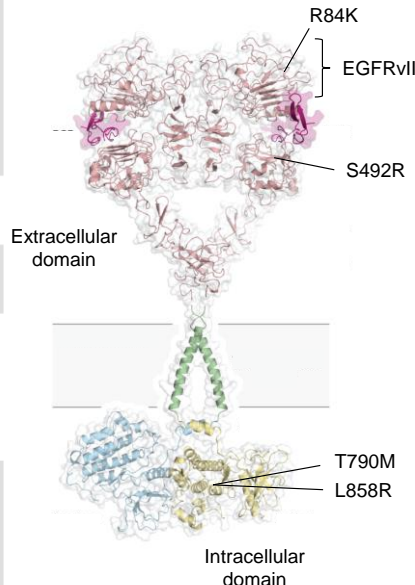


Figure 9. Overview of EGFR resistance mechanisms in lung, colorectal, glioblastoma and head cancer (left). Frequent mutations in EGFR found in cancer (right). Note that they gather in the ligand-binding domain and in the tyrosine-kinase ATP pocket region. EGFRvIII is a truncated EGFR with ligand-independent constitutive activity. Adapted from Chong, C. & Jänne, P. A.³⁸

In second place, cells can undergo “oncogenic shift” or activation of a bypass signaling pathway such as MET amplification or KRAS activation.³⁸ This countervailing mechanism may occur through any of the multiple, interacting pathways that contribute to cell proliferation. The result is the persistent activation of downstream signaling, despite EGFR inhibition, and hence cancellation of the apoptotic effects that are normally mediated by EGFR inhibition. The activation of these pathways may differ across cancer types; for example, BRAF activation triggers resistance to TKIs in EGFR-mutant lung cancer,⁷⁵ whereas KRAS mutation is responsible for the acquired resistance to anti-EGFR mAb in CRC.⁷⁶

Third, and related to the previous point, there are the genetic alterations that impair a pathway that is essential for anti-EGFR mediated apoptosis, such as loss of PTEN expression⁷⁷ or intronic deletions that remove the BH3 domain of BIM.⁷⁸ Fourth and last, histologic transformations to small cell lung cancer, as well as epithelial to mesenchymal transitions, have been identified. This resistance mechanism was reported in patients treated with EGFR TKIs and may be mediated by the activation of AXL kinase along

with AKT activation. Although epithelial to mesenchymal transitions occur in a minority patients, this rare phenomenon points to the role that EGF plays in development, as first observed by Cohen in 1962.²⁹

All in all, tumors are highly heterogeneous and may exhibit distinct mechanisms of resistance even within the same patient. For example, the degree of MET amplification with or without T790M mutation varies among metastatic sites sampled from patients with lung carcinoma.⁷⁷ This observation raises the question as to whether these alterations are novel spontaneous mutations or conversely whether they are selected from pre-existing resistant subclones by the anti-EGFR therapy. The latter explanation appears to be more plausible, meaning that the early resistant clones are selected under therapeutic pressure.⁷⁷

As we have seen, an understanding of the various resistance mechanisms to EGFR-targeted therapies is essential to develop efficient drugs to tackle the spread of cancer. Current clinical approaches to combat resistance include irreversible inhibitors that target specific mutations in EGFR, such as osimertinib. This TKI has demonstrated high efficacy in clinical trials and was recently approved by the FDA for the treatment of patients with T790M-positive carcinoma.³⁴ With respect to antibodies, mAb806 and CH12 have been developed to selectively bind to the mutated EGFRvIII (a constitutively active EGFR often found in glioblastoma) and are able to efficiently suppress the growth of EGFRvIII-positive tumor xenografts.⁷⁹ In the future, a deeper molecular characterization of the primary tumor in patients, together with a close monitoring of the molecular evolution of the tumor genotype over time, will allow the identification of specific genetic drivers and therefore the selection of targeted therapies for each cancer subtype.

Some researchers have drawn an analogy between drug resistance in HIV and targeted therapies in cancer:⁸⁰ both diseases are subject to high mutational rates that foster the onset of drug-resistant clones in response to therapy, both may benefit from combination treatment (as it is already done in HIV treatment), and both may be difficult to eradicate entirely. In cancer, the combination of targeted therapies is one of the main focuses of current preclinical and clinical studies. For instance, the concomitant administration of two TKIs, or a TKI and a mAb, has achieved promising results in particular subgroups of patients. As an example of this, the combination of cetuximab and erlotinib in wild-type KRAS metastatic CRC patients has proven highly effective clinically.⁸¹

Another potential solution to overcome the multiple mechanisms of acquired resistance is to target downstream pathways that mediate pro-apoptotic signals, thus circumventing the mutations that can arise in EGFR and in other proteins that are earlier in the signaling cascade. For example, knockdown of ROR1, a pseudokinase that is directly regulated by transcription factors, has proved sufficient to inhibit the growth of cancer cells with multiple acquired resistance mutations.³⁸ Also in this regard, being a convergence point of several upstream growth signals, the MAPK-ERK pathway may be another ideal candidate for drug intervention.

In conclusion, novel-acting drugs, used as part of combinational therapies targeting multiple pathways should be urgently developed. Furthermore, the recent advances in genomics, applied to the molecular-level analysis of cancer, should translate into highly effective personalized treatments that better control the escape of tumor cells and allow for an increased patient survival.

Peptides as covalent inhibitors of specific targets. The case of prolyl oligopeptidase (POP)

The interest in covalent peptide inhibitors

As discussed in the first part of the introduction, designed peptides are versatile structures that can allow the selective targeting of proteins involved in non-covalent PPIs implicated in disease pathogenesis. Typically, they perform their biological effect by competing with other endogenous ligands for the binding to the target protein. Due to the transient nature of most PPIs, such peptide ligands present fast off-rates from their target, contributing to moderate cell activity at micromolar concentrations.⁸² While this “soft” inhibition is desirable to modulate PPIs, site-specific binding (such as a substrate entering an enzyme active site) is governed by stronger interactions (often covalent), thus in this situation fast off-rates may represent an important limitation. As a solution, an electrophilic group can be attached to the molecule to generate a covalent-acting drug. Despite the success of some of these molecules as enzyme blockers, this approach has been ruled out owing to safety concerns regarding the potential off-target reactivity of the electrophile.

Only more recently, the quest for longer acting drugs with greater clinical efficacy has led to the resurgence of covalent drugs, especially with milder electrophiles, such as boronic esters, and more specific designs complementary to the target of interest.⁸³ In fact, there are currently 42 FDA-approved covalent drugs, and in the year 2013 alone, 3 out of 27 approved drugs were covalent inhibitors.⁸⁴ The field of oncology has been a rich source of covalent drugs, mainly due to the serious medical needs of patients and the acute administration of therapies. As we highlighted above, second- and third-generation anti-EGFR TKIs have been developed to overcome acquired drug resistance (afatinib was approved in 2013).

However, there are fewer examples of peptides and peptidomimetic drugs that incorporate an electrophilic group to make a covalent bond with a specific residue of the target protein. This is explained by the fact that these molecules have been exploited to target larger protein surfaces involved in PPIs, rather than reactive or catalytic protein domains, where small molecules have resulted more feasible. However, the selectivity of peptide-like drugs in protein recognition events can be also used to reduce off-target rates in the case of covalent inhibitors. This concept was demonstrated by

Walensky *et al.*, who developed an α -helical stapled peptide able to covalently target the BFL-1/BH3 interaction.⁸⁵ This PPI has the peculiarity that it contains a unique juxtaposition of cysteine residues at the binding interface. This feature was exploited by inserting, in the optimal position of the peptide, an acrylamide warhead that would covalently react with the cysteine side chain embedded at BFL-1 interface. With this example, they showed how the peptide backbone and side chains form a template that selectively recognizes the target PPI, whereas the electrophilic group is able to irreversibly bind to the protein.

Proteases as peptide targets

Major physio-pathological processes, including cancer, inflammatory states and thrombosis, are strongly dependent upon the fine regulation of proteolytic enzyme activity. Proteases play a central role in the synthesis and turnover of all proteins in cells and their malfunction is involved in the development of diseased states. Consequently, inhibitors of proteases are used to treat high blood pressure, human immunodeficiency, hepatitis C viral infections, type 2 diabetes, cancers, multiple myeloma, thrombosis, stroke, and inflammatory conditions, among others.⁸⁶

Proteolytic enzymes typically feature a multi-domain structure in which the catalytic domain is connected to one or more regulatory domains through flexible “hinge” regions. The regulatory region is the docking site for a number of endogenous molecules, as well as partner interacting proteins, that modulate the activity of the catalytic site.⁸⁷ Overall, far from the static picture that is often associated with enzymes, proteases are highly dynamic structures that sample a broad conformational space in order to perform their function.

A common feature in proteases, revealed by the analysis of >1,500 crystal structures, is their recognition of a peptide β -strand in their substrates and/or inhibitors.⁸⁸ Therefore, it comes as no surprise that peptides have achieved tremendous success in targeting proteases. Clues to develop active site inhibitors often come from endogenous peptide substrates or from natural product screening. Once a specific peptide sequence has been identified, subsequent optimization strategies are performed. These can include truncating polypeptide substrates to short peptides, replacing the cleavable amide bond by a non-cleavable isostere, or optimizing inhibitor potency through peptidomimetics.¹³

Prolyl oligopeptidase (POP), a protease with pleiotropic effects

Prolyl oligopeptidase (POP) is an 80-kDa cytosolic protein that hydrolyses peptides under 30 amino acids long at the carboxyl side of proline residues.⁸⁹ Proline residues, thanks to their particular cyclic side chain, impose conformational constraints and kinks in the secondary structure of peptides, which in turn require specific enzymes to process them. In addition, proline lacks the free hydrogen NH atom that is required to form a hydrogen bond with the backbone carbonyl oxygen during catalysis performed by enzymes such as chymotrypsin and subtilisin.⁹⁰ Indeed, typical serine proteases are unable to cleave peptides at proline residues. Only a few proteolytic enzymes, known as proline-specific proteases, can in fact accommodate the constrained shape of proline in their active site and hydrolyze the amide bond adjacent to proline residues. Therefore, these types of enzymes (which belong to the S9a subfamily of serine proteases) play a crucial role in the regulation of proline-containing peptide hormones and neuropeptides, such as oxytocin, vasopressin, angiotensin, bradykinin and substance P.⁸⁹

In mammalian organisms, POP is ubiquitously expressed but it reaches its highest levels in the brain.⁹¹ Since its discovery in the 1970s as an oxytocin-degrading enzyme,⁹² its biological function has been related to mental processes. The first hypothesis raised was that POP activity contributed to neuropeptide degradation, hence affecting cognitive processes.⁹³ However, this notion was in conflict with the spatial association between extracellular peptide substrates and this intracellular peptidase.⁹⁴ Further research pointed to the involvement of POP in the inositol triphosphate (IP₃) pathway. IP₃ is a second messenger involved in the transduction of neuropeptide signaling, subsequently increasing concentrations of intracellular Ca²⁺, a signaling cascade that is active in memory and learning functions.⁹⁵ Along this line, an inverse correlation between IP₃ levels and POP expression was observed in glioma cells;⁹⁶ however, the only *in vivo* study performed to confirm this hypothesis failed to show a significant modification of IP₃ levels in rat central nervous system (CNS) upon administration of a POP inhibitor.⁹⁷

All in all, these previous hypotheses failed to provide a complete understanding of the biological role of POP in the brain. In 2009, Di Daniel *et al.* were the first to postulate that POP may also act through PPIs, further showing in yeast-two-hybrid assays that POP interacts with GAP-43 (a regulator of synaptic plasticity) and with α -synuclein, the major

component of Lewi bodies in Parkinson's disease.⁹⁸ In subsequent publications, POP has been found to accelerate α -synuclein aggregation *in vitro* and in cells,⁹⁹ observations that correlate with the colocalization of POP and α -synuclein in post-mortem brain samples from patients with Parkinson's disease.¹⁰⁰

Like other enzymes, POP is a highly dynamic protein, undergoing a slow equilibrium between open and closed conformations in aqueous solution (*Figure 10*). These dynamics are essential for the biological function of this enzyme, as it has been shown in several computational and biophysical studies carried out by our group.¹⁰¹ In this regard, the addition of active-site inhibitors disrupts the conformational equilibrium by stabilizing the enzyme in the closed conformation. This effect also leads to a substantial alteration of the native PPIs of the enzyme, a mechanism that would in fact explain the positive effects of the administration of POP inhibitors to animal models of CNS disorders.

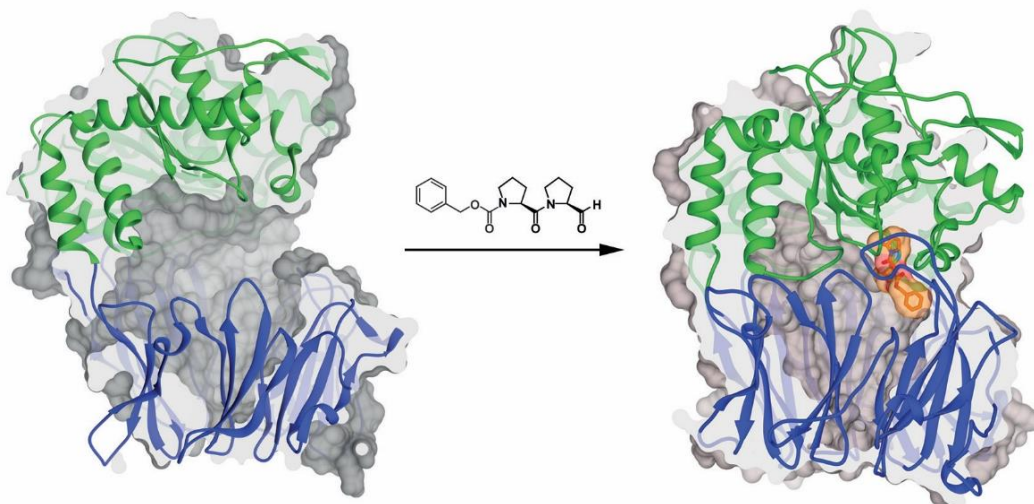


Figure 10. A) Active POP in the open conformation (left) and inactive/closed conformation (right), caused by the binding of the active site inhibitor Z-prolylprolinal (orange). The α/β -hydrolase domain is shown in green and the β -propeller in blue (PDB code: 1QFS). Extracted from López, A. *et al.*¹⁰²

Experimentally, POP inhibitors have shown anti-amnesic and memory-enhancing effects in a variety of animal models (mainly rats and monkeys) characterized by cognitive impairment.⁹⁴ In addition, the administration of such drugs to α -synuclein mutant mice led to a reduction in the levels of soluble and aggregated protein in the brain¹⁰³—an observation that was confirmed in monkeys with an induced form of Parkinson's disease.¹⁰⁴ In this direction, a recent study with mice proved that POP inhibitors were

able to accelerate the clearance of aggregated α -synuclein and, more importantly, restore the motor behavior impaired in Parkinson's disease.¹⁰⁵

In summary, all the reported evidence indicates the involvement of POP in mnemonic and cognitive processes, although the exact underlying mechanism is still not clear. These effects are probably originated by the enzymatic regulation of the IP₃ pathway and they proceed through PPIs with intrinsically disordered proteins such as α -synuclein and GAP-43—a hypothesis that may explain the relationship between POP and neurodegenerative disorders.

The medicinal chemistry of POP inhibitors

Due to the pleiotropic effects of POP and its association with CNS pathologies, the modulation of its activity by inhibitors has been a subject of interest for decades; in fact, the development of POP inhibitors has greatly facilitated the study of the *in vivo* POP mechanism. Most POP inhibitors derive from the canonical compound Z-prolylprolinal, a molecule that forms a hemiacetal bond with Ser554 of the catalytic center of POP, acting as a transition state analog.⁹⁵ This general structure has been the reference for the ligand-based development of most POP inhibitors, which share related P1', P1, P2 and P3 pharmacophores that will bind to the S1, S2 and S3 cavities of the POP active site, according to the Berger and Schechter notation.

A variety of C-terminal (P1') functionalities have also been explored, some of them (hydroxymethyl ketone, α -keto heterocycle, aldehyde or nitrile) able to react with the nucleophilic -OH group of the catalytic serine. Such inhibitors form a reversible covalent bond with the enzyme, a feature that endows them with a higher potency and a more sustained inhibition than their non-covalent analogues. This bond, despite being covalent, is hydrolyzed after some time, thus recovering enzymatic activity. In the field of POP inhibitors, there are very few examples of irreversible inhibitors; these are limited to highly reactive isoxazol¹⁰⁶ or chloromethyl ketone¹⁰⁷ electrophiles with broad antiprotozoal activity. If applied to the selective blockade of human POP, this type of irreversible inhibitor would provide slow off-rates (relative to the rate of re-synthesis of the protein) and sustained disruption of POP activity.

A detailed description of the vast diversity of reported POP inhibitors is out of the scope of this introduction; for a brief summary see *Figure 11*.

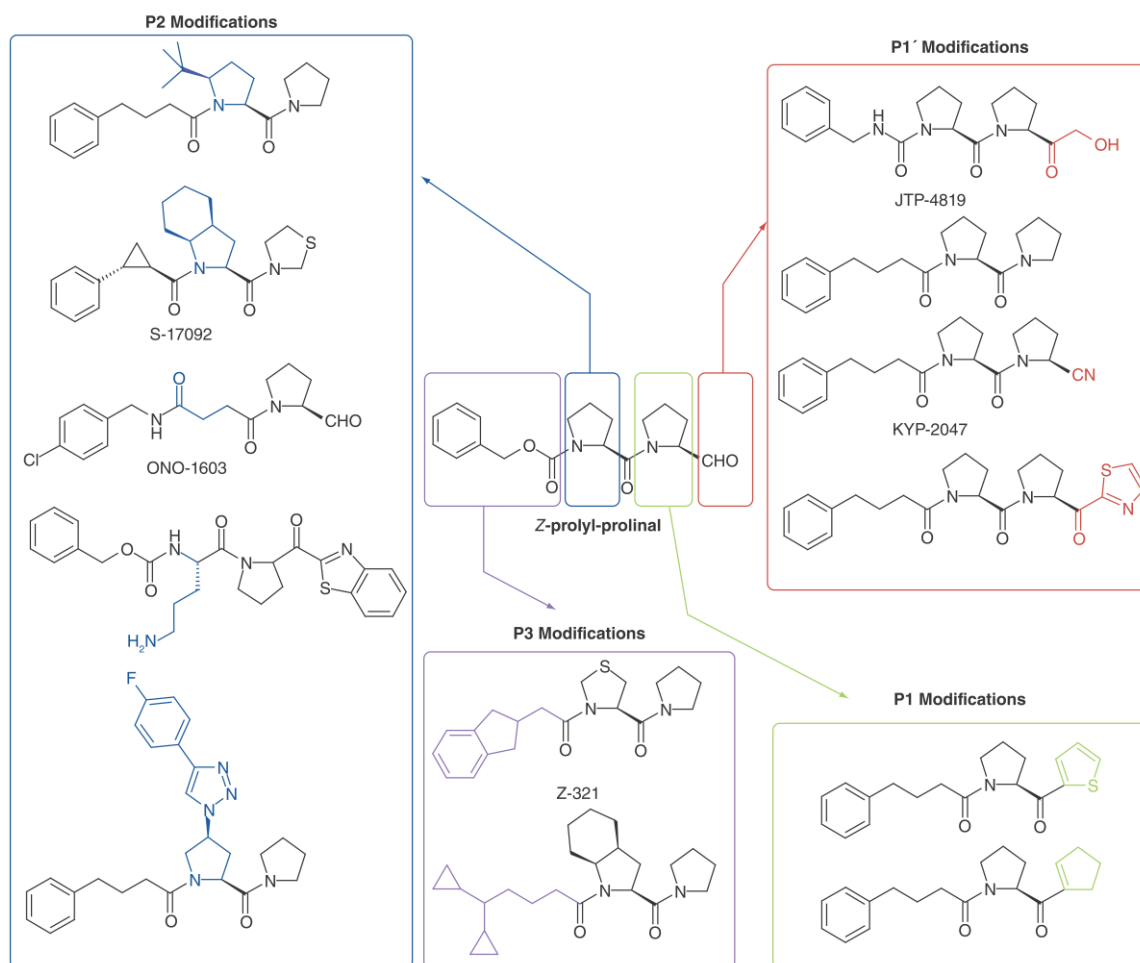


Figure 11. Examples of POP inhibitors obtained through modifications of the canonical compound Z-prolyl-prolinal in different positions. Extracted from López, A. *et al.*⁹⁵

Experimentally, POP inhibitors have shown anti-amnesic and memory-enhancing effects in a variety of animal models (mainly rats and monkeys) with cognitive impairment.⁹⁴ In addition, treatment of a type of α -synuclein mutant mice with POP inhibitors led to a reduction in the levels of soluble and aggregated protein in the brain¹⁰³—an observation that was confirmed in monkeys with induced Parkinson’s disease.¹⁰⁴

Besides preclinical studies, a few compounds have reached clinical trials in the treatment of cognitive deficits in humans. These inhibitors showed acceptable pharmacodynamics and no toxicity; however, no marked improvements in memory or cognitive tasks have been reported. A possible explanation is that so far, the design of POP inhibitors has been focused mainly on potency using a common pharmacophore, whereas other parameters, such as blood-brain barrier permeability, are not always optimized.

In conclusion, further studies are required to elucidate the biological role of POP, as well as to validate this protein as a relevant target for drug intervention in cognitive and neurodegenerative disorders.

OBJECTIVES

This PhD thesis explores the potential of designed peptides to target and modulate the function of therapeutically relevant protein targets involved in cancer (EGF) and cognitive disorders (POP).

The main part of the research work is devoted to the discovery of new peptide ligands of EGF with potential therapeutic application. Although several drugs inhibiting EGFR have already been developed, there are no reported molecules acting on EGF. As we have discussed in the Introduction, the chemical space explored by small molecules has remained insufficient for targeting such a small and structurally flexible protein. In contrast, peptides present larger and more diverse structures that can better adapt to irregular protein surfaces. In this regard, EGF peptide inhibitors could emerge as a new class of drugs for the treatment of cancer.

To accomplish this ambitious goal, the following specific objectives are established:

1. **Peptide docking tools** will be applied to the *de novo* design of EGF ligands, using as template the recently published NMR structure of the protein. In parallel, an array of suitable **biophysical techniques** will be set up to assess the potential interactions of our designed ligands with EGF.
2. As an alternative to docking-based methods, we will explore the **design of larger and structurally more complex peptides to mimic interacting epitopes of EGFR**. Computational tools will play an important role in this exploratory investigation, supporting the rational design and structural characterization of peptide hits. Moreover, we will synthesize the best candidates and assess their binding to EGF with the available biophysical methodologies.
3. In order to show a biological effect, peptides that bind EGF must also **disrupt the EGF-EGFR interaction**. In addition to being active at this level, they must display a convenient aqueous solubility, as well as chemical and **biological stability**. For our peptide hits, we will study all these features in a selection of *in vitro* settings, as well as in living cells.

In the last part of the thesis, we will explore another mechanism of action through which bioactive peptides can selectively modulate of therapeutically relevant proteins. In particular:

4. We will **design, synthesize and evaluate library of covalent peptide inhibitors that target the catalytic site of POP**, a protease that is involved in cognitive disorders. These molecules will feature a peptide scaffold endowed with a new class of sulfonyl fluoride electrophile, meant to react with the catalytic serine of POP. The interaction of these molecules with the target will be evaluated by enzymatic end-point activity assays and kinetic experiments. In addition, their selectivity against two proteases closely related to POP will be assessed. Finally, we will study the potential of our designed inhibitors to access their site of action—the brain.

RESULTS AND DISCUSSION

**Chapter 1:
Docking-based
design of
EGF peptide ligands**

Structural features of EGF

Human epidermal growth factor (EGF) is a single-chain protein consisting of 53 amino acid residues. It is synthesized by the cell as a transmembrane protein and is then secreted to the extracellular space, where it can produce autocrine (the same cell receives the input), paracrine (nearby cells) or endocrine (distant cells) effects. In all cases, EGF performs its function by binding to the extracellular region of epidermal growth factor receptor (EGFR) and triggering subsequent signaling cascades.

Like other growth factors, the primary structure of EGF comprises the $CX_7CX_{4-5}CX_{10-13}CXCX_8GXRC$ motif, where C is cysteine (Cys) and X represents any amino acid.⁴⁷ The Cys residues form three intermolecular disulfide bonds that divide the secondary structure of EGF into the A, B, and C loops. The A loop (residues 6–19) contains some α -helical structure, the B loop (residues 20–31) has two antiparallel β -strands connected by a β -turn, and the C loop (residues 33–42) comprises another short antiparallel β -sheet. Asn32, located between two Cys residues, functions as a hinge along which the two parts of the proteins can move. (*Figure 12*).

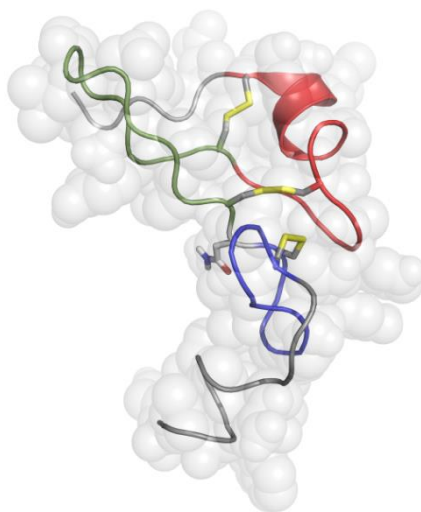


Figure 12. NMR structure of human EGF (PDB code: 2KV4). The A, B and C-loops are shown in red, blue and green, respectively. The “hinge” residue Asn32, as well as the six Cys residues, are shown as sticks.

Upon binding to EGFR, most of the surface area of EGF is buried inside the C-shaped pocket that forms the extracellular domains of EGFR.⁵² Despite the large area of contact between the two proteins, some specific amino acid residues on EGF are essential for the biological activity. Among these are the key residues involved in protein folding: all six Cys residues, G22 and G39. Mutation of these into any other amino acid leads to a total

loss of activity.⁴⁷ Extensive site-directed mutagenesis studies have also identified a few hotspots on EGF that do not play a direct role in the overall structure of the protein but make a major contribution to the free binding energy to the receptor. In particular, the largest reductions in binding affinity have been found for the R41A mutant, which shows only 0.05% of the affinity of the wild-type. Likewise, the affinity of the L47A mutant is reduced to 2% of the wild-type, whereas the affinity of I23A is reduced to 5.9%. Also, a Y13A mutant was shown to conserve only 2% of the activity of the wild-type.^{108–111}

Over the past decades, considerable attention has been paid to the structural elucidation of EGF in order to clarify the structure-function relationship. Due to the difficulty to grow good quality crystals of EGF not in complex with its receptor, most reported structures of this protein have been characterized by NMR. Early NMR studies of the solution structure of this molecule were performed with murine EGF and other analogues under highly acidic conditions. However, at pH values below 5, EGF is reported to lose its binding affinity to the receptor.¹¹² It was not until 2010 that the NMR structure of the full-length human EGF under physiological conditions was solved (PDB code: 2KV4).⁶⁹

Surprisingly, the only crystal structure of human EGF was reported to be an asymmetric dimer.¹¹² This structure (PDB code: 1JL9) was determined at pH 8.1 using a truncated protein (residues 6–47), which lacked the flexible C-terminus and thus produced higher quality crystals for X-ray diffraction. However, this dimer was probably an artefact resulting from the crystallization conditions, and it is currently accepted that EGF exists as a monomer in solution.⁵⁰

Identification of potential binding sites on EGF

As there are no reported endogenous ligands of EGF (other than its 140-kDa receptor), we decided to use a rational design approach, based on the wealth of structural data available on the EGF-EGFR system, in order to obtain new ligands for EGF with the potential to disrupt this interaction. Henceforth, the proposed plan of action was to employ computational design approaches to devise peptides that: a) would show consistent *in silico* binding to EGF; and b) would be feasible to obtain by synthetic methods in the laboratory.

To embark on this peptide discovery journey, we used the recently reported solution structure of human EGF under physiologic conditions (PDB code:

2KV4). In contrast to other structures that can be found in the PDB, this one includes the C-terminal hydrophobic core of the protein, which is reported to be relevant for binding to EGFR. As EGF is a small and globular-shaped protein, on first glance there are no *druggable* cavities on its surface.

Taking advantage of the Schrodinger molecular modeling suite available in our lab, we applied the SiteMap tool¹¹³ to identify potential binding sites on EGF. Using this algorithm, a small hydrophobic cavity was found between the A and B loops. This pocket is delimited by Tyr13 and Tyr29 and has polar acidic residues (Asp3, Glu5 and Asp27) that surround the upper side of the cavity (*Figure 13*).

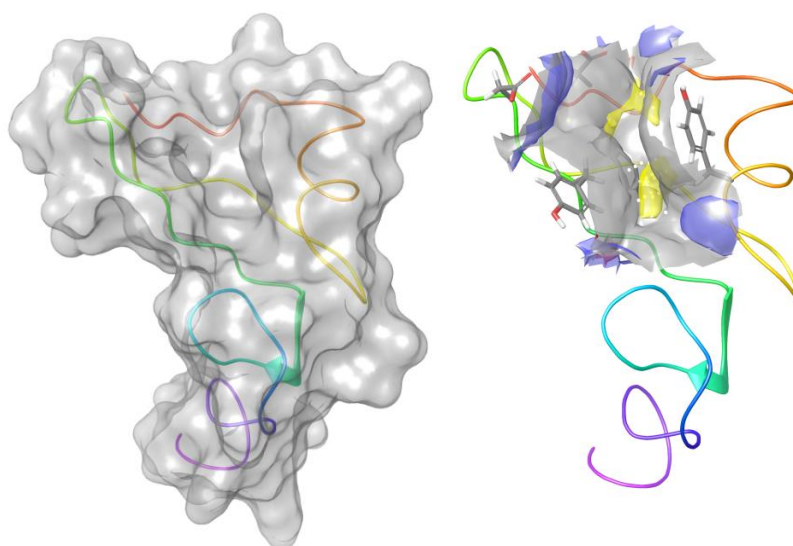


Figure 13. The solution NMR structure of EGF (PDB code: 2KV4) (left); note the narrow groove on the upper part of the protein, which could serve as binding site. A detailed view (right) of this cavity located between Tyr13, Asp27 and Tyr29. Hydrogen-bond donor groups are shown in blue, while hydrophobic groups are in grey.

To evaluate the druggability of this pocket, we took advantage of the know-how implemented in our laboratory by Dr. Martin Kotev, and performed the docking of a virtual library of 400 3-mer peptides. This library features a Pro in the central position and all possible combinations of natural amino acids (20^2) in the other two positions. Satisfyingly, the backbone of most tripeptides fit inside the tight cavity, and it was the tripeptide Arg-Pro-Arg that had a better predicted affinity (docking score: -7.70 kcal/mol, *Figure 14*).

Docking studies

Based on the virtual screening, we performed a manual refinement process, consisting of iterative rounds of docking and structural modifications on the ligand, using EGF (PDB code: 2KV4) as the docking template. Briefly, a flat aromatic ring in the central position of the peptide was found to nicely fit in the hydrophobic cavity (*Figure 14*). The Arg residue on the C-terminus, which could establish salt bridges with acidic residues of EGF, was kept, whereas a small polar residue, such as Ser, was introduced at the N-terminus. By exploring these mutations, we reached the peptide Ser-2Nal-Arg (**d1**, in *Table 5*), which yielded a docking score of -8.25 kcal/mol. The large naphthalene ring in the central position is sandwiched between the side chains of Tyr13 and Tyr29, and the basic side chain of the Arg makes strong ionic interactions with Asp3.

The addition of a Glu residue at the N-terminus of the peptide (resulting in Glu-Ser-2Nal-Arg, **d2**) slightly improved the predicted affinity for EGF (docking score: -8.44 kcal/mol), the negative charge of its side chain forming a salt bridge with Lys28 in EGF. Although we tested dozens of other modifications, we were not able to achieve lower docking scores with this type of peptide. The addition of extra amino acids, with the aim to establish new favorable interactions with the protein, did not either enhance the binding affinity. Moreover, having too many rotatable bonds on the peptide chain slows down the calculation process and is a source of inaccurate conformations.

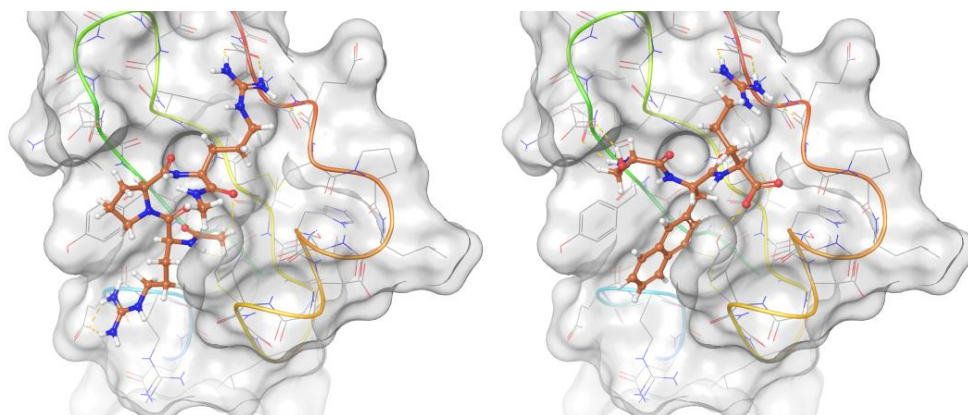


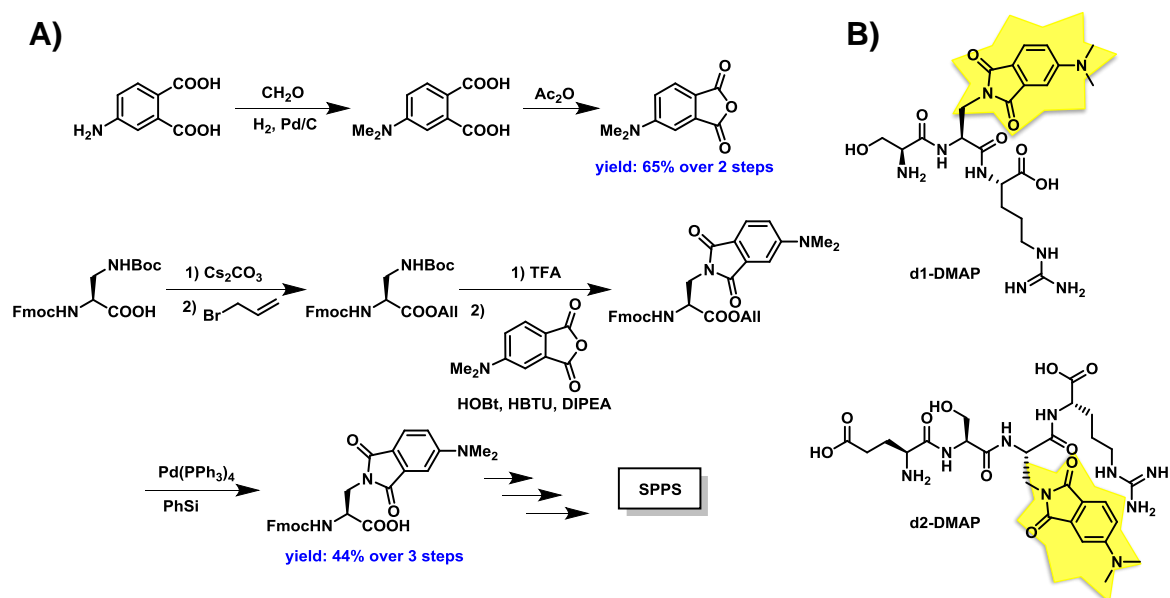
Figure 14. Representation of the lower-energy poses of the tripeptides Arg-Pro-Arg (*left*) and Ser-2Nal-Arg (*right*) docked on the surface of EGF. Hydrogen bonds are indicated with yellow dashes.

Biophysical tools to study the interaction of peptides with EGF

A fluorescence method to assess ligand binding

In order to evaluate the docking predictions, we synthesized peptides **d1** and **d2**. Given that at that time we had no biophysical method with which to study interactions with EGF in the laboratory, we decided to set up a fluorescence-based assay that was developed by Imperiali *et al.*¹¹⁴ In a nutshell, this method is based on introducing in the peptide sequence an environment-sensitive chromophore (4-dimethylaminophthalimide, 4-DMAP), which yields a dramatic increase in fluorescence upon ligand binding. Like other environment-sensitive fluorophores, which are bulkier, 4-DMAP has a similar size to tryptophan, hence rendering this dye less disruptive to native interactions.

The synthesis of the Fmoc-protected 4-N,N-dimethylaminophthalimidoalanine was accomplished following the reported procedure¹¹⁵ in five steps (Scheme 1). The fluorophore was then used as a standard Fmoc-amino acid in SPPS.



Once **d1** and **d2** and their corresponding fluorescent analogues (**d1-DMAP** and **d2-DMAP**) were obtained, a titration experiment with EGF was performed for both candidates. In this assay, the DMAP-labeled peptide (at 5 μM concentration) was titrated against increasing concentrations of EGF. An increase in fluorescence, as well a shift in the emission maximum (typically to the left), would reflect the change in the environment of the fluorophore that would occur upon ligand binding.

However, for both **d1-DMAP** and **d2-DMAP**, we observed little change in their fluorescence emission spectra upon titration with EGF (while Imperiali *et al.* report >100-fold signal increase).¹¹⁴ When plotting the fluorescent intensities against the protein concentration and fitting the points to a 1:1 binding model, we obtained an exponential-type of curve that was far from reaching the saturation plateau, thus impeding the calculation of a K_D value (*Figure 15*). In addition, we did not observe a marked shift in the emission maximum that would also be indicative of a binding event.

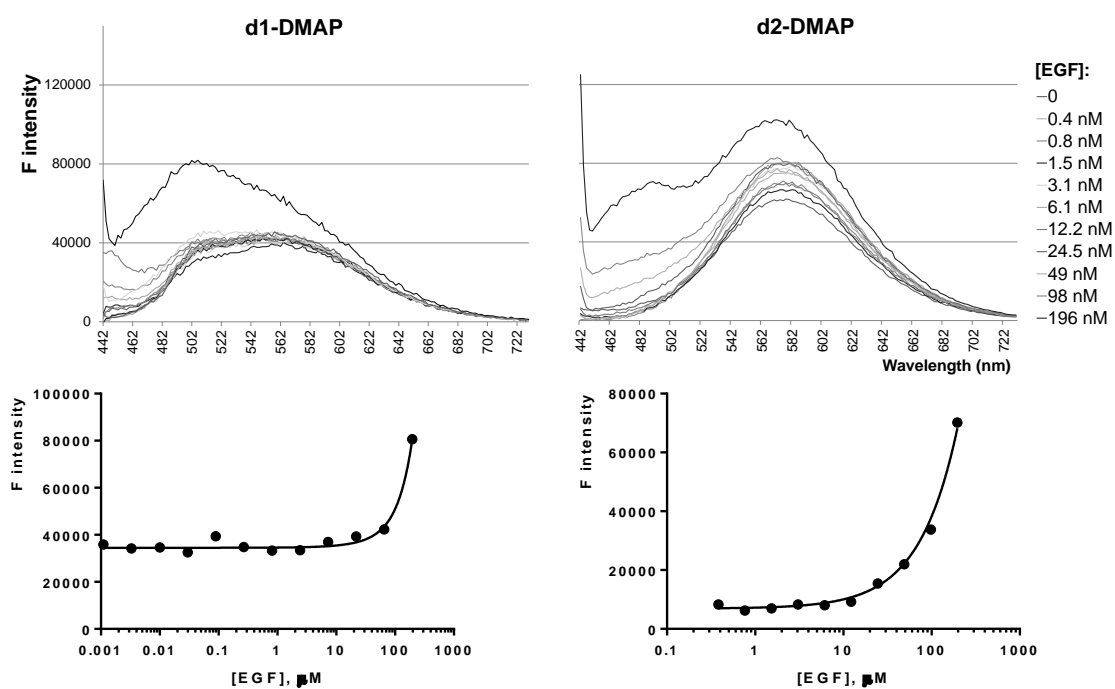


Figure 15. Fluorescence emission spectra of peptides **d1-DMAP** and **d2-DMAP**, after sequential additions of 0.4-196 μM of EGF (top). Relative fluorescence increases at corresponding concentrations of EGF (bottom).

These results casted doubts on whether the biophysical technique was appropriate for reporting interactions with a small-sized and solvent-exposed protein like EGF and whether the docking-designed peptides bound to EGF with sufficient strength to be detected by this method. On the basis of these preliminary results and with limitations to explore an alternative fluorescence-based method (such as fluorescence polarization, as a result of the small size of our protein),¹¹⁶ we decided to switch to other techniques, as described in the following paragraphs.

Isothermal titration calorimetry (ITC)

During a biomolecular interaction event, the formation of non-covalent interactions (van der Waals contacts, hydrogen bonds, salt bridges, etc.) at the interface, together with the reorganization of nearby water molecules, often results in heat exchange. ITC is the only approach that directly measures heat exchange during complex formation at a constant temperature and it provides useful information about the forces (enthalpic or entropic) that drive the interaction process.

In our case, the potential ligand **d1** was titrated (16 injections of 3 μL) into a solution containing 15 μM of EGF, both samples dissolved in the same buffer. As we can observe in *Figure 16*, the area under each peak, which corresponds to the heat evolved at each ligand injection, decreases gradually as the ligand binding site becomes saturated. However, the binding isotherm did not reach the saturation plateau, indicating a weak binding to the protein. Even so, by fixing the stoichiometry of the interaction to 1, and fitting the data to a single binding site model, a K_D value of 855 μM could be obtained for the **d1**-EGF interaction.

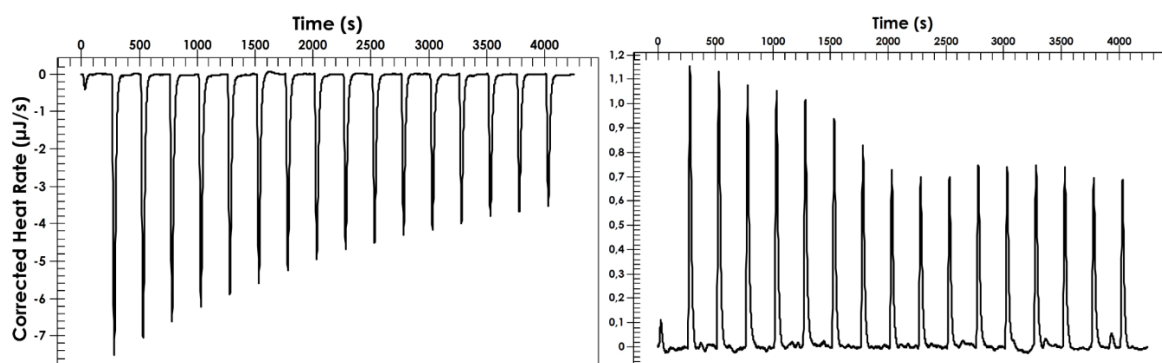


Figure 16. ITC isotherms of the titration of EGF with increasing concentrations of **d1** (left) and **d2** (right).

When the same experiment was performed for **d2**, we did not observe substantial changes in the binding isotherm upon addition of this peptide to EGF, even when the peptide was injected at high concentrations (3 mM). On the contrary, the endothermic peaks indicated that the peptide underwent aggregation at such high concentrations.

NMR methods to probe ligand-EGF interactions

As a parallel effort, with the help of Dr. Jesús Garcia and Dr. Nick Berrow, we sought to set up a robust expression system to produce recombinant human EGF, with the ultimate objective to conduct NMR experiments with an isotopically labeled form of the protein. Various expression constructs and *E. coli* strains were tested, since the main difficulties that we encountered were very low expression yields and the recovery of mis-folded species of EGF. In this regard, the formation of erroneous disulfide bonds is a frequent cause of protein misfolding and aggregation in inclusion bodies.

Eventually, we tested the SHuffle T7 strain of *E. coli*, which has an oxidative cytoplasmic environment that favors disulfide bond formation.¹¹⁷ In addition, this strain constitutively expresses the disulfide bond isomerase DsbC, which promotes the isomerization and folding of mis-oxidized species into their correct form. In addition, we added a thioredoxin (Trx) protein fusion tag,¹¹⁸ which works both as a solubility enhancer and promoter of disulfide bond formation (see Materials and Methods for details). With this expression system, we were able to obtain human EGF in moderate yields (~2 mg/L of culture). Furthermore, by growing the bacterium in minimal medium containing ¹⁵NH₄Cl as the sole source of nitrogen, we produced ¹⁵N-labeled EGF in moderate to low yields (~0.5 mg/L of culture).

In addition to Trx, other protein fusion tags were explored. When the small ubiquitin-like modifier (SUMO) protein was fused to the EGF construct,¹¹⁹ we achieved slightly better EGF yields (3-4 mg/L of LB, 1 mg/L of minimal media) than with the Trx system, probably due to reduced protein misfolding and proteolytic degradation. Therefore, SHuffle T7 *E. coli* cells transfected with the SUMO-EGF plasmid became our standard for robust EGF expression, both in LB and minimal media (see Materials and Methods for more details).

The production of uniformly labeled ¹⁵N-EGF allowed us to obtain a high-quality nuclear magnetic resonance (NMR) spectrum of the protein. The ¹H-¹⁵N HSQC spectrum was compared to that previously reported,⁶⁹ thus

confirming the correct folding and structure of EGF in our recombinant expression system (*Figure 17*).

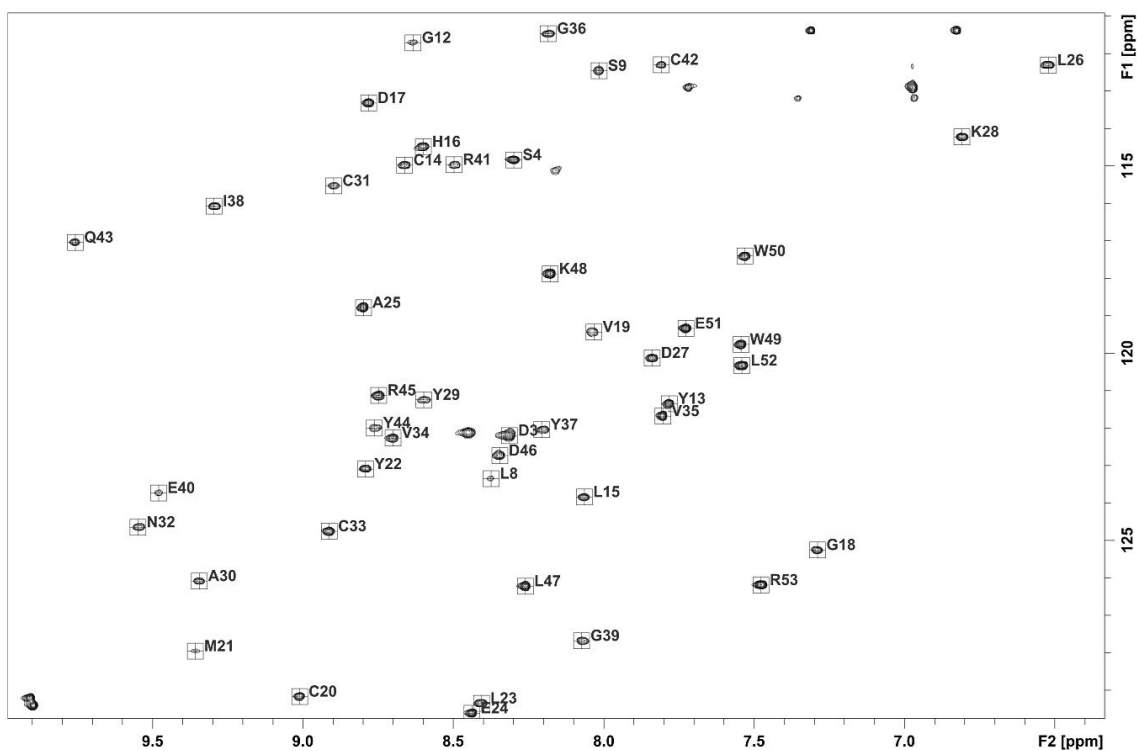


Figure 17. ^1H - ^{15}N HSQC spectrum of ^{15}N -EGF showing the assignment of cross-peaks to the protein residues. Spectra were acquired at 298 K in 20 mM sodium phosphate, 50 mM NaCl, and 0.1% NaN_3 (pH 6.8).

Furthermore, NMR methods are able to detect the binding of ligands to biomolecular targets of interest. In particular, protein-based methods have significant advantages over other biophysical approaches. The observation of discrete chemical shift perturbations (CSPs) is the strongest evidence for a specific, well-defined binding event. As a result, protein-based NMR screening is essentially immune to false positives that can arise from non-specific binding and other artifacts derived from ligand aggregation. Indeed, the unique CSP fingerprint induced by ligand binding can be used to identify the ligand binding site and even to orientate the ligand on the target protein. Therefore, we decided to apply this NMR methodology to the screening of our docking-designed ligands.

Upon incubation with EGF, both **d1** and **d2** (at 2 mM concentration) produced small but reproducible effects on the protein. The EGF residues most affected (see *Figure 18*) were those located in the A-loop of the protein (L8, S9, G12, H16, D17), which is known to interact with domain III of EGFR. Residues V34, V35 and Y44, which form part of the C-loop, but are

close to the previous ones (V34 and V35 are 5.5 and 8.7 Å from D17, respectively), were also affected (*Figure 18*, top).

In order to identify the peptide features relevant for the interaction and, in the process, potentially enhance the affinity of these ligands, we envisaged some modifications on the structure of **d1** and **d2**. As a result, in peptide **d3** (see *Table 5*), the central aromatic residue was replaced by Leu, which features a hydrophobic but much bulkier side chain, which, in principle, would not be able to fit in the narrow cleft between Tyr13 and Tyr29. This compound would therefore serve as a negative control to validate our design strategy. In **d4**, the N and C termini were capped to remove the electrostatic charges at the ends of the peptide. In **d5**, residues in positions 2 and 3 were exchanged to check for sequence specificity. Finally, in **d6**, the positive charge of Arg was removed and substituted with a Leu residue in order to guarantee a lipophilic side chain.

Upon incubation with EGF, peptide **d4** produced very similar changes to those achieved with **d1** and **d2** on the NMR signals of the protein, thereby indicating, as expected, a similar mode of binding (*Figure 18*). Peptide **d5** caused a slightly smaller CSP than **d1**, thus evidencing some sequence specificity. In contrast, **d3** and **d6**, which lack the key elements for the docking model to work, did not cause any observable change in the HSQC spectra.

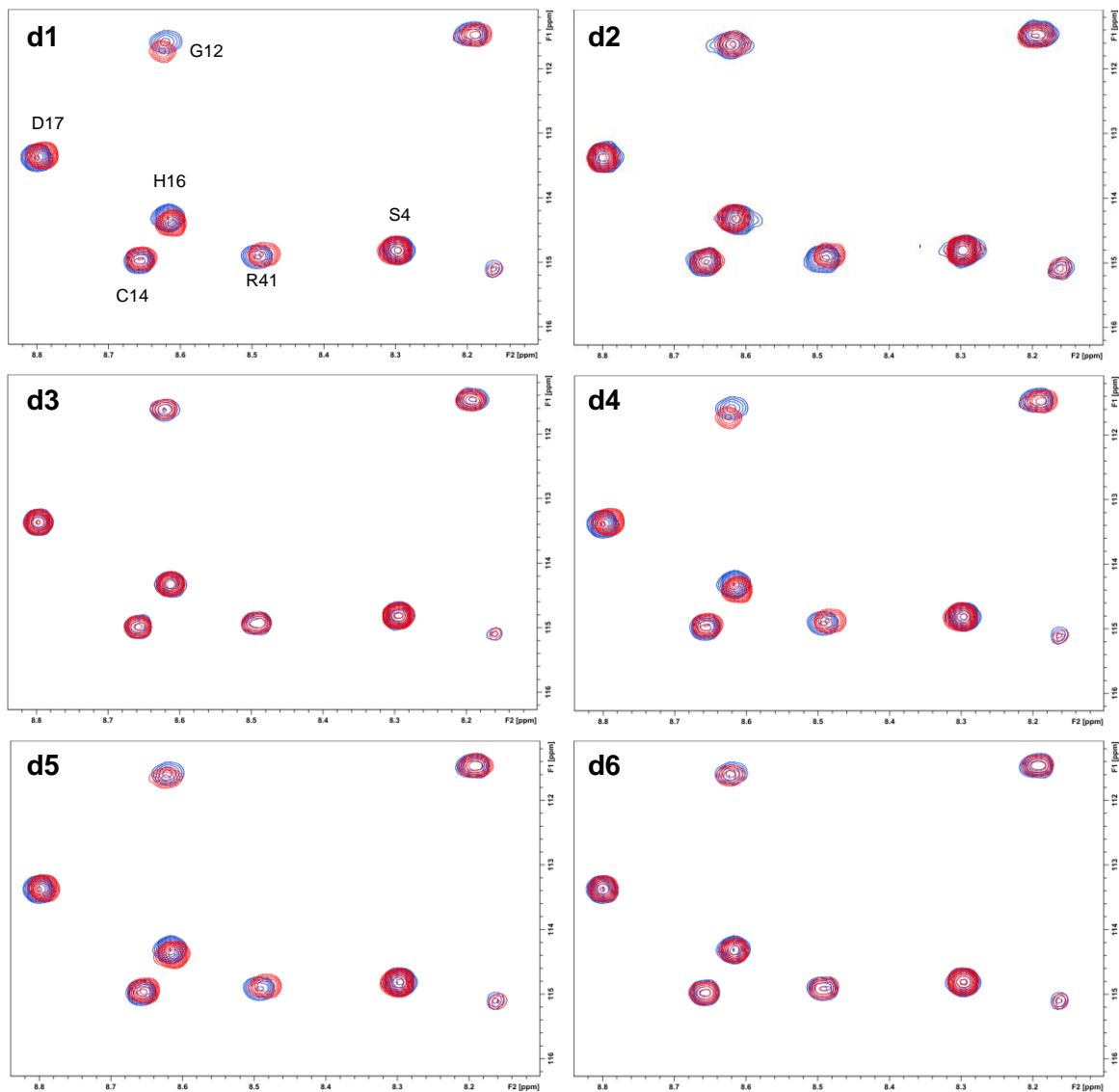


Figure 18. Overlay of the ^1H - ^{15}N HSQC spectra of ^{15}N -EGF in the absence (blue) and presence (red) of 1.5 mM of peptides **d1** (top left) to **d6** (bottom right). Spectra were acquired on samples containing 50 μM EGF in 20 mM sodium phosphate, 50 mM NaCl, and 0.1% NaN_3 (pH 6.8) at 298K.

Surface acoustic wave (SAW) for the quantitative analysis of ligand-EGF interactions

As an alternative biophysical method, we used a surface acoustic wave (SAW) device, recently purchased by the IRB Mass Spectrometry and Proteomics Core Facility, to address the quantification of the binding phenomena between our peptide candidates and EGF.

In a biosensor, we can distinguish two elements: an immobilized biomolecule, which serves as the recognition element of a specific analyte, and a transduction element, which converts the molecular events into electrical outputs, in this particular case, through the use of acoustic

waves.¹²⁰ This technique requires the immobilization of one of the interacting partners (typically, the protein), while ligands can be dissolved in buffer and flowed over the sensor surface. Changes in the phase and amplitude of the acoustic waves report on the mass loading of the chip, thus allowing the quantification of binding events (*Figure 19*).

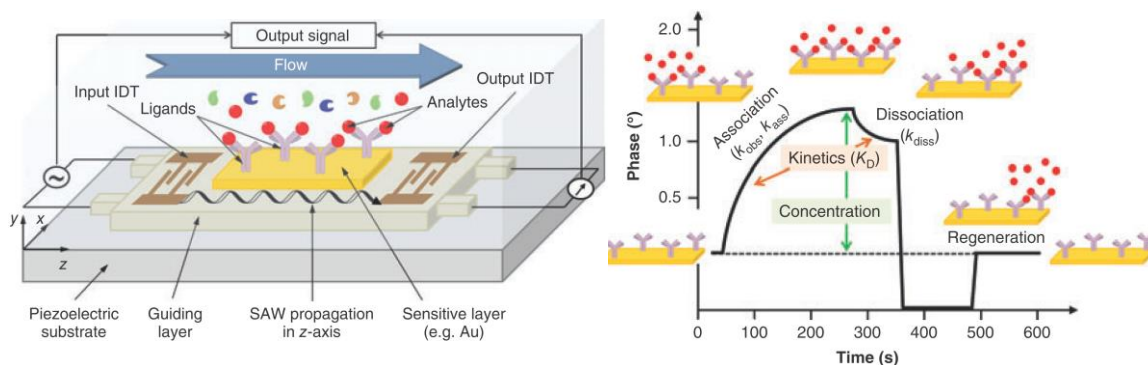


Figure 19. Scheme of a SAW device (left). An interdigital transducer (IDT) generates acoustic waves that propagate on the surface of a piezoelectric crystal. These waves travel across the chip and are modified by biochemical events that happen on its surface. A second transducer, located at the end of the chip, converts the phase and frequency of each wave back into an electric signal. Typical binding cycle performed in SAW biosensing experiment (right). Extracted from Díaz-Lobo, M.¹²⁰

In order to apply the SAW technique to the screening of our peptide ligands, EGF was first covalently immobilized on the biosensor chip (coated with a 2D-monolayer of 16-mercaptohexadecanoic acid) through the side chains of the Lys residues of the protein. We first assessed the functionality of the immobilized protein by injecting the extracellular domain of EGFR at a range of nanomolar concentrations. Saturation was observed at 75 nM, and linear regression analysis was performed to obtain a K_D of 8.2 nM, in agreement with the reported value for this interaction, calculated by a surface plasmon resonance biosensor.¹²¹ We further validated our SAW-based binding assay by injecting suramin, a polysulfonated naphthylurea and the only ligand of EGF that was previously reported. We calculated the K_D using a 1:1 binding model, obtaining a value of 0.20 mM, which matched the value of 0.29 mM reported by ITC for this interaction.⁶⁹

Once the technique was validated, peptide ligands were injected at a range of micro- to millimolar concentrations. Subsequent changes in the amplitude of the SAW were monitored, and K_D values were calculated by linear regression. As a result, weak affinities were obtained in most cases for ligands **d1-d6** (*Table 5*). The first docking-designed tripeptide **d1** showed the lowest K_D for the entire set of ligands, albeit still in the low mM range

(Figure 20 and Table 5). The addition of an extra residue to the N-terminus (Glu, meant to interact with Lys28 in the 3D-structure of EGF), as in **d2**, did not improve binding affinity.

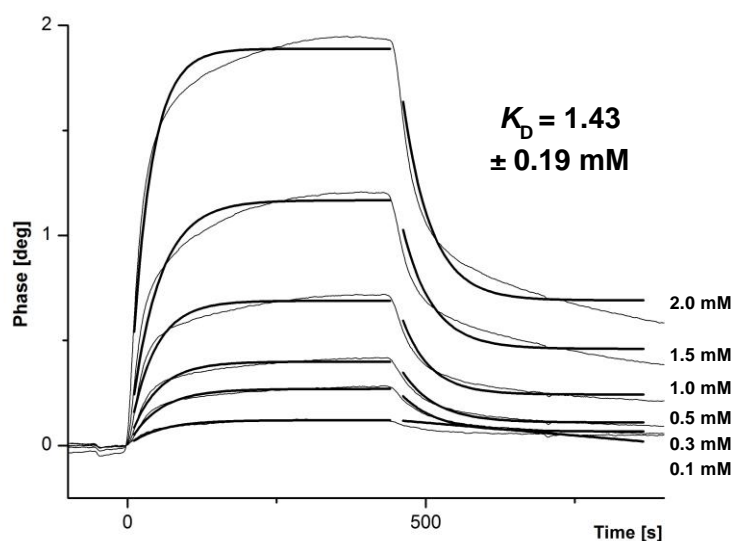


Figure 20. SAW sensorgram of **d1** injected at various concentrations.

Peptide **d3**, as expected, lost all affinity towards the protein upon mutation of the aromatic residue for a Leu residue. Moreover, position swapping of residues 2 and 3 (**d5**) resulted in decreased affinity, thereby revealing the importance of a planar hydrophobic group in the central position of the peptide. The removal of the C-terminus Arg (**d6**) was detrimental for the interaction, as a weaker K_D was obtained for this peptide. In the docking model, the basic guanidine group of Arg formed a strong ionic interaction with the side chain of Asp3 from the protein. Finally, in ligand **d4**, the removal of the N and C termini electrostatic charges reduced, as expected, but did not abolish the interaction with EGF. These observations thus reveal the contribution of the peptide backbone and side chains to the interaction.

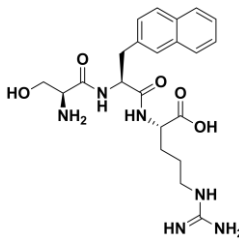
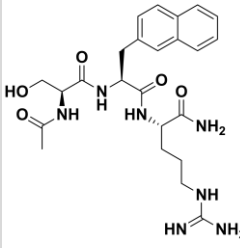
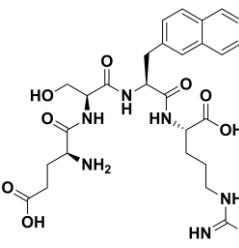
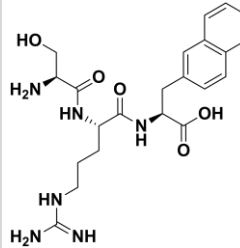
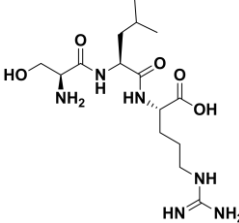
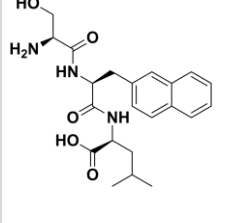
Id	Structure	Changes	K_D	Id	Structure	Changes	K_D
d1		Docking-designed original peptide	1.4 ± 0.2 mM	d4		N- and C-terminal capping	3.4 ± 0.5 mM
d2		Additional Glu to interact with Lys28 in EGF	1.5 ± 0.8 mM	d5		Residues 2-3 are swapped	2.1 ± 0.2 mM
d3		2NaI to Leu mutation (lack of aromatic ring)	No binding	d6		Arg mutated to a non-basic residue (Leu)	2.6 ± 1.4 mM

Table 5. Library of small peptide ligands **d1-d6**. Affinity values obtained by SAW.

Discussion

In the previous sections, the key role of the EGFR-EGF interaction in cancer cells has been clearly stated; therefore, all strategies investigated to inhibit this receptor activation represent a great opportunity to find new therapies in oncology. In particular, the innovative approach followed in this thesis is challenging: discovering peptide ligands that bind to the smallest partner of the interaction, EGF, a 7 kDa-protein with no catalytic or specific binding domains (as instead it will be pursued in the last Chapter focused on POP inhibitors) was an ambitious, but at the same time stimulating goal.

Due to the complexity of the case, several techniques not only for the design of the ligands but also for the characterization of the interaction have been explored, adapting the method by tuning the general parameters in function of this specific situation. In particular, computational strategies can have an invaluable impact in guiding the discovery of new peptide binders of a particular protein, especially in those cases for which no endogenous ligands or drugs (as is the case of EGF) are available.

Among other *in silico* techniques, docking experiments predict the binding orientation of two interacting partners in a faster and more cost effective way than more exhaustive methods such as molecular dynamics (MD). Generally, and here was no exception, the protein is assumed to be rigid during the docking process, and only the ligand is allowed a certain degree of flexibility.

However, proteins are flexible and dynamic entities, and such target flexibility can have a direct impact on the robustness and accuracy of docking experiments. Despite its small size and the three disulfide bonds that maintain its globular shape, EGF is a flexible protein that undergoes structural changes at different pH in solution, as well as upon binding to EGFR. Indeed, there is an RMSD of 2.7 Å between the NMR structure of EGF at physiological pH (which was used here for the computational experiments) and the bound form of EGF, present in the crystal structure of the 2:2 EGF-EGFR complex.

Even within the NMR structure of EGF, among the ensemble of 10 structures deposited at the PDB, there is substantial difference in the positions of the side chains of many residues. Considering this issue, we chose the 1st structure of the ensemble for the docking simulations, since it has the lowest energy and thus is the most representative conformer of the ensemble. In this structure, there is a narrow pocket located between Tyr13 and Tyr29, and this was the binding site onto which we docked our designed ligands. However, analysis of the 2nd and 3rd structures of the ensemble reveals that the side chains of the two Tyr residues are in different orientations and that the hydrophobic pocket in-between is smaller in the 3rd ensemble, and absent in the 2nd.

The transient character of this pocket can also be assessed by performing standard MD simulations. In *Figure 21*, we can observe that this pocket opens and closes quickly, as a result of the intrinsic flexibility of the protein. Although some authors claim that these transient pockets are an opportunity for the design of PPI inhibitors,¹²² we hypothesize that the lack of a stable binding site on the surface of EGF may explain the poor correlation between the computational predictions and the experimental results obtained for our designed peptides.

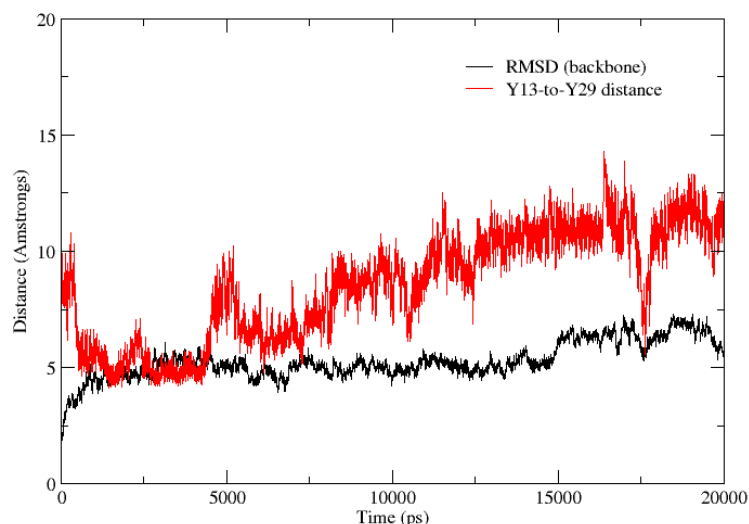


Figure 21. *Molecular dynamics simulation of EGF in explicit water. Notice how the binding pocket between Y13 and Y29 is closed during the first 5 ns of trajectory, and briefly at $t = 17$ ns; during the rest of the simulation the pocket is in a dynamic open conformation.*

In addition to the limitations underlying the computational design, we faced the challenge of setting up a robust biophysical method for detecting and quantifying the interactions of our ligands with EGF. As EGF is a small protein, many of the usual screening techniques such as fluorescence polarization were not applicable to assess protein-ligand interactions. The alternative methodology that we used, based on the changes in fluorescence emission, does not, in principle, depend on the overall size of the ligand-protein complex. However, in our case, we faced other issues associated with the specific nature of the target PPI.

EGF is a globular-shaped protein, with no unique regions confined from water molecules. In order to report a ligand-protein interaction, a solvatochromic fluorophore such as 4-DMAP must undergo considerable change in its microenvironment, going from hydrophilic (when the ligand is free in solution) to hydrophobic (when the ligand is bound in the complex). Indeed, most applications of this fluorophore consist of ligands that bind with high affinity to a hydrophobic cavity on the surface of a protein.^{123,124} In our situation, the most probable scenario is that the ligand binds to a solvent-exposed region of the protein, and possibly with weak affinity, if any. This would explain the results we obtained and prompted us to try other methods.

With regard to the use of ITC, our attempts to fine-tune the conditions of the experiment (ligand concentration, protein concentration, temperature, buffer, etc.) resulted in the best of cases in relatively flat binding isotherms.

Although a binding event could be inferred from these data, mathematical fitting often led to a lack of accuracy in the determination of thermodynamic parameters (K_D , n , ΔH , ΔG).

In a calorimetry experiment, the shape of a binding isotherm changes according to the product of the association constant (K_a) and concentration of protein ($[M]$), a relationship which Wiseman et al. summarized by the c value (*Equation 1*). Taking this into account, it is recommended to work with c values between 10 and 500, as receptor saturation can be achieved by adding as little as 2 equivalents of ligand, and the binding isotherm adopts a sigmoidal curve which allows an optimal fitting of the mathematical equations and accurate determination of the thermodynamic parameters.¹²⁵

$$c = nK_a[M]_t \quad \textbf{Equation 1}$$

At c values below 10, typical for high micromolar or millimolar interactions, the sigmoidal shape of the curve is lost, and it is necessary to add several equivalents of ligand per injection to achieve final receptor saturation. In such a challenging scenario, the n parameter must be known and fixed with accuracy in order to determine the values of K_D and ΔG .¹²⁵ Signal-to-noise ratio is also a significant problem for studying high affinity systems at low c values; all these factors thus limit the applicability of ITC for our screening purposes.

Fortunately, the use of NMR for the observation of chemical-shift changes in the protein ^1H - ^{15}N HSQC spectrum proved to be a highly valuable and sensitive tool in our binding studies. This experiment allowed us not only to probe ligand interactions but also to extract structural information about the binding mode of our docking-designed peptides. The largest chemical-shift changes were detected for residues that belong to the A-loop of EGF, an observation that would be in agreement with the docking simulations. However, the low affinity of our peptides meant that we had to use high concentrations of ligand in order to saturate the binding site and thus detect changes in the protein spectra. This also hindered the determination of affinity constants from titration experiments, as saturation of the titration curve could not be achieved. It is important to mention that the NMR screening method required the setting up and optimization of ^{15}N -labeled EGF expression for the first time in our laboratory, a feat that was successfully accomplished.

Finally, the results obtained with the SAW biosensor showed that we successfully immobilized EGF on the chip through NHS/EDC chemistry.

Crucially, the protein proved to be stable and maintained its functionality in response to the reported ligands (suramin, and especially, the extracellular domain of EGFR). Although the interactions for our ligands were weak, we were able to quantify them in a robust and reproducible way. In our hands, the SAW technology provided a low-cost (the protein on the chip was stable for weeks), label-free, and highly sensitive method for observing real-time binding events.

On the basis of the data collected, we drew several conclusions, which served as basic principles for the further development of the project (as will become evident in the next chapter):

1. Our computational design was successful in providing a set of small-sized peptides that were able to bind to EGF, a small globular protein for which no drugs of this kind had been reported. However, the binding potency of our ligands was very weak, probably due to the intrinsic flexibility of the protein and the transient character of the pocket that we chose to target.
2. The weak affinity, in addition to the small size of the protein, had consequences on the biophysical methodologies that we could apply for the screening. Although we succeeded in expressing the recombinant protein, major drawbacks were encountered in the application of fluorescence and calorimetry techniques.
3. We were able to set up an effective low-cost, label-free, and highly sensitive SAW platform in order to identify and quantify the EGF-binders interactions.
4. The SAW-based screening method complemented the use of ^{15}N -labeled EGF (expressed for the first time in our lab) in NMR experiments, which provided a more complete picture of the binding mode of our peptides.

Chapter 2:
Design of peptides
mimicking interacting
domains of EGFR

Structural analysis of the EGF-EGFR interaction

Epidermal growth factor receptor (EGFR) is a 170-kDa membrane protein composed of many flexible parts, including a large extracellular portion, a region that crosses the cell membrane, a kinase domain, and a long tail. The extracellular region of the receptor is composed of four articulated domains (I-IV) that recognize EGF. As discussed in the introduction, these domains adopt a folded or tethered conformation in the inactive state. Upon EGF binding, the receptor undergoes a conformational change that leads to an extended conformation. In this state, EGFR can bind to another receptor copy (or another member of the ErbB family), forming the dimeric complex that is necessary for intracellular kinase activation.

Being a large membrane protein with a highly dynamic structure, EGFR has proved a very challenging target to characterize from a structural perspective. In this arduous journey, researchers have used a divide and conquer strategy, splitting EGFR into several pieces and studying each one separately, mainly by X-ray diffraction of the corresponding EGFR fragments. Consequently, several PDB files, including 1NQL, 1IVO, 2JWA and 1M17, are needed to create the picture of the full receptor protein, allowing understanding of the EGF-EGFR activation mechanism at the atomic level.

The structure of the extracellular domain of EGFR bound to EGF, solved in 2002, revealed that each EGF molecule binds to an EGFR monomer, thus forming a 2:2 complex.⁵² In this complex, domains I, II, and III of EGFR are arranged in a C shape, with EGF accommodated between domains I and III. The EGF-EGFR binding interface is extense and spans several epitopes of domains I and III, with a total buried surface of 1440 Å². Structural and biochemical studies have shown that EGF binds first to domain I of the receptor (with an affinity of 50–200 μM), this region being a key recognition epitope of EGF.¹²⁶ Binding to domain III occurs subsequently and has a stronger affinity (400 nM), although this interaction is pH-dependent and is abolished at pH<5.¹¹²

In the structure of the dimeric active complex (PDB code: 1IVO), the binding interface of EGF with domain I of EGFR has a surface of 720 Å² and involves a number of specific contacts (*Figure 22*). Residues 16–18 of EGFR and residues 31–33 of EGF form a short parallel β-sheet, with several hydrogen bonds that stabilize the interaction. The side chains of Leu14, Tyr45, Leu69, and Leu98 in domain I of EGFR make a hydrophobic interaction with Met21, Ile23, and Leu26 in the B loop of EGF. In fact, the

Ile23 binding site, formed by the side chains of Leu14, Tyr45, and Leu69 of EGFR, has a shape roughly complementary to that of the Ile side chain. The Gln16 side chain of EGFR forms a hydrogen bond with the Asn32 side chain of EGF. Finally, the Glu90 side chain of EGFR is close to the Lys28 side chain of EGF, allowing the formation of a salt bridge.

Despite spanning a similar-sized surface (720 Å²) as domain I, the binding of EGF to domain III is stronger, as a result of the prevalent interactions that are formed in this region (*Figure 22*). Here, the Val350 and Phe357 side chains of EGFR hydrophobically interact with Leu15 and Tyr13, respectively, of EGF. In fact, the latter aromatic side chains are stacked upon each other in a typical π - π interaction. Also, Leu47 of EGF is involved in hydrophobic interactions with the side chains of Leu382, Phe412, and Ile438 of EGFR. In close proximity, the Gln384 side chain of EGFR hydrogen bonds with the backbone groups of Gln43 and Arg45 (EGF). Finally, the Asp355 side chain of EGFR makes a key salt bridge with Arg41 (EGF).⁵² The relevance of some of these interactions has been confirmed experimentally by site-directed mutagenesis studies (see Chapter 1, p. X).

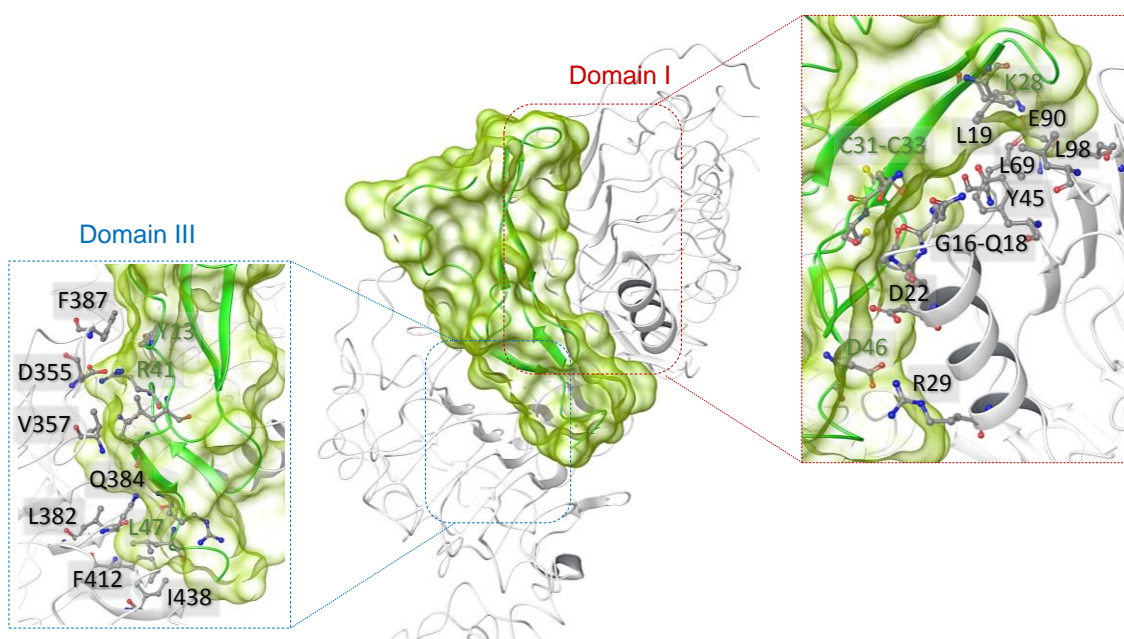


Figure 22. Detailed structure of the EGF-EGFR complex in its active state (PDB code: 1IVO). EGF is depicted in green surface, EGFR in grey cartoon. The key residues of domain I and III of EGFR (in black) and EGF (in green) are shown.

Mimicry of a 7-mer loop with constrained cyclic peptides

Following the structural analysis of the EGF-EGFR interaction, the search for alternative ligand design strategies led us to consider the mimicry of relevant binding epitopes in EGFR. We first focused our attention on continuous epitopes of EGFR, as these typically present fewer difficulties to be reproduced by conventional peptide scaffolds. In particular, we identified a 7-mer peptide (Asn12-Gly18) in EGFR that, in the bound complex, adopts a loop-like structure and forms several hydrogen bonds with the B-loop of EGF. In this loop (*Figure 23*), the side chains of Lys13, Leu14 and Gln16 contribute to the interface with EGF, while those of Asn12 and Leu17 face towards the back, their side chains at close distances.

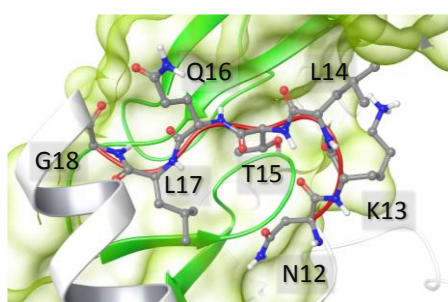


Figure 23. Close view of the interacting loop (in red) in EGFR domain I.

Assuming that we used a linear 7-mer peptide to mimic this loop, its high flexibility would very likely translate into a high entropic cost to adopt the bioactive conformation, thus hampering binding to EGF. In order to restrain its flexibility, we envisaged a cyclization strategy through the side chains of Asn12 and Leu17, as these residues do not form part of the binding interface. Towards this aim, we explored a variety of linkers that would enable the formation of the cyclic molecule, while maintaining its bioactive conformation. We also considered additional constraints, such as backbone N-methylation or the introduction of a Pro residue, which would produce a more rigid peptide structure.

Due to the elevated number of possible combinations and the challenging synthetic procedures that were envisaged to obtain the different cyclic analogues, we decided to apply computational tools to predict the structure of a set of cyclic analogues. Then, we selected those presenting the most native-like low-energy conformations for subsequent synthesis and characterization.

In order to perform the *in silico* conformational analysis, the coordinates of the atoms corresponding to this 7-mer peptide (Asn12-Gly18) were extracted

from PDB 1IVO. Then, Asn12 and Leu17 were manually mutated to Lys and Glu (or different lengths, such as Orn and Asp), respectively, and the peptide was manually cyclized through the side chains of these residues, whilst fixing the rest of the peptide. Often, an additional amino acid, such as Gly or Pro, was added in-between in order to cover the distance between Asn12 and Leu17. In all cases, the newly formed cyclic linker was subjected to a short minimization to relieve the local clashes, which arose as a consequence of the distortions introduced by the manual cyclization.

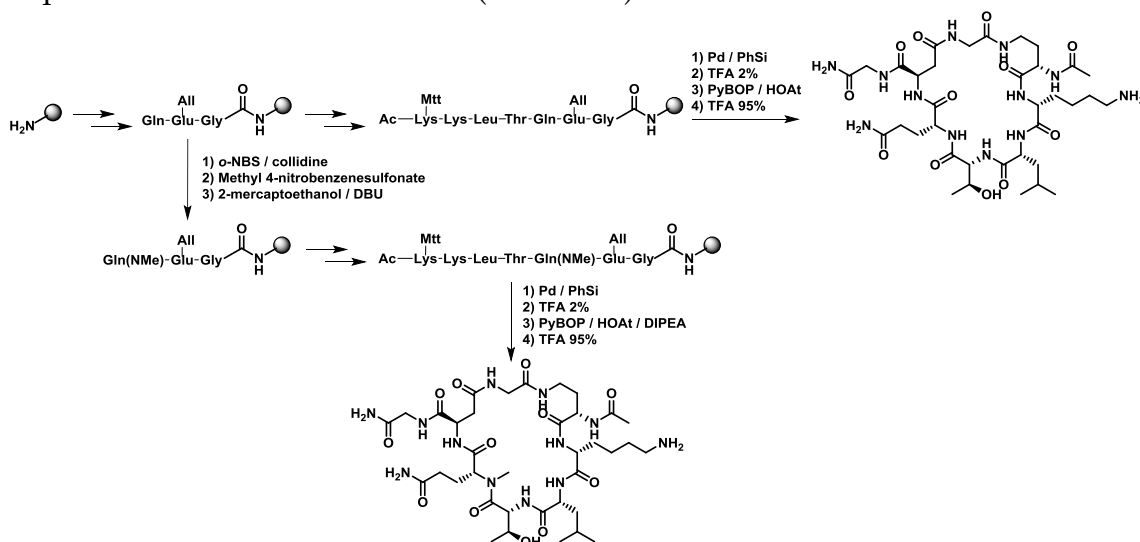
We then attempted to perform standard MD simulations on these structures, in order to identify their lowest-energy conformations. However, we realized that this technique, although very powerful for sampling local minima extensively, is highly sensitive to the starting coordinates of the peptide. Therefore, access all conformations would require very long simulation times, especially if high energy barriers were present, as was likely the case due to the constrained nature of our peptides.

As an alternative solution for our problem, we used the Macrocylic Conformational Sampling tool,¹²⁷ which is part of the Schrodinger modeling suite. This sampling protocol includes repetitive cycles of heating/cooling and energy minimization (see the Materials and methods section for more details). As a result, more conformations can be sampled faster and a final lower-energy conformation can usually be accessed. A selection of the most “receptor-like” structures that we obtained is compiled in *Table 6*. As expected, the N-methylated peptides had higher minimal energies than their non-methylated analogues (due to a smaller number of intramolecular hydrogen bonds and increased backbone hindrance). However, some of them exhibited very low RMSD values in comparison with the EGFR loop, and these were the ones that we selected for synthesis.

Id	Cyclization linker	Potential energy (kcal/mol) ¹	RMSD ²
Lp1	Asp-Gly-Dab	-1770.8	2.0
Lp2	Dab-Gly-Asp	-1753.1	1.9
Lp3	Dab-Pro-Asp	-1618.0	1.9
Lp4	Dab-Pro-Glu	-1609.0	1.9
Lp5	Dab-D-Pro-Glu	-1602.6	2.4
Lp6	Dab-Gly-Asp + NMe-Gln	-1602.0	2.4
Lp7	Pro-Dab-Glu	-1592.3	3.1
Lp8	Lys-Glu	-1592.2	1.9
Lp9	D-Pro-Dab-Glu	-1588.4	3.5
Lp10	Linear peptide	-1524.5	2.3
Lp11	Dap-Asp	-1468.2	2.0
Lp12	Dab-D-Pro-Glu + NMe-Gln	-1457.0	2.3
Lp13	Dab-Pro-Glu + NMe-Gln	-1449.5	2.0
Lp14	Lys-Glu + NMe-Gln	-1433.5	2.0
Lp15	Pro-Dab-Glu + NMe-Gln	-1426.1	3.2
Lp16	D-Pro-Dab-Glu + NMe-Gln	-1424.2	3.5

Table 6. List of cyclic peptides explored by conformational sampling; those in bold were selected for chemical synthesis. ¹Averaged for the 5 structures of minimal energy. ²RMSD for the backbone of the interacting motif (Lys13-Gln16), using the EGFR structure as reference.

The selected peptides were synthesized using orthogonal protecting groups for the side chains of the residues involved in the cyclization. This approach allowed us to carry out the cyclization reaction on-resin, thus avoiding the usual polymerization issues associated with cyclization reactions in solution. When required, backbone N-methylation was conducted following the reported Mitsunobu conditions (*Scheme 2*).¹²⁸



Scheme 2. General synthesis of the constrained heptacyclic peptides. For the N-methylated analogues (bottom part), an on-resin Mitsunobu step was carried out.

While this synthetic route allowed us to reach the desired cyclic scaffolds, it was not free of obstacles. First, we used the coupling reagent COMU for the on-resin cyclization reaction, as it was reported to provide higher coupling yields than HATU/HBTU.¹²⁹ However, this resulted in the formation of Oxyma covalent adducts (+141 Da), which decreased the overall yield and complicated the purification of the crude product. The use of a combination of PyBOP and HOAt solved this issue. Second, the efficiency of the cyclization step was highly dependent on the peptide sequence. Generally, we observed that peptide sequences containing Pro more readily underwent the cyclization step—unless the Pro itself was the reacting nucleophile. In some cases, the reaction did not proceed to completion and the remaining linear peptide often co-eluted with its cyclic analogue under semi-preparative HPLC conditions. All in all, we were able to obtain and purify a small library of 8 cyclic peptides that shared the Lys13-Gln16 interacting motif. Five of these peptides contained proline rings and 3 featured an N-methylation on the Gln residue (*Figure 24*).

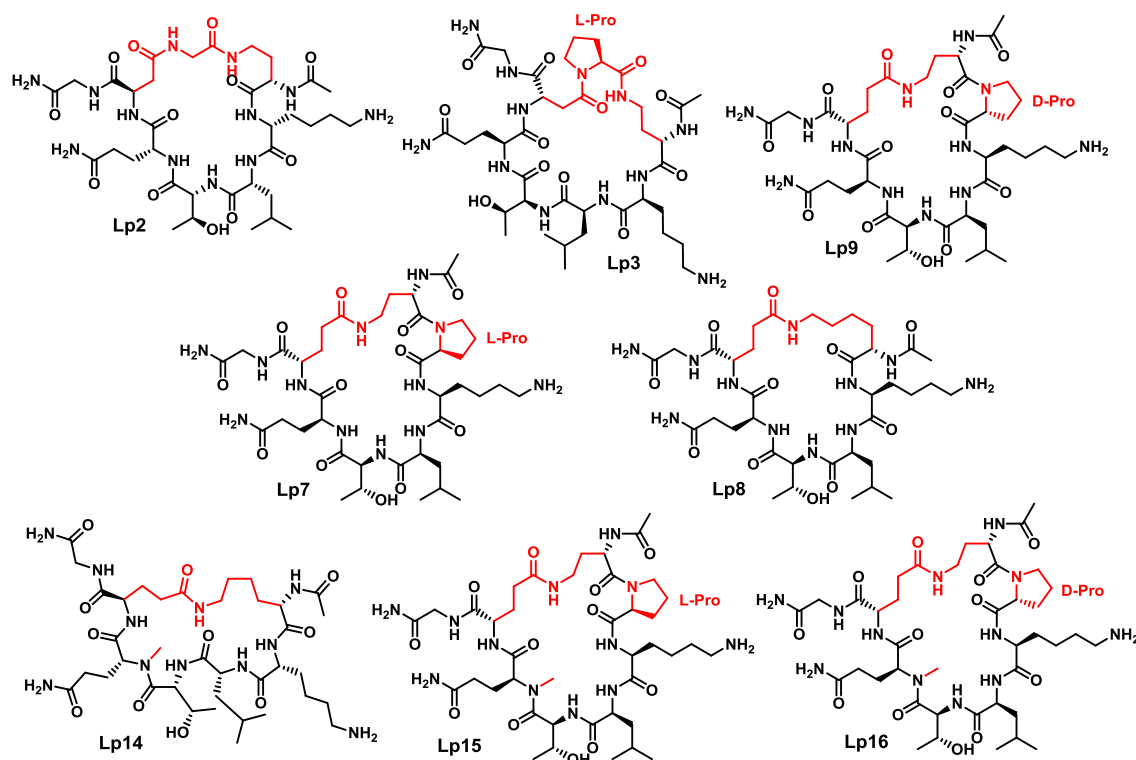


Figure 24. Family of cyclic heptapeptides synthesized as mimetics of an interacting loop in EGFR. The cyclization linker and other constraining elements are shown in red.

Despite the synthetic efforts dedicated to the optimization of their synthesis, NMR and/or SAW showed very weak or no binding to EGF. Very small chemical shift perturbations were observed on the ¹H-¹⁵N HSQC

spectrum of EGF, even when concentrations of up to 1.5 mM of ligand were added. The SAW technique confirmed these results, as none of the compounds produced clear association-dissociation responses on the biosensor. Given the lack of positive results with this approach, we decided to focus our design strategy on the mimicry of larger epitopes of EGFR.

Mimicry of a 28-mer cyclic domain using designed mini-proteins

After a careful look at the structure of domain I of EGFR, it can be appreciated that the 7-mer loop described above is part of a larger 28-mer cyclic subdomain, which spans residues 7 to 34 of the receptor (*Figure 25*). As structural features, this subdomain comprises a 3-turn α -helix and a long flexible loop containing a disulfide bridge (Cys7-Cys34) that closes the cycle. In the crystal structure of the complex, this region of EGFR makes contact with a large surface on EGF. In addition to residues 16–18 of EGFR, which form a short parallel β -sheet with EGF—discussed above, several charged residues on the interacting face of the α -helix, such as Arg and Asp, are at close distances to EGF in the complex structure and may also contribute to the binding process. Moreover, this region is highly conserved in other proteins of the ErbB family, thereby suggesting that it may be a relevant epitope for the recognition of growth factor proteins.

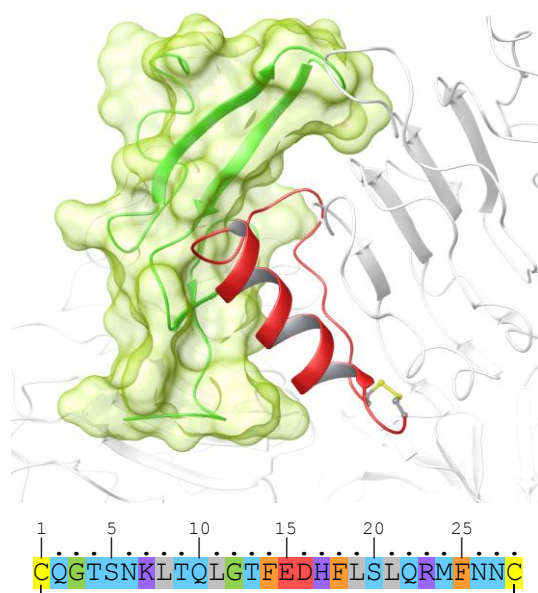


Figure 25. Structure and sequence of **cp28** (in red) as part of the EGFR domain I. The α -helix features 3 turns and spans residues 14 to 23 of the sequence. Note the disulfide bond connecting the two Cys residues that leads to the cyclic structure.

With the aim to study the solution structure of this 28-mer cyclic subdomain of EGFR (which will be referred to as **cp28** hereafter), we applied a series of computational simulation methods. First, we explored three secondary structure prediction servers (*Figure 26*). Despite following independent algorithms in their calculations, all three methods conferred residues 15-23 of the peptide a high propensity of α -helix, in agreement with the topology present in the crystal structure of the receptor.

For a more thorough assessment of the structure of **cp28** in solution, we took advantage of the expertise of Dr. Jesús Seco to perform replica exchange molecular dynamics (REMD), a computational method that overcomes the multiple-minima problem of molecular dynamics in predicting peptide and protein folding. In other words, standard MD simulations ran at single low temperatures (typically 300 K) tend to get trapped in local minimum-energy states. By exchanging non-interacting replicas of the system at several temperatures, REMD allows for a wider exploration of the conformational space.¹³⁰

Among the ensemble of conformations explored during the REMD experiment, the most populated ones (which as a cluster accounted for >75% of all conformations) preserved the initial α -helix motif (*Figure 27*). The rest of the peptide structure, despite its intrinsic flexibility, underwent relatively small changes during the simulation, as revealed by the RMSD metrics of the peptide backbone. These results show, at least *a priori*, that the structure of the peptide would be retained outside its membrane receptor environment. Going a step further, we ran a free MD simulation of the

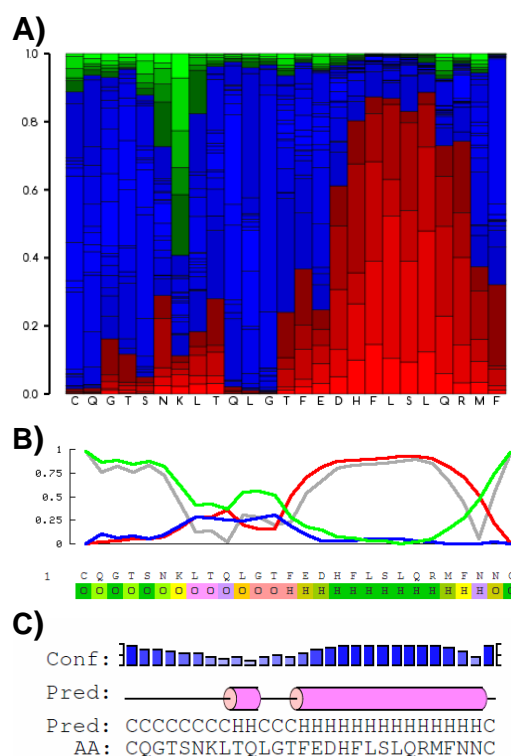


Figure 26. Structural prediction of peptide **cp28**, according to three different computational algorithms. *PepFold* (A) uses a *de novo* approach followed by energy minimization; *Sparrow* (B) is based on a statistical learning procedure, and *PSI-PRED* (C) performs sequence homology on proteins held in the PDB. Color code: α -helix, red; β -sheet, green; random coil, blue. Note that all of them confer a highly helical character to the C-terminal region of the peptide.

cp28-EGF complex, with the purpose of studying this interaction *in silico*. During the 12-ns simulation time, we observed that both molecules moved closer and that **cp28** was able to reproduce the same interacting pose with EGF as that depicted in the crystal structure (*Figure 27*).

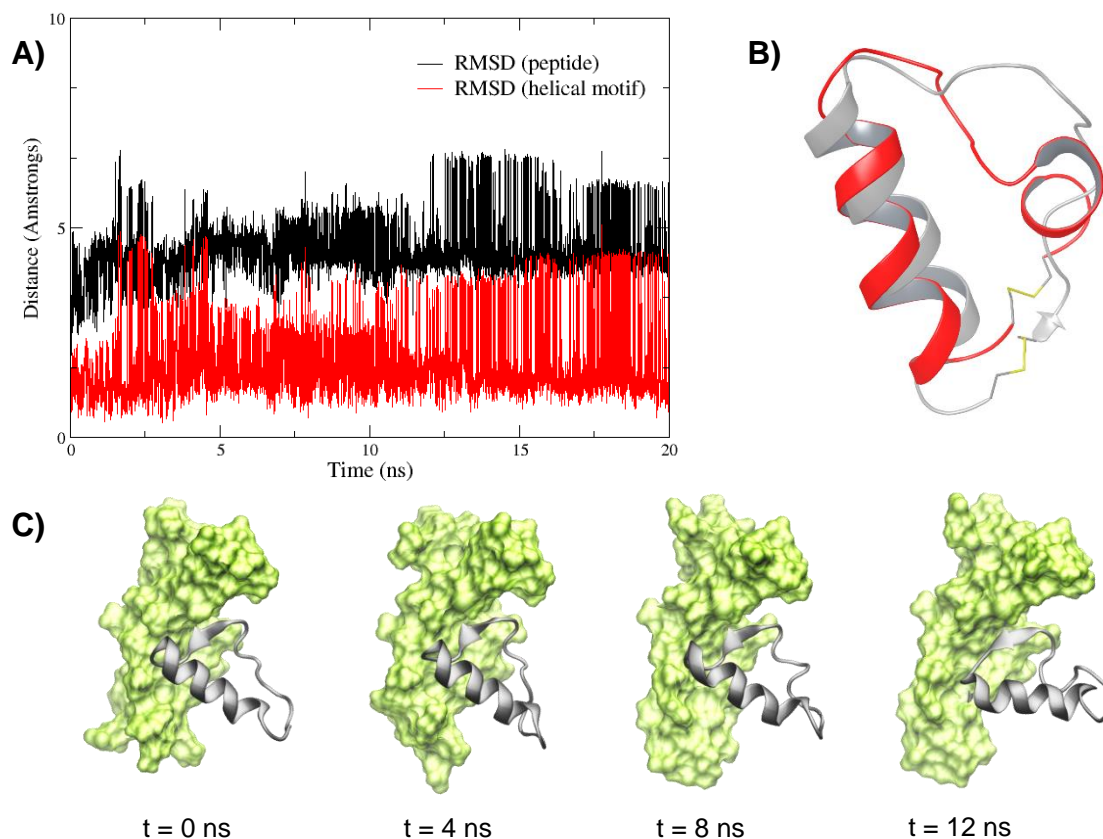


Figure 27. A) RMSD of the **cp28** backbone and of the α -helix motif for the 300 K trajectory during the REMD simulation. B) The most populated cluster of conformations for **cp28** during the REMD experiment (78% of all conformations, in red) is superimposed with the initial structure (in grey, coordinates extracted from PDB 1IVO). C) MD simulation frames of the interaction of **cp28** (in grey cartoon) with EGF (green surface).

Given these encouraging findings, we decided to pursue the chemical synthesis of **cp28** and perform its experimental validation. For the synthesis of this 28-mer, the peptide chain was manually grown on solid support (see Materials and methods). Once cleaved from the resin, the cycle was formed by the intramolecular oxidation of the cysteine thiol groups in aqueous solution. The structure of **cp28** in solution was assessed by circular dichroism (CD) spectroscopy, which provides a rapid and accurate fingerprint of the peptide folding in solution. In the experiment, 10% TFE was added to better simulate the environment close to the protein surface and favor the folding of the peptide. In agreement with the predictions, the

CD spectra of **cp28** showed two negative minima at 208 and 222 nm (*Figure 29*), the typical markers of α -helix.

We then applied the SAW technique to assess the potential interaction of **cp28** with covalently immobilized EGF. As shown in *Figure 28A*, **cp28** produced concentration-dependent association and dissociation curves in the SAW sensorgram, thus clearly reporting a binding event between the peptide and EGF. The calculated affinity for the interaction ($K_D = 286 \pm 46 \mu\text{M}$) was about an order of magnitude stronger than that calculated for **d1**, thus validating our design concept of mimicking a particular region of EGFR.

Using ITC, we also evaluated the **cp28**-EGF interaction in aqueous solution, which constitutes a more native environment for the protein than the SAW sensor chip. As previously experienced for ligands **d1** and **d2**, this experiment resulted in relatively flat and featureless isotherms, due to the already discussed limitations of ITC for measuring low-affinity interactions (*Figure 28B*). Nevertheless, by fixing the stoichiometry value of the interaction to 1, we calculated a K_D value of 664 μM , which is in the same range as the results of the SAW experiment.

In the ^1H - ^{15}N HSQC spectrum of EGF, some of the signals of the protein were clearly modified upon the addition of **cp28** (1.5 mM), thereby allowing us to map the interacting region on EGF. Significant perturbations occurred on residues spanning a wide region of the protein (G12, V19, V34, Q43, D46, W50, L52), including the flexible C-terminus tail (*Figure 28C*).

Going a step further in this mimicry strategy, we envisaged some modifications on **cp28** with the ultimate goal to enhance its binding potency. We designed an analogue peptide (**cp28mut**), maintaining the same size and the disulfide bridge as in **cp28**, but introducing two conservative mutations in residues placed at the interface with EGF: K7R and F14W (*Figure 29*). Although these modifications had little effect on the overall structure (as shown by the very similar CD spectra of **cp28** and **cp28mut**), both NMR and SAW techniques showed that this analogue did not improve the affinity of the parent peptide.

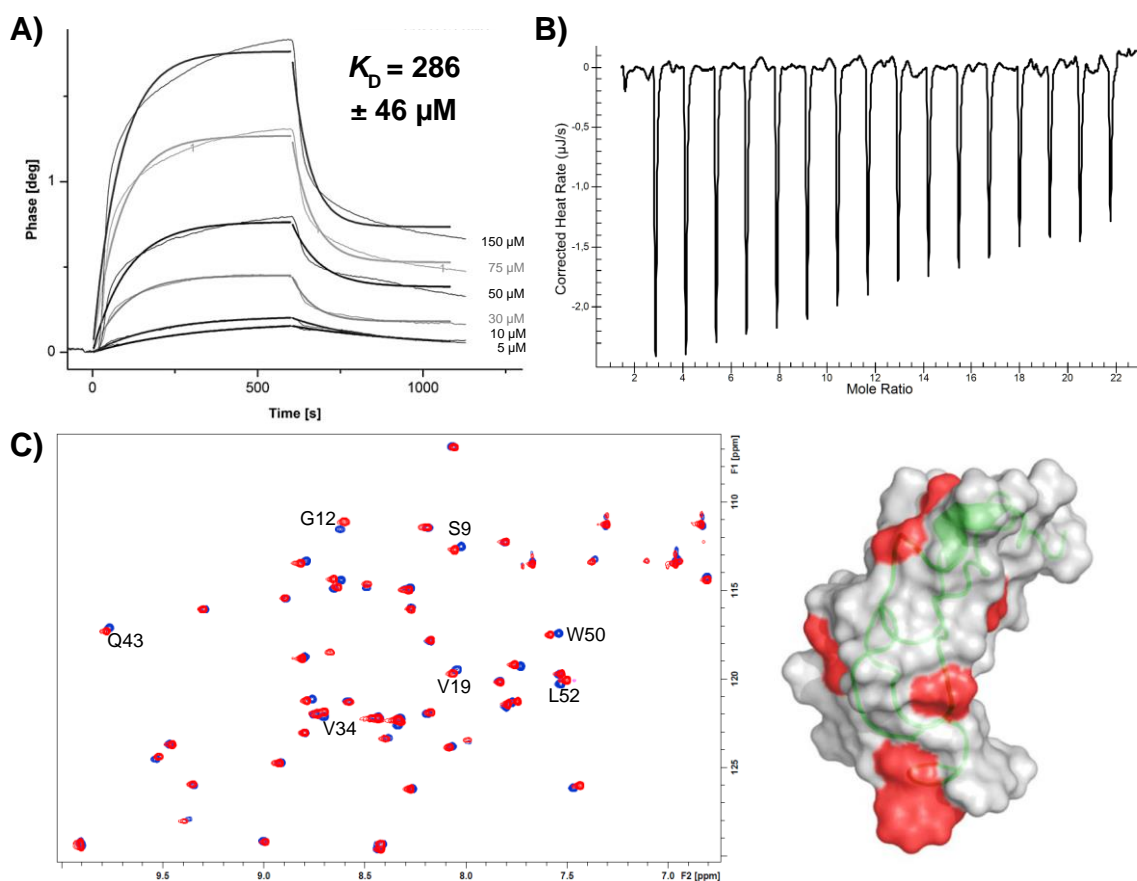


Figure 28. A) SAW sensorgram of **cp28** injected at a range of concentrations. B) ITC titration curve of **cp28** with EGF. C) ^1H - ^{15}N HSQC spectrum of EGF in the absence and presence of **cp28** (1.5 mM); residues with significant changes are labelled, and depicted in red in the surface representation of EGF.

Furthermore, to assess the stability and role of the helical structure in the interaction, we synthesized peptide **h**, which comprises the residues of the α -helix. Surprisingly for a linear peptide with no structural constraints, the CD spectra of **h** showed clear negative bands at 208 and 222 nm, typical of the α -helix conformation. As in the case of **cp28mut**, however, no significant binding of **h** to EGF was observed, thereby revealing that the α -helix alone is not sufficient to mediate the interaction with the protein.

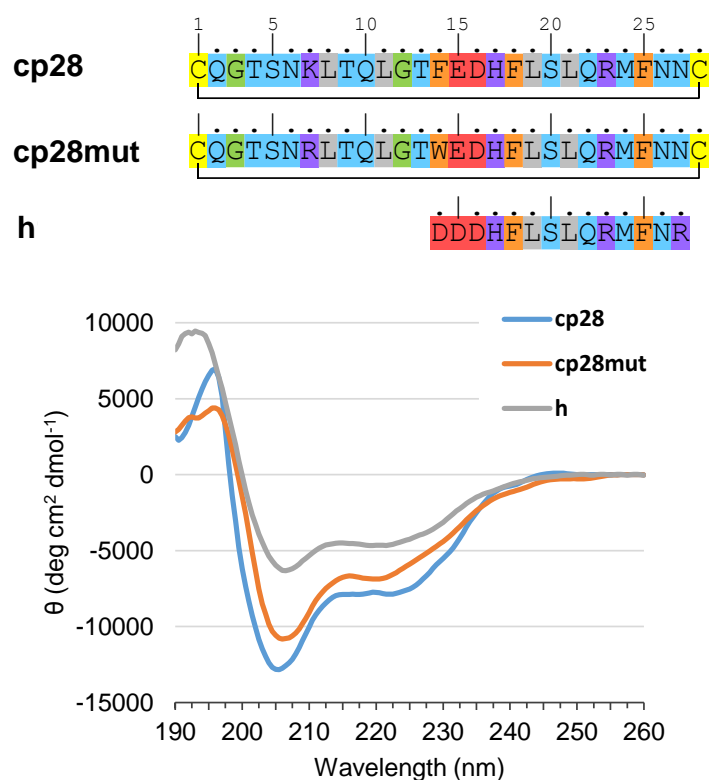


Figure 29. Amino acid sequence and CD spectra of peptides **cp28**, **cp28mut** and **h**, recorded at 20°C in phosphate buffer (pH = 6.8) with 10% TFE.

Mimicry of a β -sheet epitope of domain III of EGFR

After the encouraging results obtained with **cp28**, we were keen to explore the mimicry of other EGFR domains that may be relevant for EGF recognition. As discussed above, domain III of EGFR binds EGF second to domain I, albeit with a stronger affinity. The domain III binding interface comprises mainly hydrophobic interactions and involves several residues located along a four-stranded parallel β -sheet.

In this structure, the side chains of the alternate amino acids in the β -strands face opposite sides of the β -sheet, meaning that only one side of the sheet contributes to EGF recognition. The overall structure is stabilized by networks of hydrogen-bonded 12-membered rings, as well as by interstrand interactions between the side chains of neighboring residues.¹³¹

In particular, we focused our efforts on the mimicry of a two-stranded β -sheet segment of this region that comprises sequences GQFSLAVV and GDVIISKV (Figure 30). As mentioned above, some of these residues interact with the flexible C-terminal tail of EGF. In the crystal structure, Phe412, Ala415, Val417 and Ile438 in EGFR form a hydrophobic cavity that constitutes the Leu47 binding site and also interacts with the long aliphatic

portion of Arg45. Nearby, Lys443 forms a key ionic interaction with Glu51 of EGF.

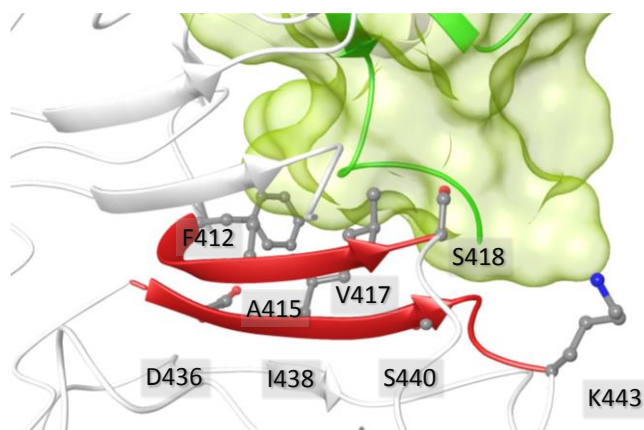


Figure 30. Close view of the four-stranded parallel β -sheet, part of EGFR domain III (PDB code: 1IVO). The two strands that were selected for peptide design are shown in red, and the key interacting residues are highlighted.

We decided at the beginning to mimic the individual strands with the corresponding linear peptides to test whether the interaction with EGF could already be detected. Although the synthesis of the peptides GQFSLAVV and GDVIIS proceeded smoothly, we could not undertake their purification or binding studies because they were highly insoluble in aqueous media, probably due to peptide aggregation. Consequently, we decided to explore the incorporation of artificial templates that would nucleate the formation of a β -sheet structure in water.

Unlike antiparallel β -sheet structures, which can be created by a single peptide chain containing artificial turn units (such as D-Pro-L-Pro) in the middle of the sequence, parallel-oriented strands must be linked either via their N-termini or their C-termini, or both. Thus, the formation of a parallel β -sheet is synthetically more demanding, as it requires the use of non-peptide units in the linking segment. In contrast to the variety of templates that promote antiparallel β -sheet folding, parallel β -sheet structures have been less explored.

We came across the work of Gellman *et al.* on parallel β -sheet mimetics, in which they describe the use of two non-peptide linkers to generate macrocycles that promote parallel strand interactions.^{132,133} A chiral linker containing D-proline and 1,2-diamino-1,1-dimethylethane (D-Pro-DADME) is used for C-to-C linkage; while, on the other side of the macrocycle, a flexible N-to-N linker is introduced (*Figure 31A*). Both linkers can be readily accessed via a few steps of synthesis in solution. The final macrocyclic

scaffold is reported to display substantial β -sheet conformational stability in solution.¹³²

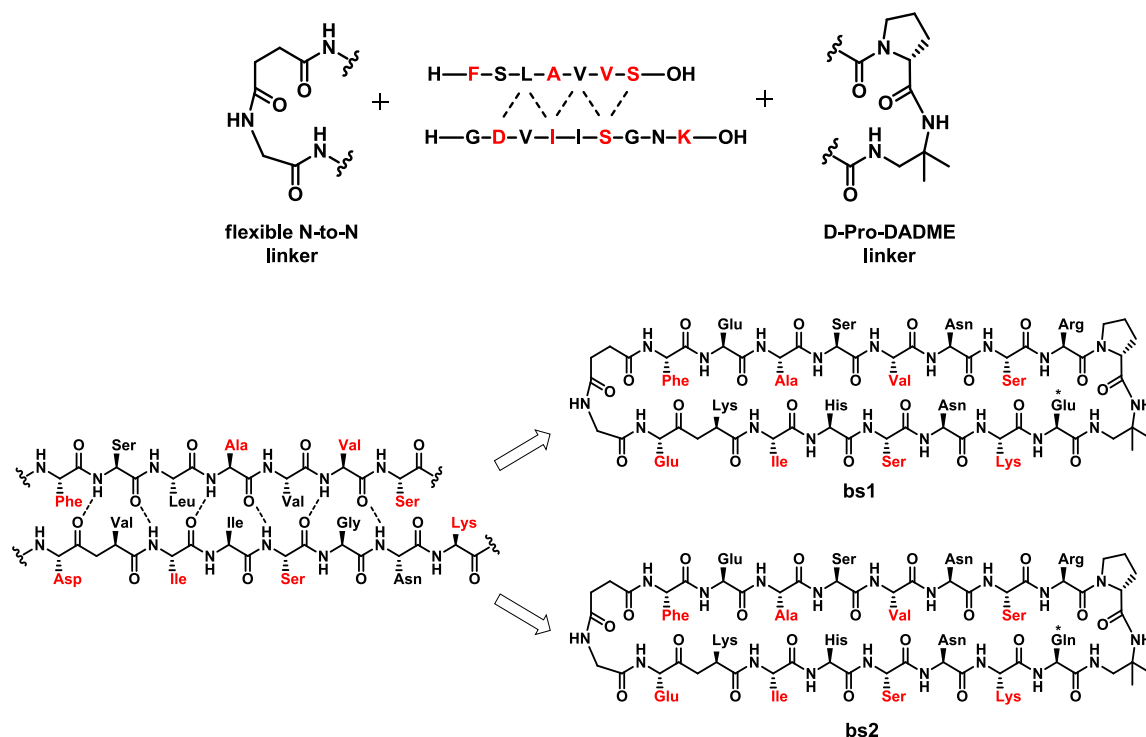


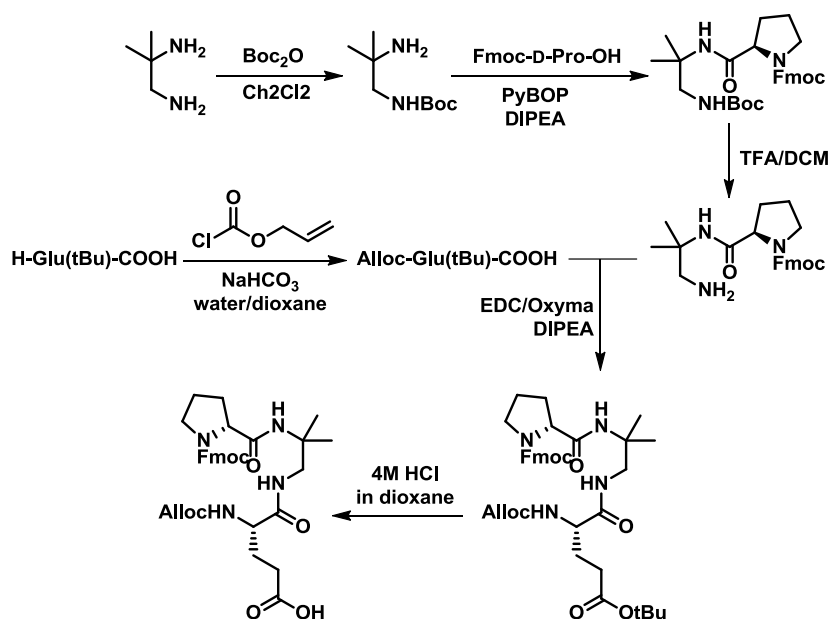
Figure 31. A) Elements of a designed parallel β -sheet mimetic of EGFR domain III. B) Sequence and hydrogen-bonding pattern of the two strands, which gave rise to the two macrocyclic β -sheet mimetics **bs1** and **bs2** (the anchoring side chain to the resin is marked with a star). Residues in red were those with their side chains projecting towards the EGF interface.

Given this precedent, we decided to apply a similar strategy for the mimicry of this specific discontinuous β -sheet epitope of EGFR. From the two strands forming our macrocyclic template, we used the same interacting residues as those in EGFR, with the only exception of Asp, which was mutated to Glu. In this position, a longer side chain appeared to achieve a better interaction with EGF. For the non-interacting residues, we did not follow the sequence of this EGFR epitope, as it was highly hydrophobic, and selected them instead based on several considerations:

(1) As we had experienced, peptides forming β -sheets are often prone to aggregation; therefore charged and polar residues were incorporated in the non-interacting face of the β -sheet, so that our molecule would bear a net positive charge at physiological pH. (2) We selected residues with a high propensity to participate in β -sheet structures and, whenever possible, we paired them in the most favorable way. Regarding last point, pairs of oppositely charged residues strongly favor β -sheet conformation, especially

when the negatively charged residue is on the hydrogen-bonded position of the pair. As well as acid-base pairs, the Asn-Asn pair is known to strongly favor parallel β -sheet formation, due to the formation of stable hydrogen bonds when their side chains are in their preferred conformation.¹³⁴

Therefore, we sought to condense this intricate structural data into a macrocyclic peptide (**bs1**) that would be able to mimic a singular binding epitope of EGFR (*Figure 31B*). The synthesis of cyclic peptide started with the preparation in solution of the Alloc-protected D-Pro-DADME fragment, following the reported procedure (*Scheme 3*).¹³² This fragment was then attached to a 4-(Hydroxymethyl)benzoyl-aminomethyl (HMBA-AM) polystyrene resin via the Glu side chain, and residues corresponding to one of the strands (Phe to Lys) were progressively appended via standard Fmoc-based SPPS. The N-to-N linker was also coupled to the Phe residue on solid phase, using standard conditions. Then, several attempts to remove the Alloc protecting group were made, using PhSiH_3 and $\text{Pd}(\text{PPh}_3)_4$ as catalyst. However, no deprotection occurred even when adding more equivalents of catalyst and increasing the reaction temperature to 80°C . It could be hypothesized that the bulkiness of the peptide chain attached to the resin prevented the action of the Pd catalyst on the Alloc group.



Scheme 3. Synthesis of Fmoc-D-Pro-Dadme-Glu(OH)-Alloc, which was used as an artificial β -turn C-to-C linker in peptides **bs1** and **bs2**.

To address this problem, we envisaged changing the solid support to ChemMatrix, which is known to have a higher swelling capacity than polystyrene supports. Due to the RinkAmide linker used on the resin, the anchoring Glu residue became Gln upon cleavage (peptide **bs2**). However,

this modification occurred on the non-interacting side of the β -sheet, and it thus should not impact the binding of the peptide. By using this solid support, the Alloc deprotection proceeded smoothly under the conditions reported. Then the amino acids of the second strand were appended via standard Fmoc-based solid-phase synthesis. After attachment of the last residue (Gly) and removal of the silyl protecting group, the peptide was cyclized on-resin using PyBOP/HOAt. Finally, the cyclic structure was cleaved from the resin and purified using the conventional procedure.

Although the synthesis and purification of this β -sheet macrocycle was completed successfully, the CD spectrum of the peptide in water failed to reveal a β -sheet signature (*Figure 32*). Instead, the spectra presented a minimum at 200 nm, which indicates a predominantly random coil state. Not surprisingly, no relevant interaction with EGF was observed by either NMR or SAW, possibly because the peptide lacked secondary structure.

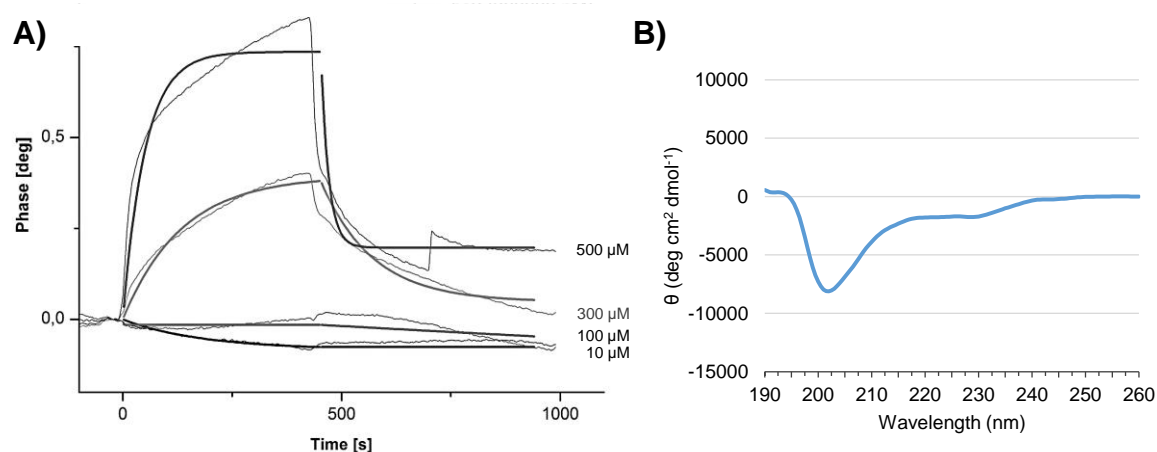


Figure 32. A) SAW sensorgram of *bs2* injected at a range of concentrations (no K_D could be determined). B) CD spectra of peptide *bs2* at 20°C in phosphate buffer (pH = 6.8) with 10% TFE.

Bicyclic peptides as improved analogues of **cp28**

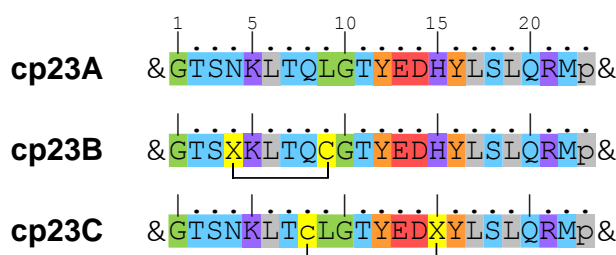
First generation bicyclic peptides

Taking the ensemble of results into consideration, we focused our efforts on improving the affinity and drug-like properties of **cp28**, as this was the only molecule that had consistently shown the capacity to bind to our target protein, EGF. Although **cp28** is a cyclic peptide, its large size is translated into elevated conformational flexibility, which may be detrimental for activity due to the high entropic penalty of adopting the bioactive conformation. High structural flexibility, combined with the presence of natural amino acids, is the ideal recipe for proteolytic degradation and poor stability in physiologic media. In addition, disulfide-bridged peptides are intrinsically unstable in reducing environments, thereby also hampering their *in vivo* applicability. Last, but not least, the hydrophobic character of some parts of the **cp28** sequence (three Leu, three Phe and a Met, among others) was the root cause of its low solubility in water. Overall, these were all limiting factors for the future development of this peptide, either as a chemical biology tool or as a potential drug, which needed to be addressed.

Using a rational design approach, we conceived two analogues of **cp28** (named **cp23A** and **cp23B**) that included the following changes: (1) the non-interacting region—²⁵FNNC²⁸—C¹—would be removed, thus shortening the overall size of the peptide; in its place, a D-Pro-Gly motif would be introduced to force the turn in the peptide chain; (2) the two Phe residues on the ‘back’ part of the α -helix, which do not contribute to the binding interface with EGF, would be mutated to Tyr, with the aim to improve the overall solubility of the construct; and (3) the disulfide bridge would be replaced by a more stable head-to-tail covalent bond, through a D-Pro-Gly motif (*Scheme 4*).

In addition to all these modifications, an extra constraint, namely a disulfide bridge, would be introduced into **cp23B**, between residues 6 and 11, in order to decrease the flexibility on this region of the peptide. For the introduction of this disulfide bridge, we explored several side chain lengths and C _{α} chirality, and selected those that seemed to better preserve the original structure of **cp28**. As a result, L-homocysteine and L-cysteine were selected to replace Asn6 and Leu11, respectively, resulting in the bicyclic peptide **cp23B** (*Scheme 4*). A practical advantage of this choice is the possibility to directly compare the cyclic and bicyclic forms of the peptide by simply reducing the intramolecular disulfide bond.

Finally, we envisaged a third analogue (**cp23C**) with a different topology for the disulfide bridge. In this peptide, we would link one of the positions in the α -helix (His15) with a non-interacting residue in the loop of the peptide (Gln8), with the objective to constrain the mobility of this loop and obtain a more “receptor-like” conformation. On the basis of the 3D structure of the peptide, we selected L-homocysteine and D-cysteine as the optimal connectivity for the cyclic constraint.

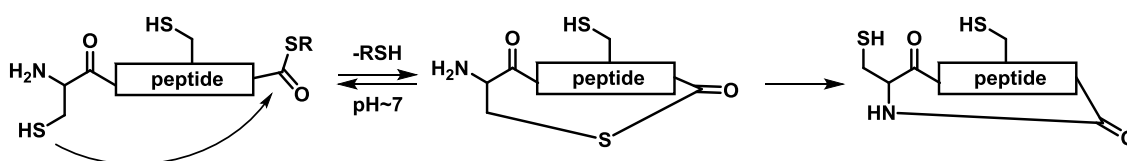


Scheme 4. Amino acid sequence of peptides **cp23A**, **cp23B** and **cp23C** ($X = L$ -homocysteine).

For the synthesis of **cp23A**, we first attempted to perform the head-to-tail cyclization of the fully protected peptide in solution, but all the conditions that we tested failed to yield the cyclic product. As an alternative, we turned to a solid-phase approach using Kenner’s safety-catch linker,¹³⁵ which allows the cyclization of the peptide with concomitant cleavage from the solid support. This methodology is reported to eliminate the risk of cyclodimerization or polymerization and has been used to synthesize cyclic decapeptides in good yields.¹³⁶ In our hands, however, no cyclic peptide was isolated after the final cyclization/cleavage treatment.

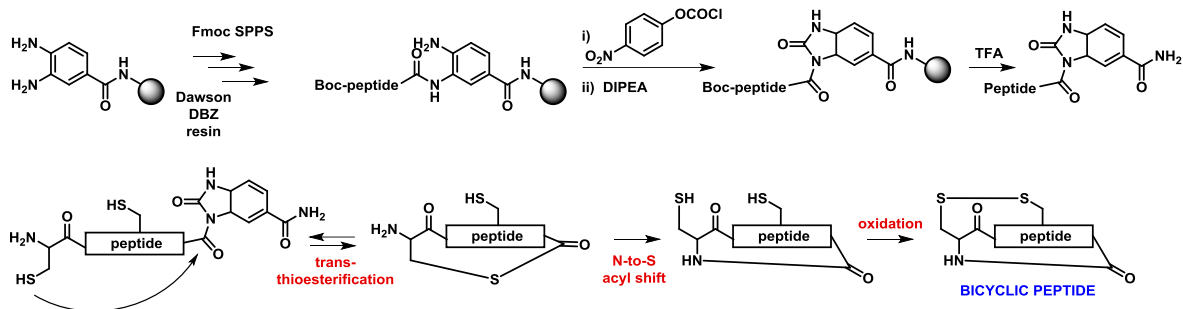
As a third strategy to obtain **cp23A**, we followed a parallel approach to that used for the synthesis of the β -sheet mimetic **bs1**. This approach relied on side-chain anchoring of the peptide through a Glu residue and final cyclization on the solid phase. First, we prepared the tripeptide fragment Fmoc-Glu(OH)-Leu-Gly-OAll, which was attached to a ChemMatrix RinkAmide support through the side chain of Glu. The peptide chain was then elongated via standard Fmoc-based synthesis. After the last residue (Thr) was incorporated, the C-terminal Alloc protecting group was removed and the cyclization was attempted using PyBOP/HOAt. However, little conversion to the cyclic product was achieved, even after 4 cycles of reaction were performed. Probably, the conformation of this 23-mer peptide attached to the resin was not favorable for the intramolecular reaction. Nevertheless, a sufficient amount of the product was isolated from the reaction crude product for further testing.

For the synthesis of **cp23B** and **cp23C**, we took advantage of the native chemical ligation (NCL) reaction, introduced in 1994 by the Kent laboratory. This reaction involves a N-terminal cysteine and a C-terminal thioester that react to form a native amide bond,¹³⁷ and it was originally developed for the ligation of peptide segments, as a way to facilitate the chemical synthesis of proteins. By introducing the N-terminal cysteine and the thioester into the same peptide sequence, a circular peptide can be produced (*Scheme 5*). Indeed, this methodology has proved highly effective for the preparation of several naturally occurring and artificial cyclic peptides.¹³⁸



Scheme 5. The native chemical ligation (NCL) reaction takes place in an unprotected peptide between the thiol moiety of the N-terminal cysteine and the C-terminal thioester. The first and rate-determining step is a reversible transthioesterification, which thermodynamically favors the intramolecular reaction. Under the same conditions, the intermediate thioester undergoes a rapid S→N acyl intramolecular rearrangement to form the more stable amide bond. Note that additional unprotected cysteine residues in the peptide do not interfere in the reaction.

As an alternative to the preparation of the C-terminal peptide thioester, which requires the use of special Boc-based protocols, we followed Blanco's methodology for the formation of a C-terminal N-acylurea.¹³⁹ This acylating moiety functions as a thioester precursor and, importantly, behaves as a stable amide during the process of chain assembly and cleavage, which can be done using standard Fmoc-SPPS conditions. Treatment of the unprotected peptide in solution with a thiol additive, such as 4-hydroxybenzenethiol, then catalyzes *in situ* the intramolecular NCL, yielding the desired cyclic peptide (*Scheme 6*). Of note, the thioester conversion occurred at pH 7.0 in under an hour, with no polymeric species observed.¹³⁹ To complete the synthesis of our bicyclic peptides, the disulfide bond between Cys and homocysteine residues was formed in solution under oxidative conditions.



Scheme 6. Synthetic strategy for the synthesis of the bicyclic peptides. The peptide chain is grown on one of the amines of the DBZ linker. The N-terminal residue is introduced as a Boc-amino acid. Prior to peptide cleavage, the linker is activated via cyclization. The intramolecular NCL reaction yields the head-to-tail cyclic peptide in aqueous solution. Finally, the thiols are oxidized to form the final bicyclic molecule.

The CD spectra of the cyclic peptide **cp23A** showed a sharp negative band at ca. 208 nm, and a second minimum at ca. 222 nm, consistent with a high population of α -helical secondary structure (Figure 33). Regarding the bicyclic peptides, similar structural traits were observed in the **cp23B** spectra, thereby showing that the introduction of the second cyclic constraint does not affect the overall conformation of the peptide in solution. Indeed, the two peptides shared a similar CD spectrum as that of the parent peptide **cp28**. In contrast, the alternative cyclic topology that we tested in peptide **cp23C** led to a distortion in the overall secondary structure of the peptide, as can be inferred from the lack of the typical α -helical bands. Instead, the negative minimum at ca. 200 nm is indicative of a predominantly random coil structure.

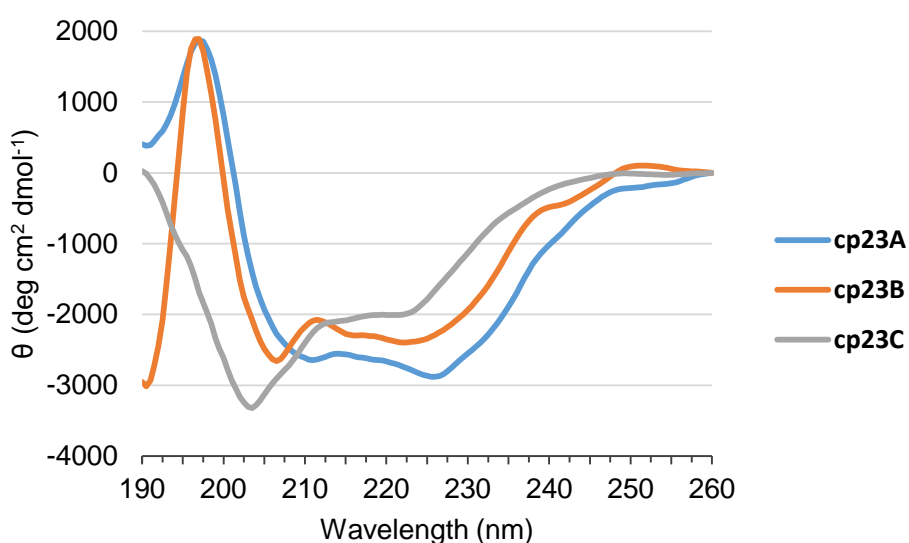


Figure 33. CD spectra of peptides **cp23A**, **cp23B** and **cp23C** recorded at 20°C in phosphate buffer (pH = 6.8) with 10% TFE.

As for the previous inhibitors, the binding of the three new analogues (plus the reduced version of **cp23B**) to EGF was assessed by recording the ^1H - ^{15}N HSQC spectra of the protein in the absence and in the presence of ligand (1 mM). The induced chemical-shift changes indicated that the three analogues bind to the same region of the protein. In fact, the most affected residues (L8, S9, G12, H16-V19) cluster on the same binding epitope as **cp28**.

By calculating the average weighted ^1H - ^{15}N chemical-shift differences for all analogues, we clearly observed larger differences in the oxidized form of **cp23B** compared to monocyclic analogues **cp23A**, **cp23B** (reduced) and the less-structured **cp23C** (Figure 34). This result shows how a more restrained scaffold, if appropriately selected for stabilizing the bioactive conformation, can enhance the molecular recognition of the target protein.

The trends observed by NMR spectroscopy were confirmed in the SAW binding assay. In this experiment, all analogues produced concentration-dependent responses in the SAW sensorgram, but it was the bicyclic peptide **cp23B** that displayed lower affinity values for the interaction (Figure 35, $K_D = 575 \pm 93 \mu\text{M}$). Although the affinity of **cp23B** was slightly weaker than that of the parent peptide **cp28**, the straightforward synthetic procedure, together with the unique features of its bicyclic scaffold (e.g. higher stability *in vivo*), highlighted the opportunity to explore optimized analogues based on this candidate.

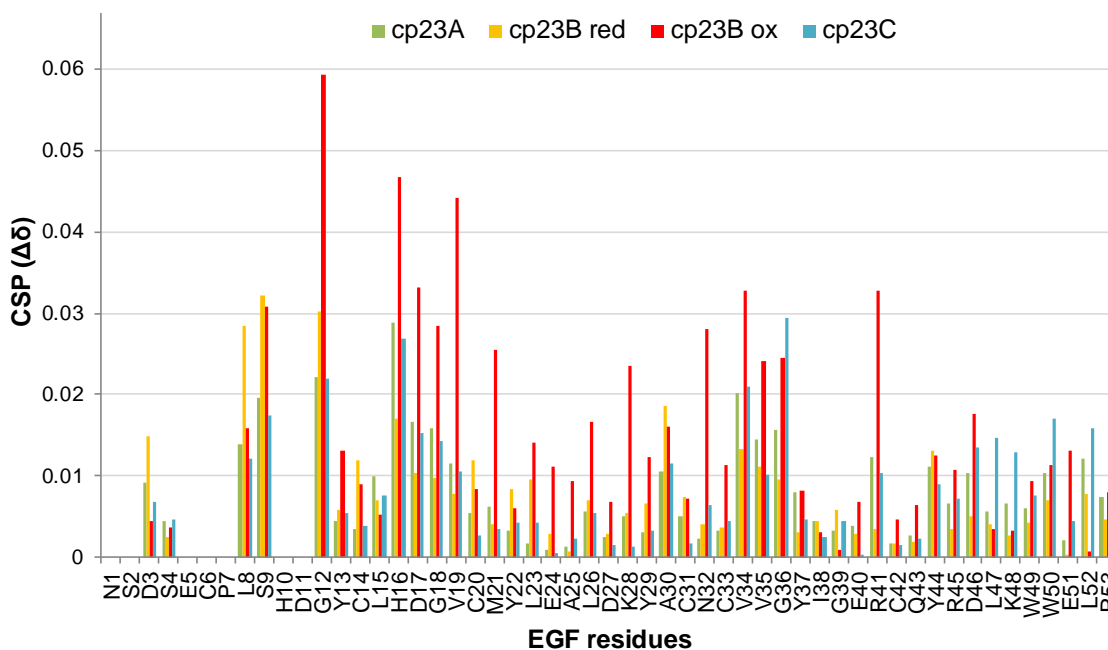


Figure 34. Bar graphs of the average chemical shift perturbations (CSPs) induced by peptides **cp23A**, **cp23B** (reduced and oxidized) and **cp23C**, as a function of EGF sequence.

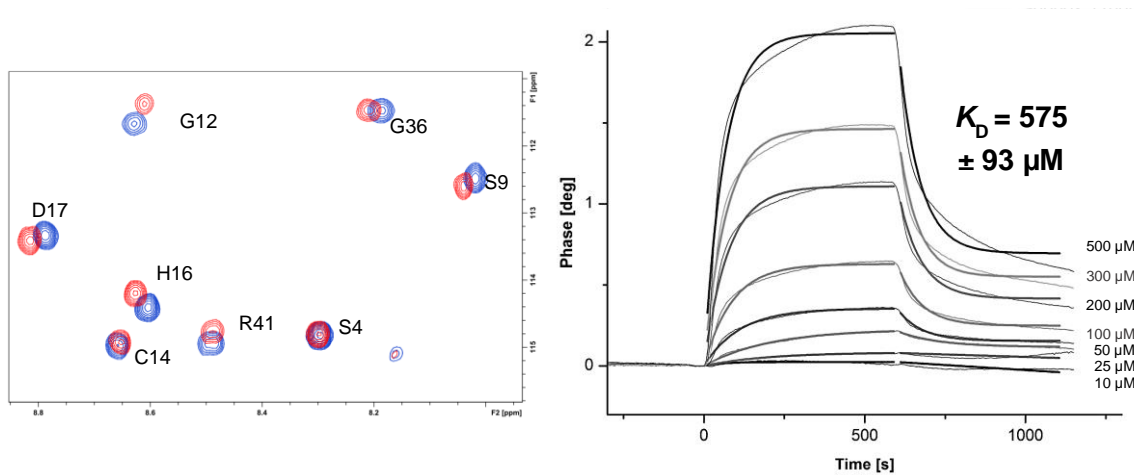


Figure 35. Detail of the ^1H - ^{15}N HSQC spectrum of EGF in the absence and presence of **cp23B** (1 mM); residues with significant changes are labelled (left). SAW sensorgram of **cp23B** injected at a range of concentrations (right).

Computational methods for lead optimization

Second generation bicyclic peptides

The main objective in structure-based peptide design is to find specific topologies that can bind to a target of interest. In recent years, a plethora of computational methods have been developed with the aim to reduce the time and cost associated with experimental design campaigns. Docking and scoring, which predict the binding mode of a drug and its estimated affinity, are by far the computational methods most widely used in drug design. . However, virtual docking tools have the drawback of inadequate scoring functions, which treat the protein as a rigid or semi-rigid construct.¹⁴⁰ Consequently, as we saw in the first chapter, a major problem with docking is the poor correlation with experimental binding data—a problem that can be exacerbated when flexible proteins such as EGF are involved.

With the aim to estimate the affinity of ligands, as well as to dissect the key interactions that contribute to the binding process, various *in silico* algorithms have been developed that take into account the flexibility of proteins and the induced fit of ligands. Some of these algorithms are based on molecular dynamics (MD) simulations, which are able to extensively sample the interacting complex in a solution environment. End-point free-energy methods, such as molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) and generalized Born surface area (MM/GBSA), are well-established methods to predict the free energy of binding (ΔG) of ligand-protein interactions.¹⁴¹ These two methods differ only in the mode of calculation used to estimate the solvent contribution to ΔG . In general terms, the MM/PBSA method yields more accurate absolute free-energy values; however, the MM/GBSA method is computationally more efficient and regarded as superior at estimating relative differences between analogues in the set.^{142,143}

In the previous section, we showed how we could condense the features of peptide **cp28** into a smaller, more rigid and synthetically accessible bicyclic scaffold (**cp23B**), without compromising its binding activity to EGF. Here, this constrained scaffold will serve as the starting point of an *in silico* structure-based design optimization process, which pursues the discovery of new bicyclic analogues with enhanced binding affinity to EGF.

First, we conducted a 50-ns free MD simulation of the **cp23B**-EGF complex. The RMSD for the peptide was low at the beginning of the simulation, indicating a stable interaction with the protein (*Figure 36*).

However, during the second half of the simulation, there was a gradual deviation in the peptide backbone, presumably caused by alterations in the peptide structure and dissociation from the protein. This observation is in agreement with the analysis of some key interactions, as well as the center-of-mass distance between the two molecules, along the collected trajectory. Whereas the Cys27-Gln8 hydrogen-bond interaction was stable, the key Glu36-Lys5 hydrogen bond was mostly absent during the last part of the simulation. Also, important fluctuations were observed in the center-of-mass distance between the two molecules, evidencing a weak interaction of **cp23B** with EGF. From 500 equally spaced snapshots, taken during the first 20 ns of the simulation, we were able to calculate the mean free energy of binding (ΔG) for the **cp23B**-EGF interaction, which was -31.4 kcal/mol according to the MM/GBSA method.

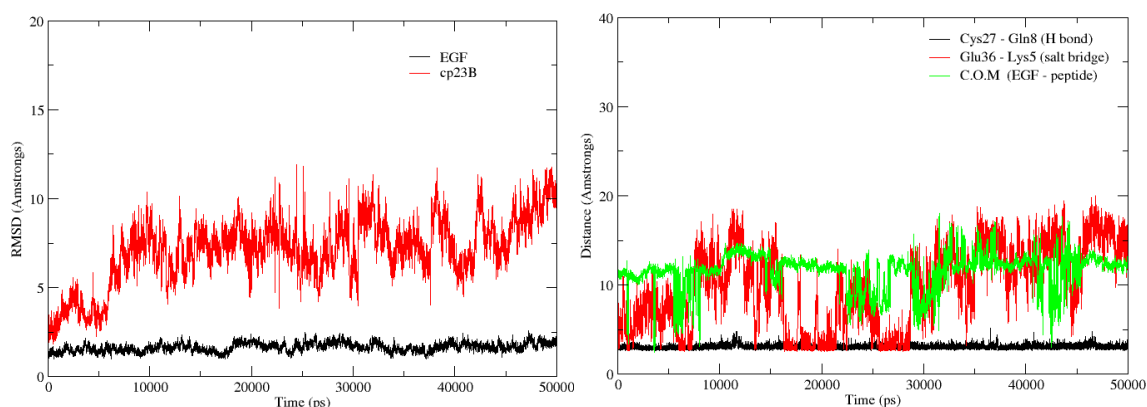


Figure 36. The RMSD values during the MD trajectory of **cp23B** (in red) and EGF (in black), calculated for the backbone atoms (left). B) Distance metrics of two key intermolecular interactions (Cys27-Gln8 and Glu36-Lys5), as well as the center-of-mass distance between the two molecules (right).

We then performed virtual point mutations on the sequence of **cp23B**, with the goal to enhance some of the interactions that we had visualized during the MD trajectory. Several mutations were tested for Met22, as the side chain of this methionine became frequently oxidized at different stages of the synthesis. As well as pursuing a stronger affinity, a major aim at this stage of design was to endow our peptide with a more stable secondary structure. In some of the simulations with **cp28** and **cp23B**, we had observed that the α -helical motif of the peptide tended to unwind, especially towards the C-terminal end of the helix. To tackle this issue, we explored the introduction of a Pro residue at different positions of the C-terminal end of the helix. In total, we generated 28 peptide structures (Table 7), which were placed in complex with EGF and subjected to free MD simulations.

From these simulations, regular snapshots were taken to calculate the theoretical mean ΔG values, using both GBSA and PBSA approximations.

By comparing the values obtained for each of the mutated analogues, we were able to assess the effect of the new mutation on the interaction with EGF. The absolute free-energy values differed considerably depending on the method used, GBSA giving lower absolute energy values (*Table 7*). In this regard, it has been reported that calculated affinities strongly depend on the parameters used in the algorithm, such as the solvation method, the charges, and the dielectric constant.¹⁴⁰ Nevertheless, the relative differences for a given ligand showed an acceptable level of correlation ($r^2 = 0.52$) when both methods are considered.

Marked differences in ΔG were found along the series of **cp23B** analogues. Sometimes minor mutations caused surprising effects on the theoretical affinities. For instance, the L17A mutation increased the relative free energy of binding by 6 kcal/mol. On the other hand, the most advantageous set of mutations, such as Y12P_K5A_L6E_M22Nva, resulted in up to 15 kcal/mol gain in affinity. In this particular case, the introduction of an acidic residue in the loop region of the peptide (L6E) produced a significant increase in binding potency. The mutation of the adjacent Lys to Ala (K5A) further enhanced this positive effect, due to the removal of the intramolecular competing salt bridge.

As already discussed, we also sought to restrain the peptide conformation by stabilizing the α -helical domain. For this purpose, we explored the effect of introducing a Pro residue on the C-terminal end of the helix. As depicted by the ΔG values of the different Pro analogues, the most favorable position to introduce the proline ring was in the 12th residue, substituting a Tyr. In most cases, the Y \rightarrow P mutation generated an 8-10 kcal/mol boost of interaction energy (*Table 7*). Along with these effects, we found that a conservative mutation of Met22 to norvaline (Nva) yielded good results in replacing its oxidation-prone side chain. This observation thus reveals the importance of keeping a hydrophobic contact in this interaction site.

Analogues	MM/GBSA (kcal/mol)	MM/PBSA (kcal/mol)	Avg. RMSD peptide (Å)
Y12P_H15D_polyA (cp23D)	-46.8	-12.7	3.3
Y12P_K5A_L6E_M22Nva (cp23E)	-46.6	-9.2	3.1
Y12P_M22Nva (cp23F)	-42.1	-2.9	3.6
Y12P	-42.0	-3.7	3.8
Y12P_L6E	-41.2	-3.0	3.7
Y12P_K5A_M22Nva	-41.0	-6.0	3.3
M22Q	-40.1	-1.2	3.2
L6E	-39.7	1.0	4.6
Y12P_L6E_T11P	-35.7	4.0	3.4
Y12P_M22Orn	-34.2	0.2	4.1
T11P_M22Nle	-33.8	9.9	5.3
T11P	-32.3	4.6	4.0
T11P_D14L_M22Orn	-31.9	1.4	4.6
cp23B	-31.4	6.8	5.5
M22K	-29.4	14.3	3.9
Y12P_L6E_E13A	-29.0	2.1	4.1
Y12P_polyA	-29.0	11.5	5.9
Y12P_K5A_L6E_H15D_M22Nva_polyA	-28.8	-0.3	5.2
M22E	-27.9	4.8	4.5
T11P_M22Orn	-27.0	2.2	5.6
Y12P_L6E_E13P	-26.5	10.7	3.5
L17A	-25.5	8.7	5.0
Y12P_M22Q	-25.2	3.5	4.2
E13P	-24.8	5.2	5.3
Y12P_K5A_L6E	-24.2	3.9	5.9
Y12P_M22Dab	-23.0	6.8	4.0
T11P_M22K	-22.9	3.7	7.8
K5A_L6E_M22Nva	-21.1	5.8	7.5

Table 7. Mean backbone RMSD values and relative free energies (ΔG), using either PB or GB solvation methods, for all the designed analogues. Dap = diaminopropionic acid; Dab = diaminobutyric acid; Orn = ornithine; Nva = norvaline; Nle = norleucine. PolyA refers to multiple mutations to alanine performed on the non-interacting residues of the α -helix: E13A, Y16A, L17A, L19A and Q20A.

Analysis of the simulation of this last analogue (Y12P_M22Nva, **cp23F**) with EGF showed a much more stable interaction than that simulated above for **cp23B**, with minimal peptide RMSD alterations and highly stable side chain interactions (*Figure 37*).

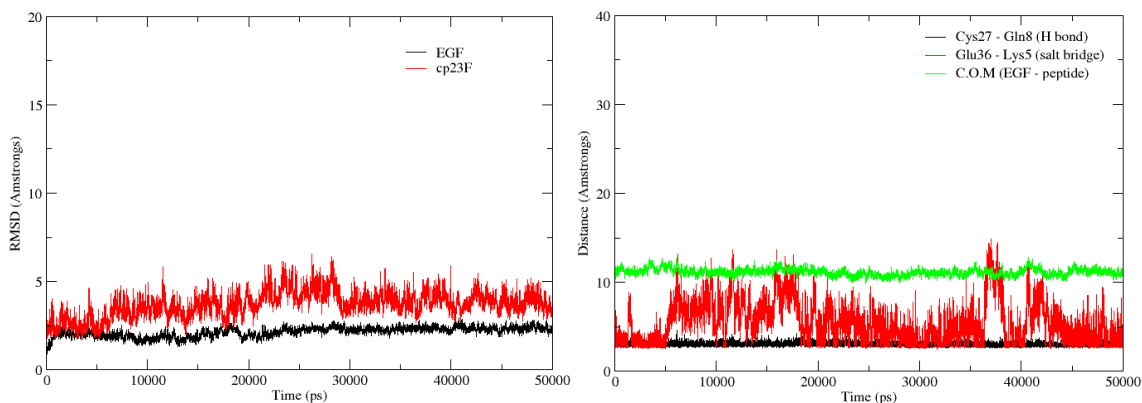
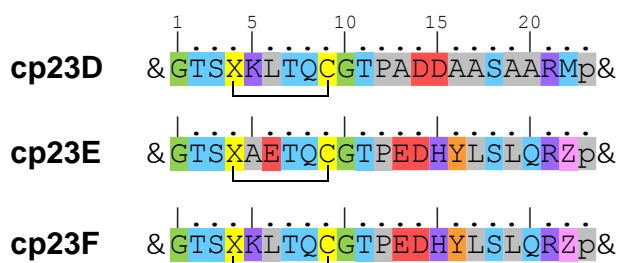


Figure 37. The RMSD values during the MD trajectory of **cp23F** (in red) and *EGF* (in black), calculated for the backbone atoms (left). B) Distance metrics of two key intermolecular interactions (*Cys27-Gln8* and *Glu36-Lys5*), as well as the center-of-mass distance between the two molecules (right).

With the results obtained in this computational campaign, we selected the three **cp23B** analogues with the best affinity values, so as to proceed with their chemical synthesis and biophysical evaluation: Y12P_polyA (**cp23D**), Y12P_K5A_L6E_M22Nva (**cp23E**) and Y12P_M22Nva (**cp23F**); see *Scheme 7*.

The synthetic methodology described above for the synthesis of **cp23B** was also applied in these cases. Namely, the linear peptide was synthesized with a reactive N-acylurea functionality on the C-terminus. An intramolecular native chemical ligation reaction, which formed the head-to-tail cyclic product, was then followed by an oxidative cyclization, between the cysteine and homocysteine side chains, to yield the bicyclic peptide. The three analogues were obtained in a milligram scale, although more complications were encountered for the synthesis of **cp23D** and **cp23E**, due to slow coupling rates for some of the mutated amino acids.



Scheme 7. Amino acid sequence of the designed bicyclic peptides (*X* = *L*-homocysteine; *Z* = *L*-norvaline).

The structural features of this new set of analogues were evaluated, first computationally by replica-exchange MD and then experimentally by CD. For the *in silico* studies, the initial structures of **cp23B** and **cp23F** were subjected to 200 ns of REMD at temperatures ranging from 300 to 600 K. The most populated clusters of conformations, for both peptides, conserved the α -helix motif that is originally found in the domain I of EGFR (*Figure 38*).

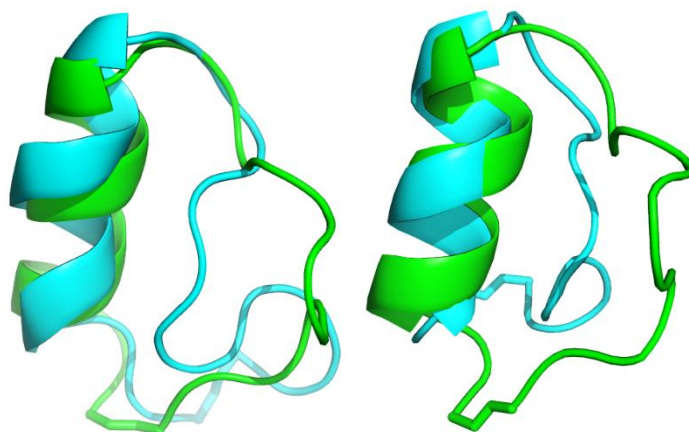


Figure 38. The most populated clusters of conformations for **cp23B** (left, 27% of all conformations, backbone RMSD of 4.8 Å) and **cp23F** (right, 58% of all conformations, backbone RMSD of 5.0 Å) during the REMD experiment (in blue cartoon) are superimposed with their respective initial structure (in green, coordinates extracted from PDB 1IVO). Note how the extra Pro residue in peptide **cp23F** yields a more structured helical motif.

For peptide **cp23F**, more than half of the sampled conformations were clustered into a single group, which featured an RMSD of 5.0 Å relative to the starting “receptor-like” structure (*Figure 39*). Peptide **cp23B**, in contrast, visited and populated a higher number of conformations, as could be expected from its less restricted nature. From these observations, we can thus deduce that the addition of a Pro residue at the C-terminal end of the α -helix effectively stabilizes the peptide scaffold. However, RMSD values in the order of 5.0 Å were obtained for both analogues, mostly due to the large fluctuations of the flexible loop of the peptide.

No substantial differences were found between the CD spectra of the three analogues (*Figure 39A*), thereby indicating that the bicyclic scaffold is well-structured and can tolerate a certain degree of sequence modification without undergoing major conformational changes.

The activity of this new set of analogues was evaluated by SAW and NMR. In the ^1H - ^{15}N HSQC spectrum of EGF, all three peptides produced clear

perturbations on some of the NMR signals of the protein (*Figure 39B*). As expected, the most affected residues (L8, S9, G12, H16-V19) matched the reported epitope of peptide **cp23B**, thus confirming the same mode of binding for these peptides. Due to the minor differences in CSP found between the three analogues, we resorted to the use of the SAW methodology for quantifying the affinity values. In this experiment, we obtained well-defined association and dissociation curves for peptide **cp23F**, which allowed for an accurate measurement of the K_D value of the interaction ($K_D = 279 \pm 79 \mu\text{M}$, see *Figure 39C*). In contrast, the results of the other two analogues (**cp23D** and **cp23E**) in the SAW assay were not as satisfactory. They produced weak sensorgrams that translated into imprecise fittings of the binding curves and poor K_D values.

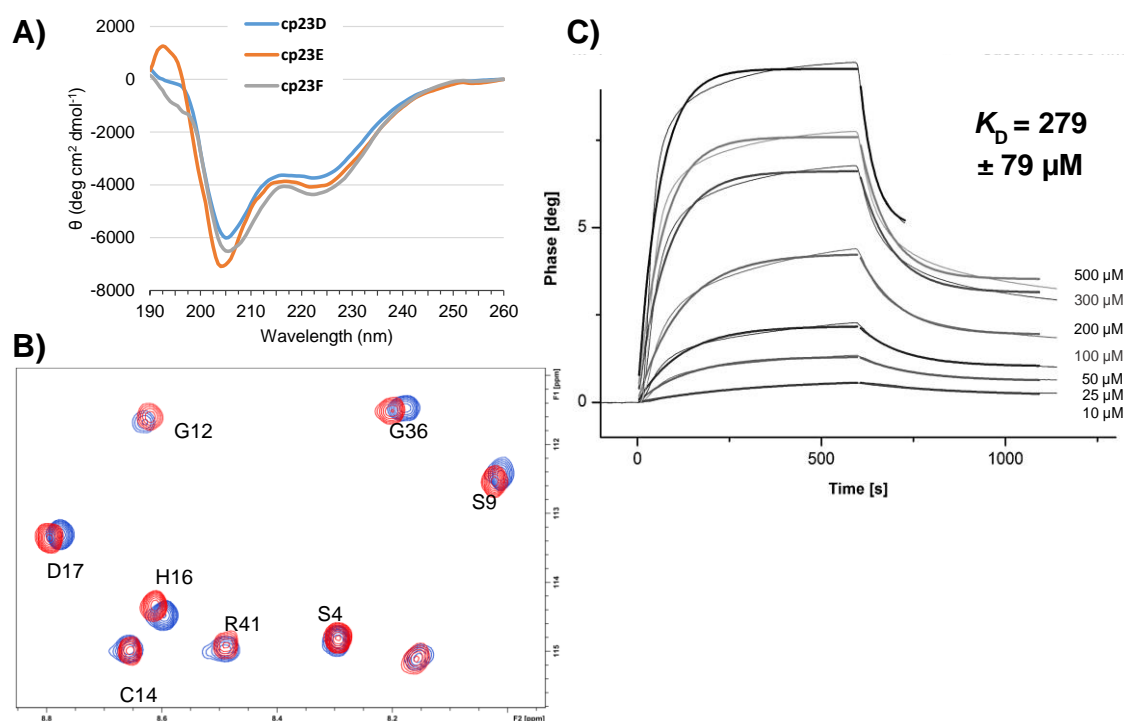


Figure 39. A) CD spectra of peptides **cp23D**, **cp23E** and **cp23F** recorded at 20°C in phosphate buffer (pH = 6.8) with 10% TFE. B) Detail of the ¹H-¹⁵N HSQC spectrum of EGF in the absence and presence of **cp23F** (1 mM); residues with significant changes are labelled. C) SAW sensorgram of **cp23F** injected at a range of concentrations.

Third generation bicyclic peptides

Given the foregoing results, we implemented a final cycle of computational modeling for the optimization of peptide **cp23F**. Briefly, we generated a set of heterochiral mutants that shared the key features of **cp23F**, but incorporated new L- and D-amino acids in some of the interacting positions

along the sequence. Again, the complex of these new analogues with EGF was subjected to free MD simulation, and the theoretical binding free energies (ΔG) were calculated for each peptide-EGF complex (*Table 8*).

As a matter of fact, most new mutations had a detrimental effect on the theoretical binding energies with EGF (*Table 8*). Still, the replacement of Gly10 for larger apolar residues, such as Met or Tyr, improved the negative score of **cp23F**. This can be explained by the hydrophobic interactions that these residues establish with Val29 and Val30, which form part of the B loop of EGF. Moreover, the introduction of D-amino acids in the place of Thr2 led to more potent binding *in silico*, with the T2r mutant having the lowest MM/PBSA score of the entire set of compounds (-8.1 kcal/mol).

As we can observe in the detailed analysis of the **cp23G**-EGF interaction, minimal fluctuations in the backbone of the peptide and protein occurred during the simulation. The key interactions, such as the backbone hydrogen bonding between Cys27 (EGF) and Gln8, were stable throughout the simulation time. Importantly, the guanidine group of the newly introduced D-Arg residue formed a highly stable salt bridge with the carboxylate of Asp42 (EGF). In addition, as we found by REMD simulation, the D-Arg side chain preferentially pointed towards the middle of the peptide, with the hydrocarbon part of its chain partly filling the void volume at the center of the cyclic peptide. Otherwise, this void space would produce the collapse of the loop towards the α -helix, thus disrupting the overall conformation, as we saw in the most visited structures of **cp23F** (*Figure 41*).

Consequently, we decided to pursue the synthesis of the best-fitted heterochiral candidates: **cp23G** (T2r), **cp23H** (G10M) and **cp23I** (T2r_G10M). Of note, all three peptides had the prototypical α -helical signature on their CD spectra, indicating that, despite the new mutations on different sites of the sequence, their overall secondary structure was preserved (*Figure 42*).

Analogues (based on cp23F)	MM/GBSA (kcal/mol)	MM/PBSA (kcal/mol)	Avg. RMSD peptide (Å)
G10M (cp23H)	-48.1	-6.0	4.6
G10S	-45.9	-2.3	4.1
G10Y	-45.2	1.6	3.5
G10N	-44.7	4.3	7.2
T2A	-44.6	-2.7	3.8
T2l	-44.5	0.0	4.9
Q8N	-44.1	-2.7	4.3
T2r_G10Nle	-43.0	-6.4	5.1
T7A	-42.5	3.2	3.6
cp23F	-42.1	-2.9	3.6
T2r_G10M (cp23I)	-41.8	-1.2	6.0
T2r (cp23G)	-41.4	-8.1	4.9
Q8D	-37.8	7.2	3.8
T2y	-37.4	0.8	4.8
H15A	-36.8	-0.3	5.0
G10A	-36.7	7.8	3.9
G10Nle	-36.7	7.8	4.1
G10R	-35.9	-5.1	4.5
T2r_S3L	-35.8	1.4	5.4
H15D	-35.6	6.6	4.8
Q8A	-34.9	4.8	4.4
G10K	-34.4	0.8	5.7
T54N	-34.4	2.2	4.4
G10L	-34.2	0.0	5.5
G10Q	-33.6	4.2	7.7
G10W	-33.6	5.7	3.7
Q8E	-33.5	4.6	4.4
T54Q	-33.3	3.1	5.3
T2Dab	-33.0	0.0	10.8
T54S	-32.8	8.2	4.0
Q8K	-31.3	3.0	4.3
H15E	-28.9	7.5	7.3
T2r_S3E	-28.7	5.7	8.1
G10T	-28.3	1.7	5.8
T2f	-27.7	3.8	9.7
Q8R	-27.5	0.9	7.7
G10I	-26.8	8.7	7.6
T2k	-26.6	5.5	6.0
T2r_G10V	-25.6	4.0	6.5

Table 8. Mean backbone RMSD values and relative free energies (ΔG), using either PB or GB solvation methods, for all the designed analogues. Dap = diaminopropionic acid; Dab = diaminobutyric acid; Nva = norvaline; Nle = norleucine.

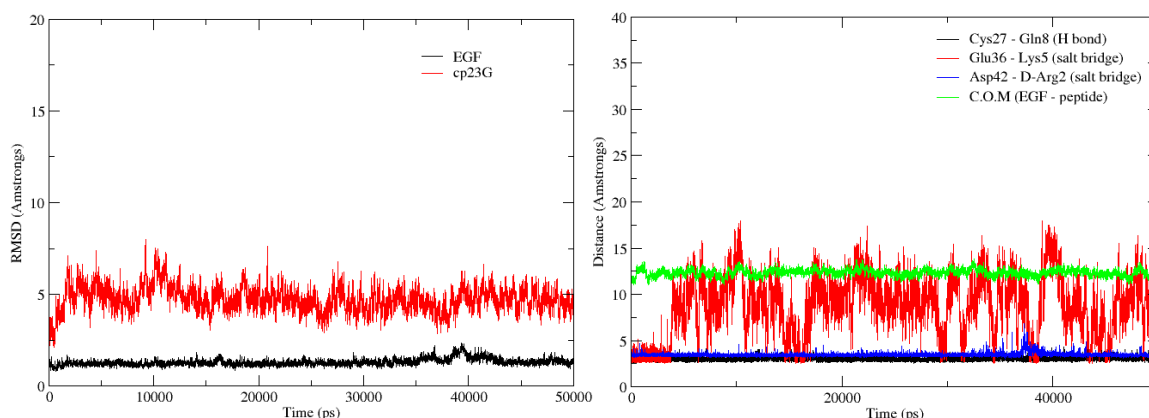


Figure 40. The RMSD values during the MD trajectory of **cp23G** (in red) and **EGF** (in black), calculated for the backbone atoms (left). Distance metrics of three key intermolecular interactions (Cys27-Gln8, Glu36-Lys5 and Asp42-D-Arg2), as well as the center-of-mass distance between the two molecules (right).

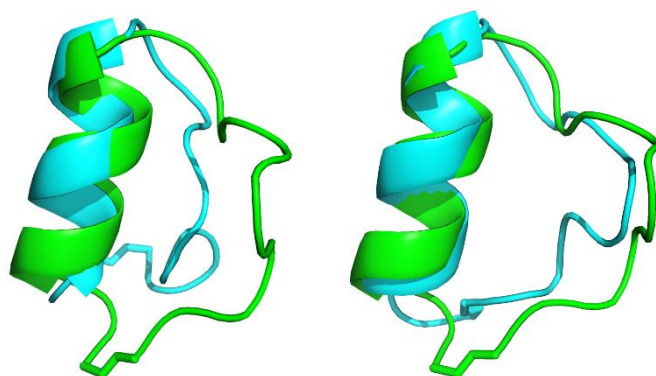


Figure 41. The most populated clusters of conformations for **cp23F** (left, 58% of all conformations, backbone RMSD of 5.0 Å) and **cp23G** (right, 28% of all conformations, backbone RMSD of 2.7 Å) during the REMD experiment (in blue cartoon) are superimposed with their respective initial structure (in green). Note the higher similarity of **cp23G** to the bioactive structure, especially concerning the loop region.

In the SAW binding assay, the calculated affinities of **cp23G** ($K_D = 252 \pm 45 \mu\text{M}$) and **cp23H** ($K_D = 115 \pm 8 \mu\text{M}$) were superior to that of **cp23F**. Likewise, the shifts on the EGF HSQC spectra caused by these peptides were comparable to those observed for **cp23F**, indicating the same binding mode (Figure 43). Surprisingly, peptide **cp23I** displayed very weak binding to EGF, both by means of NMR and SAW ($K_D = 960 \pm 200 \mu\text{M}$).

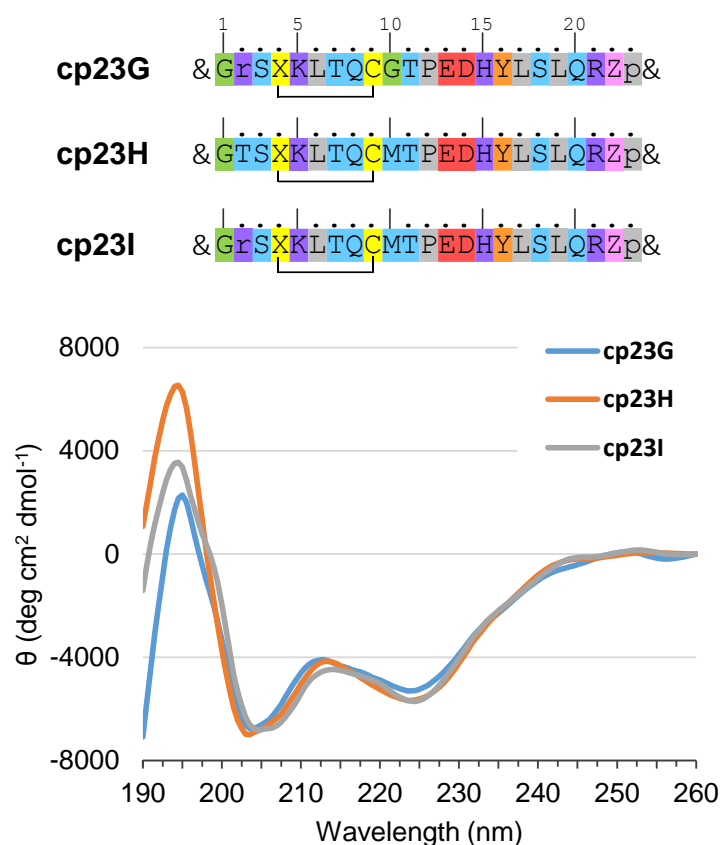


Figure 42. Amino acid sequence (top) and CD spectra of peptides **cp23G**, **cp23H** and **cp23I**, recorded at 20°C in phosphate buffer (pH = 6.8) with 10% TFE. X = L-homocysteine; Z = L-norvaline.

Discussion

The discovery of peptides capable of binding to a protein of interest (EGF) is a challenging endeavor. Among the different strategies to generate hit structures, in this chapter we have explored the mimicry of the partner protein (EGFR) in the interaction. This rational design approach, which is grounded on the wealth of high-quality data available on the complex, aims to generate molecules that mimic the binding epitopes of EGFR and that are thus able to target EGF. However, the EGF-EGFR interface spans a large contact surface (approx. 1600 Å²) that is scattered along two of the extracellular domains of EGFR (domain I and III). In this situation, mimicking large discontinuous protein-binding sites is a challenging task. In order to achieve a successful outcome, the most relevant hotspots in the partner protein have to be unequivocally identified, and a specific scaffold, which presents the respective peptide fragments in their bioactive conformations, has to be designed.

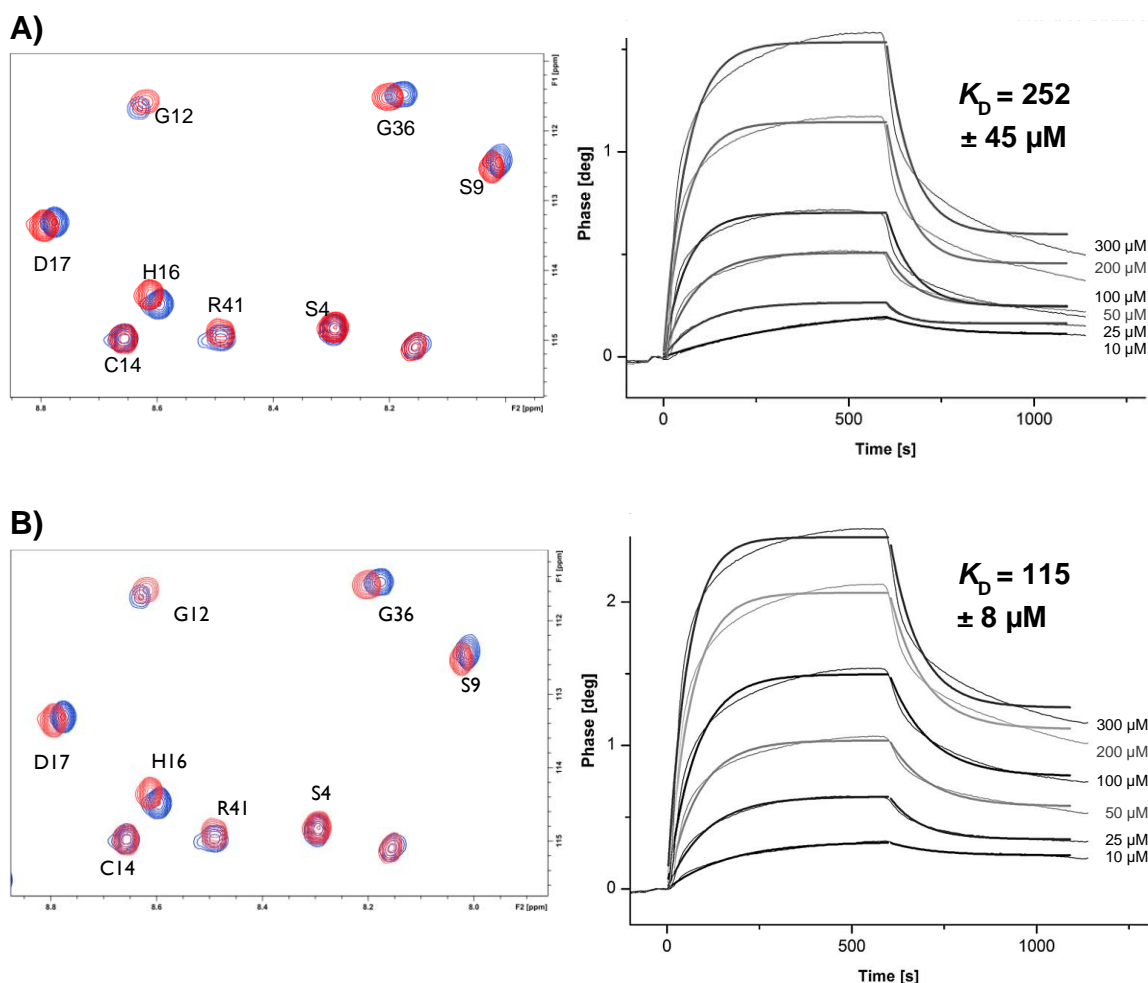


Figure 43. A) Detail of the ^1H - ^{15}N HSQC spectrum of EGF in the absence and presence of **cp23G** (1 mM); residues with significant changes are labelled (left). SAW sensorgram of **cp23G** injected at a range of concentrations (right). B) Detail of the ^1H - ^{15}N HSQC spectrum of EGF in the absence and presence of **cp23H** (1 mM). SAW sensorgram of **cp23H** injected at a range of concentrations (right).

The use of linear peptide segments as protein binders has a major disadvantage, namely that their high degree of flexibility results in negligible or random structure, thereby severely impairing recognition by a well-structured protein target. In addition, linear peptides are prone to protease degradation, low membrane permeability, and metabolic instability. To overcome these limitations, we have exploited the use of cyclic peptides, some bearing non-natural and D-amino acids in their sequences.

Based on the 3D structure of the EGF-EGFR complex, our first attempt to obtain mimetic molecules of EGFR was focused on an interacting loop, located in domain I of the receptor. This loop contains several hotspot amino acids that are directly involved, through hydrogen-bond interactions, in the molecular recognition of EGF. However, despite evaluating a library of

various analogues, our efforts to generate a bioactive peptide were unsuccessful. One possible interpretation is that our designed models failed to effectively mimic the bioactive conformation of this EGFR loop. On the other hand, we cannot exclude that, if isolated from the receptor structure, this small loop of EGFR might fail to bind EGF, thereby indicating that other regions of EGFR are also required for the interaction.

As a result, we switched to the mimicry of larger, albeit more complex, EGFR binding epitopes, such as a specific parallel β -sheet domain that is chiefly responsible for the interaction of EGF with domain III of EGFR. To accomplish the parallel β -sheet arrangement of the peptide chains, we selected a D-Pro-DADME linker for the C-termini of the peptide chains and attached a flexible linker on the N-termini. The resulting macrocyclic peptide (**bs2**), however, presented a random coil structure in solution and failed to bind to EGF.

Fortunately, more positive results were achieved with the mimicry of a cyclic subdomain of EGFR, located in the domain I of the receptor. For this purpose, we designed a disulfide-bridged 28-mer cyclic peptide (**cp28**). The design was validated by computational simulation: its lowest-energy conformations matched the receptor conformation, and its binding to EGF was stable in molecular dynamics experiments. In agreement with the predictions, **cp28** displayed a well-defined structure in solution with a prominent α -helical motif. More importantly, this mini-protein was able to bind EGF at micromolar concentrations, and we succeeded in studying this interaction using complementary biophysical techniques, such as NMR, SAW and ITC.

Based on the CD evidence of cyclic inhibitors, we interrogated whether the helical motif of this peptide was sufficient to mediate binding to EGF. However, peptide **h** failed to interact with the target protein, despite its α -helical conformation. This α -helix probably served as a structural requirement that limited the conformational flexibility of the entire cyclic peptide and stabilized its bioactive conformation. All in all, **cp28** represented the first hit of an EGF peptide ligand and, despite its modest binding potency, provided a very valuable reference for further structure-based optimization.

As well as having a moderate affinity, in fact **cp28** has the drawback of low aqueous solubility at the concentrations at which it performed its activity. For example, after the 4-h ^1H - ^{15}N HSQC experiment, a fine powder appeared at the bottom of the NMR tube, indicating precipitation of the

peptide. Also, given that the **cp28** sequence is formed by all L natural amino acids, a major weakness of this peptide was (as we will see in the next Chapter) its susceptibility to proteolytic degradation. Moreover, the disulfide bridge that yielded the cyclic structure is widely reported to be unstable in the reducing environments commonly found in biological media.

In order to overcome these issues, we followed a structure-based strategy that led to the design of a first generation of **cp28** analogues. In these peptides (**cp23A-C**), we sought to remove some of the hydrophobic residues in **cp28**, with the goal to increase water solubility. However, the improvement in solubility is typically associated with a concomitant decrease in inhibitory activity. To avoid this, we carefully selected and mutated or removed the hydrophobic residues that did not play a major role in the interaction. We further decreased the size of **cp28** (in ≈ 600 Da) by removing the non-interacting region of the peptide. At the same time, we increased its conformational rigidity by introducing a second cyclic constraint that linked residues in the loop (**cp23B**) and in the α -helix (**cp23C**). Finally, in order to improve its metabolic stability, we replaced the disulfide bridge by a covalent head-to-tail amide bond and introduced a D-Pro residue that would impede the action of proteases.

In this first set of constrained analogues, peptide **cp23B** displayed the most promising profile of the series. Indeed, this bicyclic peptide was able to maintain the key structural features of the parent peptide **cp28**, as revealed *in silico* by REMD, and experimentally by CD. The SAW binding experiment revealed a slightly lower K_D , but in the same order of magnitude as **cp28**, thereby proving that the affinity towards EGF could be maintained. The NMR experiments, performed with ^{15}N -labelled EGF, showed that **cp23B**, in its oxidized bicyclic form produced larger perturbations on the EGF spectrum, compared to the reduced monocyclic version. In this experiment, the most affected residues (L8, S9, G12, H16-V19) are located in the A and B loops of EGF and match part of the binding epitope of the parent **cp28**. However, when comparing the maps of the protein residues modified by peptide binding (*Figure 44*), only in the case of **cp28** were other distant regions of EGF also affected. This difference may be due to the larger size and more flexible structure of the parent peptide, which enables it to establish contacts with different parts of EGF.

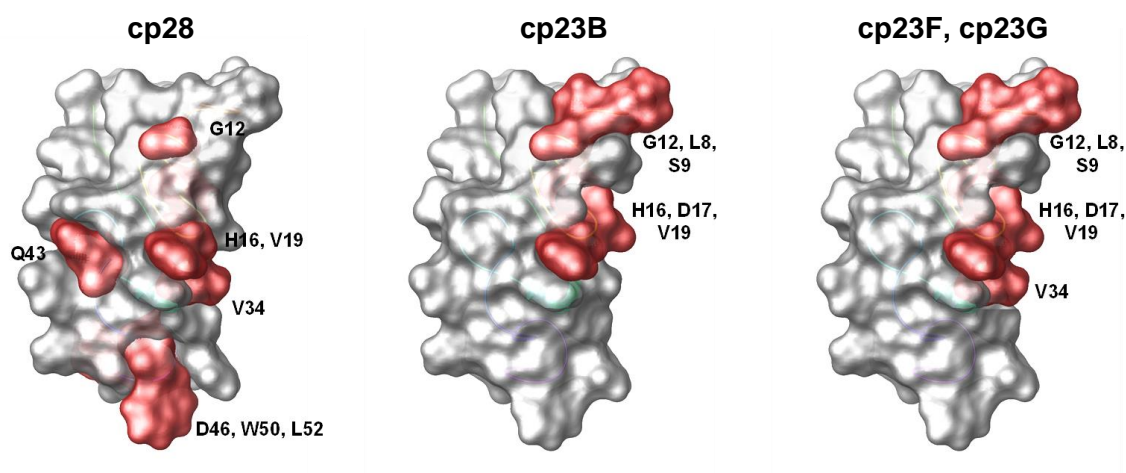


Figure 44. Surface representation of EGF, residues having significant perturbations by NMR upon addition of ligand are shown in red (**cp28**, left; **cp23B**, middle; **cp23F** and **cp23G**, right).

With regards to the synthetic methodology, we overcame the initial head-to-tail cyclization problems and achieved the family of **cp23** peptides. In this regard, the key step was the intramolecular native chemical ligation reaction, and we have been successful in applying this methodology in a straightforward manner. Indeed, the simplicity and efficiency of this reaction has been extremely useful to access this type of bicyclic ligand (the total synthesis and purification can be performed in ≈ 2 weeks' time).

Among the first series of analogues, **cp23B** was selected as hit peptide and was the starting point in a series of computational studies. With this approach, we pursued the design of more diverse and active analogues, via systematic mutation and *in silico* calculation of their theoretical binding energies to EGF. Among the different mutants, the substitution of Tyr12 residue for Pro at the C-terminal end of the α -helical motif restrained the bioactive conformation of the peptide and enhanced the theoretical binding affinity to EGF. In addition, the mutation of the oxidation-prone Met22 for norvaline also proved beneficial for the interaction. As a result, peptide **cp23F** showed slightly better affinity ($K_D = 279 \pm 79 \mu\text{M}$) than **cp23B**, which was nearly identical to that obtained for the parent peptide **cp28**.

However, peptide **cp23F** required a certain degree of structural optimization. Although this peptide had a highly stable helical motif, REMD simulation showed that the preferred conformation of this peptide in solution was rather distant from the original conformation found in EGFR. In detail, the void cavity at the center of the peptide, originated by the lack of bulky side chains, induced the folding of the flexible loop towards the α -helix (see *Figure 40*). This caused a marked deviation in the RMSD of the

flexible loop and consequently of the whole peptide (backbone RMSD = 5.0 Å).

To explore other analogues with a more related conformation, a second round of computational simulation was applied to peptide **cp23F**, this time also taking D-amino acids into account. Although most of the substitutions that we introduced resulted in poorer ΔG values, mutations T2r (**cp23G**) and G10M (**cp23H**) were experimentally validated, achieving slightly improved affinity values to EGF. Curiously, the concomitant mutation of T2r and G10M (**cp23I**) did not yield favorable results. In NMR studies with the ^{15}N -labelled protein, the same binding mode was confirmed for the most active analogues (**cp23F** and **cp23G**, see *Figure 45*). Importantly, all the bicyclic peptides that we generated displayed an almost identical CD signature, thereby indicating the robustness of this constrained scaffold to tolerate various L- and D-amino acid mutations. Indeed, the large side chain of D-Arg in **cp23G**, besides establishing a key ionic interaction with Asp42 of EGF, sits in the central cavity of the peptide and prevents the folding of the flexible loop on the helix, thus resulting in a conformation that nearly matches that of the natural receptor (see *Figure 41*).

From the exhaustive screening and optimization process described in this chapter, peptides **cp23B**, **cp23F** and **cp23G** were selected as representative hits of the first, second and third generation of bicyclic ligands, respectively. Together with the parent peptide (**cp28**), these four ligands are the first peptides reported to interact with EGF. At the next stage and in order to assure their suitability for *in vivo* applications, they were selected for a comprehensive evaluation of biological activity and ADME (absorption, distribution, metabolism and excretion) properties.

Chapter 3:
Towards an anti-EGF
peptide drug:
biological activity
and ADME properties

In the previous chapter, several of our constrained bicyclic peptides have been shown to bind EGF with affinities in the micromolar range. However, in order to block the activation of EGFR and cause a biological effect, these ligands need to disrupt the interaction of EGF with its receptor. Among the different growth factors that are known to activate the EGFR pathway, EGF has the highest affinity of all—with a dissociation constant in the low nanomolar range. As we have discussed in the introduction, the binding of one molecule of EGF to one molecule of receptor triggers the association of EGFR in the form of inactive dimers. In this transient state, the binding of a second molecule of EGF to the partner receptor in the dimer is still more favorable, having a K_d of ~ 50 pM.⁶⁰ Once the active dimer is fully established, the interaction is so strong that no dissociation of the complex practically occurs on the cell membrane. Instead, the active dimer is internalized in vesicles and dissociated under the acidic pH of endosomes for protein recycling.⁶²

Therefore, molecules disrupting the EGF-EGFR interaction must compete against the very strong affinities that naturally govern the EGFR activation process. On the other hand, such molecules will favor from acting on an extracellular target, where higher effective concentrations can be reached. In the work reported in this chapter, we studied the effect of our peptide ligands on the EGF-EGFR interaction, both in fluorescent-based specific assays and in living cells representative of human EGFR(+) tumors. In addition, the ADME properties of our peptides were characterized, and the best candidate (**cp23G**) was subjected to a complete structural characterization by NMR and CD.

PPI inhibition assay

In order to study the behavior of our peptides on a robust system that reproduces the natural interaction, without the heterogeneity and complexity of cells, we set up a fluorescence-based AlphaScreen assay that efficiently reported on the EGF-EGFR interaction. In this experiment, one of the partners of the PPI is attached to polystyrene (donor) beads containing phthalocyanine, a photosensitizing agent that, when irradiated at 680 nm, excites oxygen to a singlet state.¹⁴⁴ The partner protein is immobilized on acceptor beads that contain a fluorescent dye that reacts with the singlet oxygen and emits light at long wavelengths (520-620 nm). Since the lifetime of the singlet oxygen species in aqueous solutions is very short (approx. 4 ms), donor and acceptor beads need to be bound to one another to generate a signal, thus minimizing the background noise.

For this experiment, recombinant EGF was expressed as a His-tagged protein and was immobilized on acceptor beads coated with nickel chelate (*Figure 45A*). Then, the Fc-EGFR fusion protein (purchased) was attached through the Fc domain to complementary donor beads coated with protein A. With this set-up, which can be miniaturized in 96-well plates, the EGF-EGFR interaction was quantified in a precise and straightforward manner, by measuring the fluorescence emission intensity at 615 nm.

To select the optimal assay conditions, a cross-titration experiment with a range of EGF and EGFR concentrations (and fixed concentration of beads) was first performed. As we can see in *Figure 45B*, the fluorescence signal increases proportionally with the concentration of protein in the assay, until a saturation point is reached, after which the signal decreases again (this phenomena is known as the hook effect). At 3 nM concentration of EGF and EGFR, the fluorescence signal is high enough to perform a competitive displacement assay using our peptides. Despite using such low concentrations of protein, the signal amplification that occurs during the irradiation and emission of singlet oxygen allows for an accurate measurement of the interaction. By using the same set of experimental conditions (e.g. same functionalized beads), the strength of different PPI inhibitors can be compared in this assay, which mimics the protein setting found in cells, with the advantage that here the EGF-EGFR complex does not follow the normal biological course (e.g. internalization and recycling).

In order to validate the experiment, we studied the activity of the peptide ligands that we had first developed against EGF: the tripeptide docking hit (**d1**) and the parent mini-protein (**cp28**) and the doubly-mutated analogue **cp28mut**. In these experiments, peptide **d1** was only able to disrupt the EGF-EGFR interaction at milimolar range ($IC_{50} = 1.6$ mM), showing a weak inhibitory effect in agreement with its low affinity for EGF ($K_D = 1.43$ mM). Mini-protein **cp28** proved to be more potent in disrupting the interaction between EGF and its receptor, with an IC_{50} two orders of magnitude lower than **d1** ($IC_{50} = 61$ μ M). As expected from the results obtained with the SAW and NMR, the mutant version **cp28mut** was less active ($IC_{50} = 129$ μ M) than the parent compound.

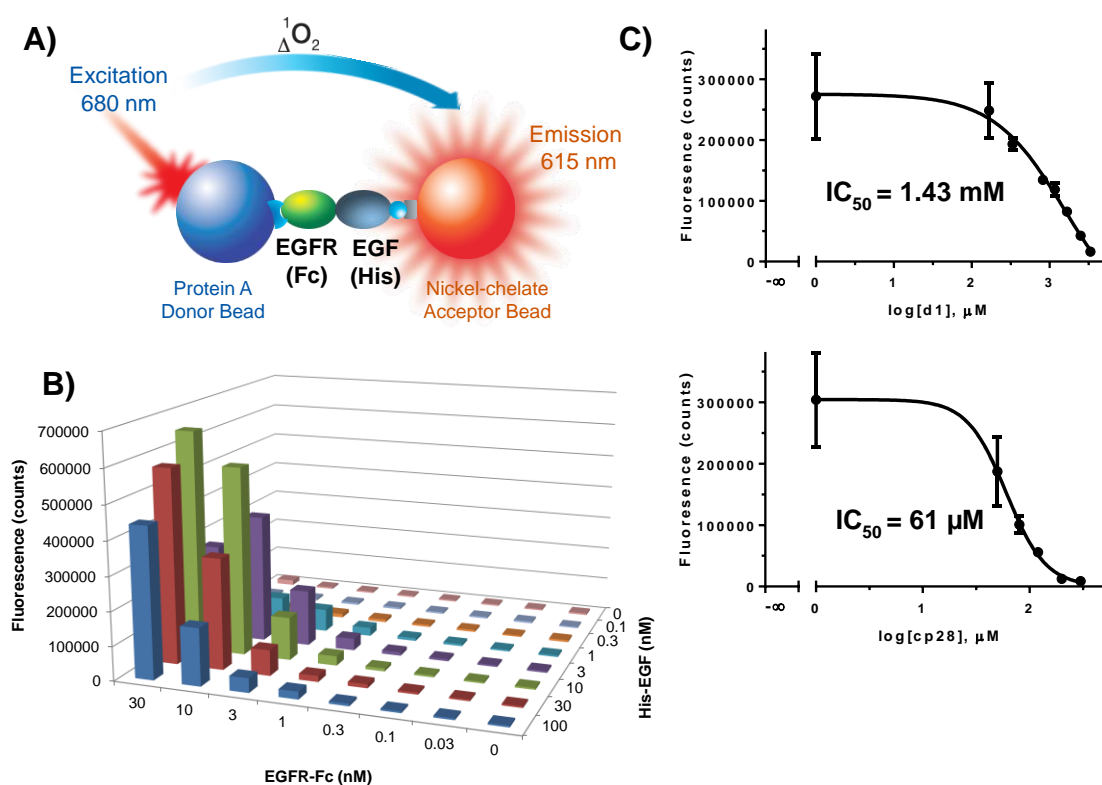


Figure 45. A) The AlphaScreen assay principle. B) Cross-titration experiment with a range of His- EGF and Fc-EGFR. C) Competitive inhibition experiment with a range of **d1** and **cp28** concentrations. Data represent mean values \pm SD ($n = 2$).

When the new library of constrained peptides was developed (**cp23A-I**), we repeated the assay in a different equipment available at that time in another lab, introducing again the best inhibitor of the first round (**cp28**) to make possible the comparison. Among this second set of experiments, the third-generation peptide **cp23G** showed the strongest PPI-disrupting activity ($IC_{50} = 149 \mu M$), while **cp28** and the rest of bicyclic analogues did not produce a strong enough effect on the EGF-EGFR interaction for the IC_{50} values to be accurately determined (*Figure 46*).

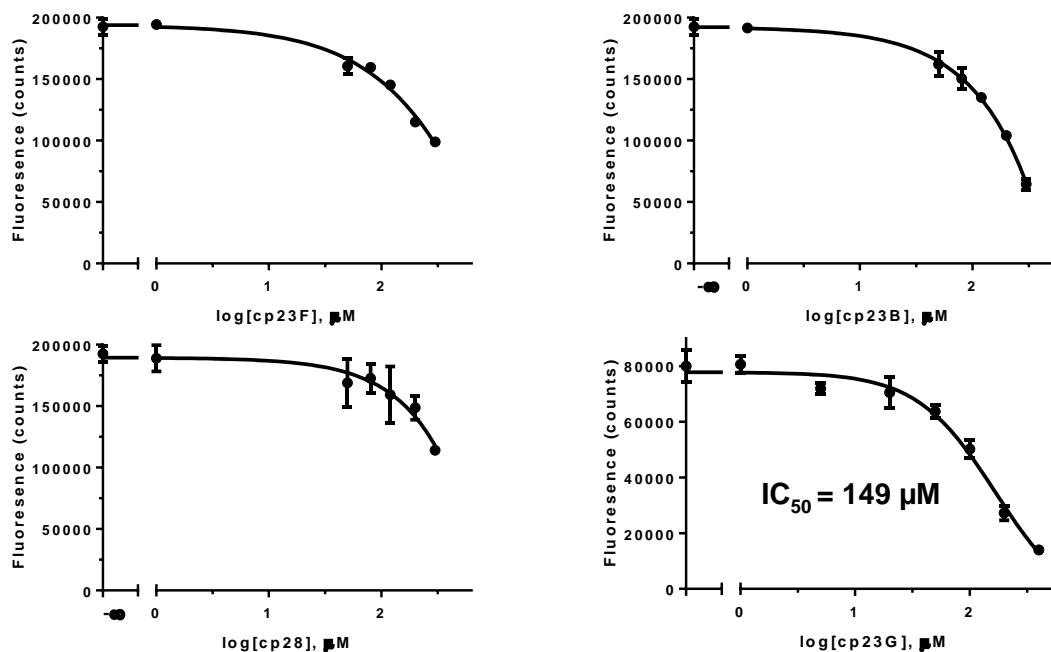


Figure 46. EGF-EGFR inhibition assay with a range of **cp28**, **cp23B**, **cp23F** and **cp23G** concentrations. Only for **cp23G** the fitting of the sigmoid allowed for the accurate determination of the IC_{50} value. Data represent mean values \pm SD ($n = 2$).

Binding of EGF to cells

The AlphaScreen assay is a simplified reproduction of the EGF-EGFR system that is found *in vivo*, being the main advantage its use in high throughput screenings. However, the complexity found in cellular environments, as well as the huge molecular diversity that exists in biological media, pose a greater challenge to the modulation of this specific PPI by our peptide molecules. In order to take this complexity into consideration, we explored the use of human cells with EGFR(+) cancer phenotype.

If appropriately selected, experiments in living cells can provide more relevant *in vivo* biological information than biophysical and biochemical assays, at the same time reducing the number of animal tests and accelerating the overall drug discovery process. Cell-based assays are widely employed to assess the activity of compounds in front of disease-specific cells. Due to the fact that they are performed in a biologically more relevant media, they also provide direct information on cell permeability and enzymatic stability.¹⁴⁵

Among the different types of immortalized cell lines that can be cultured under laboratory conditions, human cancer cell lines represent the original

malignancy and are typically used for anticancer drug screening. These lines are accessible, easy to grow, reliable and reproducible, although they contain mutations that can affect the experimental results if not taken into account.¹⁴⁵ In this regard, the skin carcinoma cell line A-431 expresses abnormally high levels of EGFR, and is often selected as a model for EGFR(+) tumours in biomedical research.¹⁴⁶ Indeed, this cell line perfectly suited our requirements for studying the activity of our EGF inhibitors in a more realistic setting.

First, we wanted to develop a sound methodology in order to quantify the binding of EGF to the A-431 cells, a technique that could then be used to study the inhibitory effect of our peptides. Towards this aim, there are related examples in the literature in which the labelling of EGF is performed with radioactive isotopes, which yields a high sensitivity down to very low concentrations.¹⁴⁷ However, the use of radioactivity is subjected to a restricted regulation and requires the use of hazardous chemicals. As an alternative, we envisaged the labelling of EGF with a fluorescent dye, as it represents fewer manipulation concerns. For this purpose, we selected Alexa Fluor® 488 (AF488), a green-fluorescent dye with a good chemical stability and high quantum yield.

The AF488 was coupled to the Lys side chains of EGF by reaction of the protein with 5 equivalents of the NHS-activated fluorophore in a pH=8 buffered solution. Due to the fact that some of the EGF remained unreacted, the crude product was subjected to HPLC purification in order to isolate the Alexa-labelled EGF. Mass spectrometry analysis confirmed the formation of a fluorophore to protein ratio of 1:1, with minor presence of the 2:1 species (*Figure 47*).

The AF488-EGF sensitivity in aqueous media was first assessed on a standard microplate reader. In this set-up, concentrations as low as 10 nM of AF488-EGF were detected with a good signal-to-noise ratio. Then, the EGFR recognition properties of AF488-EGF were evaluated in A-431 cells. In this experiment, the cells were incubated for 2 h with different concentrations of labelled EGF. Then, the unbound labelled protein was separated by performing several washing-centrifugation cycles, and the cell-bound AF488-EGF was quantified in a fluorescence reader. By plotting the fluorescence intensity of the cells at different concentrations, a K_d value of 12.3 nM could be estimated, in agreement with the reported affinity of the EGF-EGFR system (*Figure 47*).

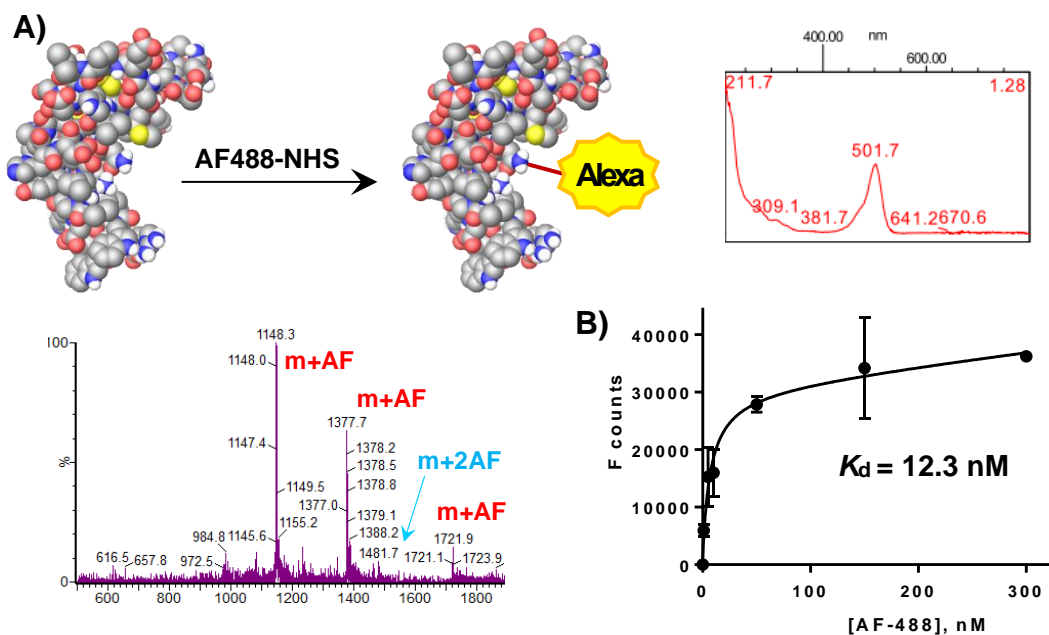


Figure 47. A) Conjugation reaction of Alexa Fluor® 488 to EGF. The incorporation of the fluorophore was confirmed by UV (right) and MS (bottom-left). B) Binding of AF488-EGF to A-431 cells (total fluorescence of the sample quantified in a plate reader, data represent mean values \pm SD, $n = 2$).

In this experiment, we were able to detect a specific and high-affinity binding of Alexa-EGF to living cells. However, due to changing number of cells in each sample, added to the manipulation error introduced during the wash-centrifugation cycles, the total fluorescence measures were subject to a high variability. As a solution to this problem, we resorted to the use of flow cytometry for a more precise quantification of the cell-bound fluorescence.

In particular, fluorescence-activated cell sorting (FACS) allows for the separation of a mixture of cells, one cell at a time, based upon the specific light scattering (a measure of the cell size) and the fluorescent characteristics of the cells.¹⁴⁸ In contrast to the determination of the total fluorescence of the sample, this methodology is able to record the fluorescent signal of each individual cell, thus providing a highly accurate measurement of the cell-bound AF488-EGF. In our experiment, we also stained the cells with propidium iodide to discard the damaged or necrotic cells.

By repeating the incubation experiment of Alexa-EGF with cells, this time performing the quantification by FACS, we obtained a K_d value of 12 nM, matching the result obtained with the plate reader albeit with far less inter-experimental variability. The specificity of the binding was verified by adding different concentrations of non-labelled EGF; at 200 nM the

fluorescent signal was markedly reduced, and it was completely abolished when a large excess of EGF (15 μM) was added to the cells. Also, the binding was disrupted at an acidic pH, confirming the pH-dependence of the EGF-EGFR interaction (*Figure 48C*).

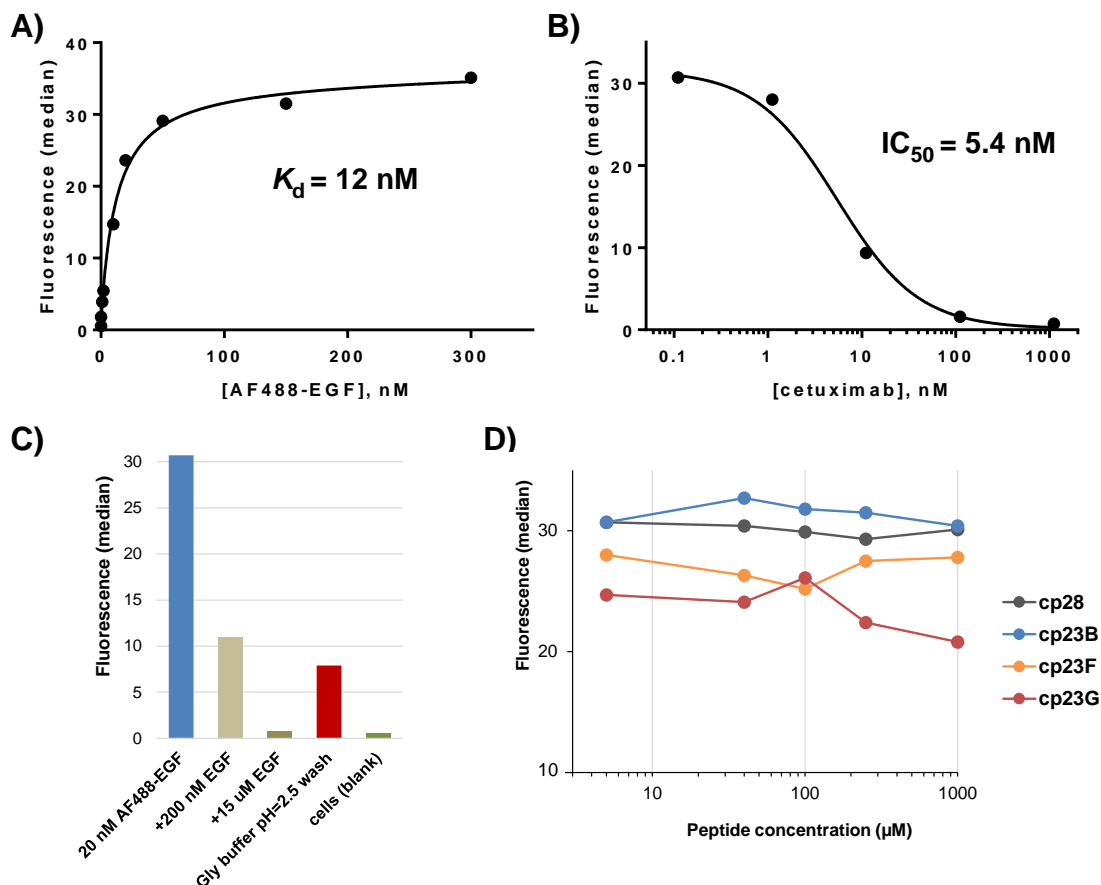


Figure 48. A) Binding of AF488-EGF to A-431 cells (the median fluorescence of 10,000 viable cells was quantified by FACS). B) Inhibitory effect of cetuximab on the binding of EGF to A-431 cells. C) Median fluorescent of cells treated with 20 nM of AF488-EGF, and after various treatments were applied. Note how the addition of non-labelled EGF, as well as the wash with an acidic buffer, disrupts the binding of AF488-EGF to the cells. D) Median fluorescent signal after incubating the cells with different concentrations of peptides.

Finally, the experimental set-up was validated by using cetuximab, an FDA-approved mAb against the EGF binding epitope of EGFR. With this drug, a concentration-dependent decrease in the fluorescent signal was clearly observed, resulting in an IC_{50} value of 5.4 nM (*Figure 48B*).

Once the conditions of the experiment had been optimized, our best peptide candidates were evaluated in this cell-based assay. Unfortunately, none of our designed peptides was able to significantly disrupt the binding of AF488-EGF to the EGFR(+) A-431 cells. Only at the higher concentrations

of **cp23G**, a slight decrease in the fluorescent signal was observed, but the majority of EGF remained bound to the cells (*Figure 48D*).

Cell viability assays

Despite the disappointing results obtained with the cell binding assay, the *in vivo* effect of our peptides was estimated by measuring the inhibitory effects on the proliferation of A-431 cells. As mentioned earlier, these cells are known to overexpress EGFR and their growth and proliferation is highly dependent on the activation of this pathway, thus being a convenient cell model to study EGFR inhibition. To measure cell viability, we used the XTT assay, which is based on the ability of living cells to reduce the tetrazolium XTT yellow salt to its orange formazan product. To prove that the assay conditions were appropriate, the EGFR tyrosine-kinase inhibitor erlotinib was used as a control.

As shown in *Figure 49*, our designed peptides were able to suppress cell proliferation at their highest concentrations. The most potent effect was observed for peptide **cp23G** ($IC_{50} = 24 \mu\text{M}$), which was able to inhibit the proliferation of A-431 cells to a greater extent than the parent peptide **cp28** ($IC_{50} = 30 \mu\text{M}$). Analogues **cp23B** and **cp23F** were less active than the other peptides, in agreement with the relative binding potencies obtained in the biophysical experiments.

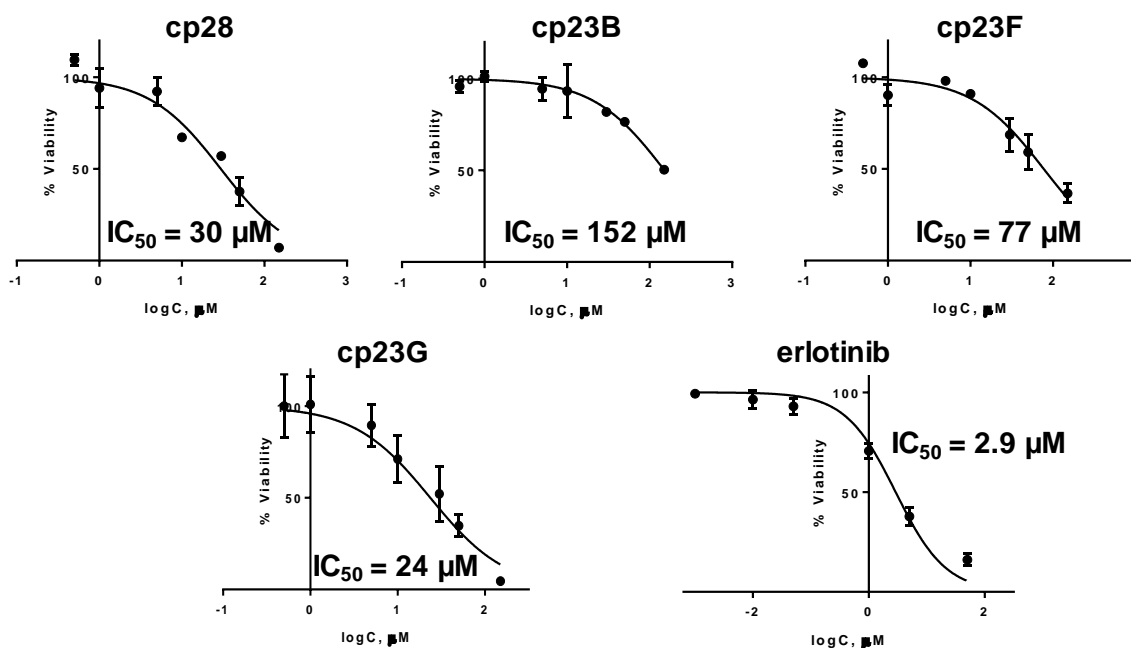


Figure 49. Viability of A-431 cells overexpressing EGFR incubated with a range of peptide concentrations. Data represent mean values \pm SD ($n = 3$).

A parallel assay was carried out in MCF7 cells (a breast cancer cell line that overexpresses estrogen receptor, but normal levels of EGFR) and in mouse fibroblasts, a line representative of non-tumoral healthy cells. In both cells types, lack of toxicity was observed for all the peptides tested, even at the high micromolar concentrations (*Figure 50*). This suggests that the A-431 anti-proliferative activity is produced by the inhibition of the EGFR pathway, and is not due to a general toxic effect of the peptides on the cells.

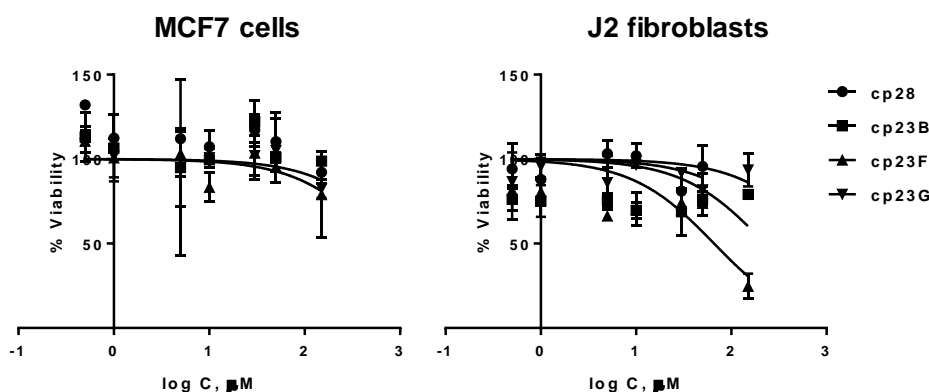


Figure 50. Viability of MCF7 cells (cancer cells with normal EGFR levels) and J2 fibroblasts (healthy cells) incubated with a range of peptide concentrations. Data represent mean values \pm SD ($n = 3$).

Stability in human serum

One of the main obstacles in the development of therapeutic peptides is that they are easily degraded by proteases in the plasma and in target cells. In consequence, they present short half-lives in systemic circulation, resulting in a poor bioavailability. The most common strategies to increase the plasma half-life of peptides is to attach them to macromolecular carriers, such as synthetic polymers or nanoparticles.¹⁴⁹ Although these strategies have proved useful in some cases, their main limitation is that they tend to encumber the peptide bioactivity.

During the design and optimization of the different EGF peptide ligands developed during this thesis, a major objective was to endow our molecules with an enhanced resistance to proteolytic degradation. As far as the stability in biological fluids is concerned, the cyclic nature of the parent peptide **cp28**, which results in a compact structure despite its large size, is expected to hamper the action of proteases. However, the cycle in this molecule is formed by a reversible disulfide bridge connecting the first and last residues. Such S-S bonds are intrinsically unstable in the reducing

environments that are often found *in vivo*. In addition, the fact that **cp28** is an all-L natural peptide further threatens its overall stability.¹⁵⁰

To study the stability of **cp28** in a protease-rich media, we incubated this peptide with 90% human serum at 37 °C, and aliquots were taken at different time-points. Proteins were immediately precipitated with methanol (in which the peptide was fully soluble) and centrifuged. The supernatant was analyzed by UPLC-UV (220 nm) and UPLC-MS. A half-life of 89 ± 6 min was obtained for **cp28**, which indicates a reasonable stability for an all L-amino acid peptide. In comparison, the half-life obtained for ACP, a linear natural peptide used as control, was of only 9 min.

The different proteases found in the serum typically recognize specific cleavage sites on their substrates, and catalyse the selective cleavage of the amide bond. The specificity of proteases depends on the primary amino acid sequence of the substrate, as well as on the three-dimensional conformation that it adopts. Proteases preferentially cleave substrates within extended loop regions, while residues that form part of more compact and structured domains are usually inaccessible to enzyme the active site.

The PROSPER bioinformatics server, which is available on the website, is a useful tool to predict the potential cleavage sites along the sequence of a peptide or a protein.¹⁵¹ For **cp28**, the amide bonds after Thr4, Thr9, Met24 and Phe25 were identified as specific cleavage sites of serine proteases found in the serum, whereas the Arg23-Met24 amide bond can be *a priori* cleaved by metalloproteases such as MMP-9 (*Figure 51*).

In order to gain a better understanding of the metabolically vulnerable points of **cp28**, we performed an UPLC-MS/MS analysis of the major degradation products found at the 4-hour time-point of serum incubation. Contrary to what we expected, the three main metabolites of **cp28** conserved intact the intramolecular disulfide bridge (*Figure 51*). Actually, the cleavage of disulfide bonds preferentially occurs in the reducing environment of the cytosol,¹⁵² and not in extracellular media like the serum. Despite the stability of the disulfide bridge, the three main metabolites that were identified by mass spectrometry suffered from enzymatic hydrolysis of Arg23 (metabolite #1) and Arg23-Met24 (metabolite #3). Moreover, oxidation of the Met side chain occurred in metabolite #2.

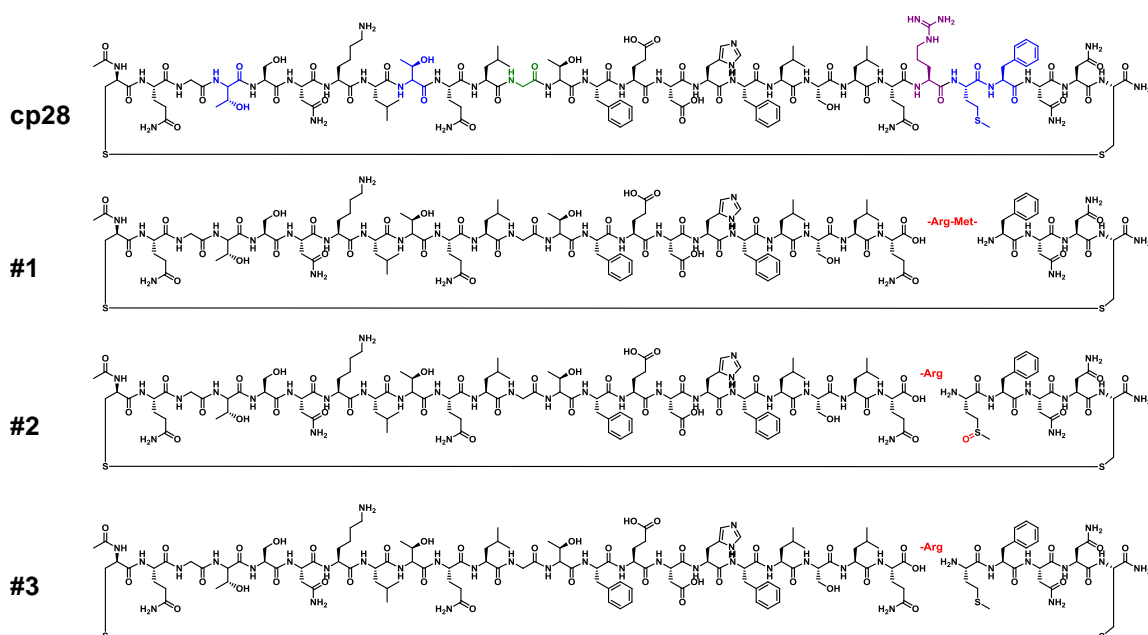


Figure 51. Structure of **cp28** (top); the specific cleavage sites, as predicted by PROSPER, are shown in blue (serine proteases), green (cysteine proteases) and purple (metalloproteases). Structures #1, #2 and #3 correspond to the main **cp28** metabolites detected in serum after 4 h of incubation at 37 °C. The metabolized residues are labelled in red.

It should be noted that the cleavage sites all cluster on the loop region of **cp28**, which is known to have a negligible structure in solution. In the different generations of **cp23** analogues, the disulfide bridge of the parent peptide was replaced for a head-to-tail amide bond, which is completely stable towards reducing conditions. At the same time, the peptide N- and C-termini were removed, with the goal to elude the possibility of exopeptidic cleavage. Finally, the second cyclization further constrained the structure of the peptide, while the introduction of non-natural amino acids would *a priori* hamper the action of proteases.

To determine the real effects of these structural modifications on the stability of our hit peptides, we performed the serum incubation assay for our bicyclic hits **cp23B-G**. Remarkably, no degradation was appreciated after 8 h of incubation at 37°C (Figure 52A), and still more than 60% of the peptide remained intact after 24 h (UPLC-UV quantification, the identity of the peptide was confirmed by MS). The exceptional endurance of our designed bicycles was even more pronounced in peptides **cp23F** and **cp23G**; in both cases, more than 70% of the peptide survived intact at the 24-hour time-point.

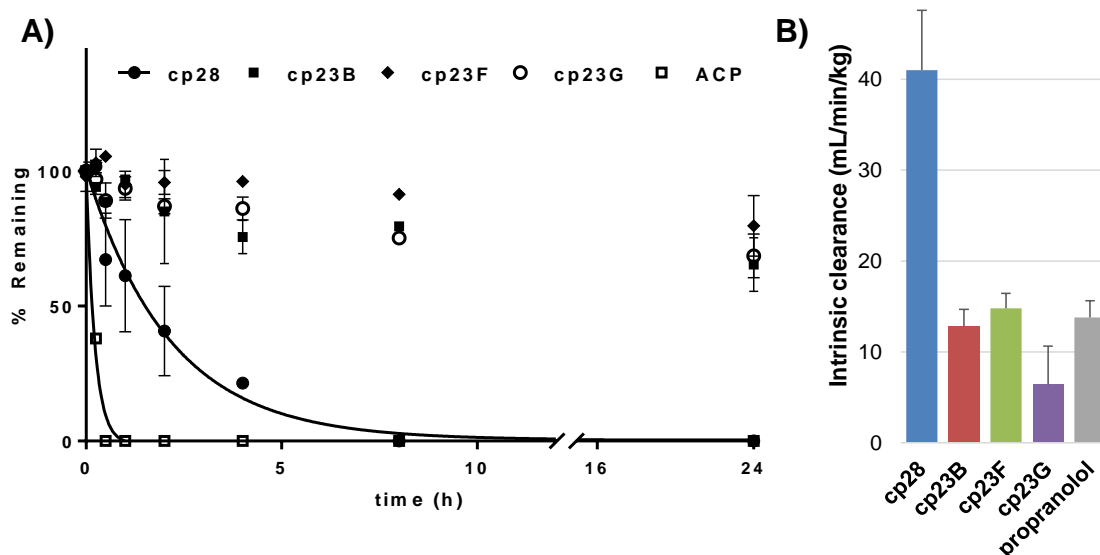


Figure 52. A) Stability of the peptides in human serum at 37 °C. The linear natural peptide ACP was used as control. B) Metabolic stability of the peptides in human hepatic microsomes. Propranolol was used as control. Data represent mean values \pm SD ($n = 3$).

Compared to the parent **cp28**, the enhanced stability of the bicyclic peptides can be explained by their more rigid structure, as well as by the introduction of non-natural amino acids (such as D-Pro, D-Arg and norvaline) in the sequence, a phenomena that is widely accepted to improve the stability in front of proteases.¹⁵³ Indeed, if we pay close attention to the main degradation products of **cp28**, we can observe that they formed as a result of hydrolysis at the Arg23-Met24 loop region of the peptide. In the bicyclic analogues **cp23F** and **cp23G**, this region is shortened and features a D-Pro-Gly turn adjacent to Met24, an artificial motif that hampers the action of proteolytic enzymes. Furthermore, in these peptides the metabolically exposed Met24 is replaced by the artificial amino acid norvaline, contributing to their overall biological stability of the peptides.

Stability in human hepatic microsomes

In addition to the proteolytic degradation catalyzed by enzymes in the plasma, peptides may also be removed from the systemic circulation by renal clearance and hepatic metabolism. Renal clearance takes place in the kidney, where hydrophilic molecules with a mass <5 kDa are filtered out through the glomeruli.¹⁵⁴ For lipophilic peptides, hepatic metabolism may also play an important role. Besides several types of proteases, many other drug—metabolizing enzymes, such as Cytochrome P450 enzymes—which catalyze oxidative reactions,¹⁵⁵ are present in the liver.

Regarding metabolic degradation in the liver, most of the data found in the literature concern small-molecule drugs, reports of hepatic stability for peptides being scarce. To investigate the effects of liver metabolism on the stability of our designed peptides, we incubated our hit compounds with homogenates of human liver microsomes. Prior to incubation, liver microsomes were activated by addition of NADPH and a NADPH-regenerating system. In order to avoid saturation of the hepatic enzymes, 1 μ M of peptide was used for the assay (at such low concentrations, UPLC-MS was used for quantification).

At short time intervals, aliquots were taken and prepared for analysis via precipitation in methanol and centrifugation. The amount of intact peptide in the supernatant was measured by UPLC-MS, and the elimination rate constant was calculated from the slope of the mono-exponential decline phase of the curve. From this *in vitro* kinetic data, the *in vivo* metabolic clearance can be estimated in a straightforward manner (*Scheme 8*).¹⁵⁶ The intrinsic clearance (which can be defined as the volume of plasma from which a drug is completely removed per unit of time and body weight) of propranolol (13.8 ± 1.8 mL/min/kg), a β -blocker drug used as a control, corresponded to that previously reported for this compound (13.2 mL/min/kg),¹⁵⁷ thereby proving that the enzymes were active and the assay conditions were appropriate (*Figure 52B*).

Elimination rate constant (k) = slope exponential phase
$\text{Half-life } (t_{1/2}) = \frac{\ln 2}{k}$
$\text{Intrinsic clearance } (Cl_{\text{int}}) = k \times \frac{\text{volume of incubation (mL)}}{\text{enzyme in the sample (kg)}}$

Scheme 8. Main parameters and equations that define hepatic metabolism.

In this metabolic assay, all three **cp23** analogues present a much lower hepatic clearance than the parent peptide **cp28** (*Figure 52B*). This can be explained by the higher hydrophobicity of **cp28**; this peptide features several apolar residues, such as Phe and Leu, which are highly susceptible to oxidation by the Cytochrome P450 enzymes. In addition, disulfide bonds are known to be reduced and cleaved in the liver to their sulfhydryl components by the enzyme glutathione reductase, also explaining the higher instability of **cp28**. Peptide **cp23G** (half-life of 3.6 h and intrinsic clearance of 6.4 mL/min/kg) exhibited a \sim 7-fold enhanced stability than **cp28**, being the more resistant of all the analogues.

Structural characterization of cp23G

Taking into consideration the results so far presented in this thesis—regarding binding potency, biological activity and metabolic stability, the third-generation peptide **cp23G** presents a very similar affinity, PPI-disrupting capacity and cell activity as the parent peptide **cp28**. However, our designed bicycle has a main competitive advantage in terms of better aqueous solubility, as well as a superior stability in different biological environments. In addition, the interest of **cp23G** lies not only in its biological activity but also in its exotic combination of amino acids which, according to the preliminary inspection by CD and REMD, confer a characteristic structure to this bicyclic peptide.

Thanks to the rigidity imposed by the structural elements of the peptide (the bicyclic scaffold and the two Pro residues), the bioactive conformation that the peptide adopts, upon interaction with EGF, is expected to be highly similar to its conformation in solution. In order to have a detailed picture of this structure (which could serve as a future guide for the rational design of enhanced EGF binders), we decided to conduct an NMR structural analysis of **cp23G**. The experiments—done by Dr. Monica Varese under the supervision of Dr. Jesús García—were performed using the same phosphate buffer (pH 6.8) as the NMR binding experiments with EGF (Chapter 2), 10% of TFE was added in order to promote the formation of secondary structure. In particular, 1D ^1H NMR spectra, in combination with homonuclear (TOCSY, NOESY) and heteronuclear (^1H - ^{13}C HSQC) 2D experiments, were acquired for characterization.

The secondary chemical shifts of both H_α and C_α showed consecutive deviations from random coil values, negative for H_α and positive for C_α , along the Pro12-Gln20 segment (*Figure 53A*), clearly indicating the presence of an α -helix conformation in this region. The presence of multiple d_{NN} ($i, i+1$), $d_{\alpha\text{N}}$ ($i, i+3$) and $d_{\alpha\beta}$ ($i, i+3$) NOEs in this segment further confirmed the α -helical folding (*Figure 53B*). Besides, chemical shift differences between the proline C_β and C_γ atoms for Pro12 ($\Delta\text{C}_{\beta-\gamma} = 3.99$) and D-Pro23 ($\Delta\text{C}_{\beta-\gamma} = 4.37$) indicate that, in both residues, the Xaa-Pro bond predominantly exists in the *trans* conformation.

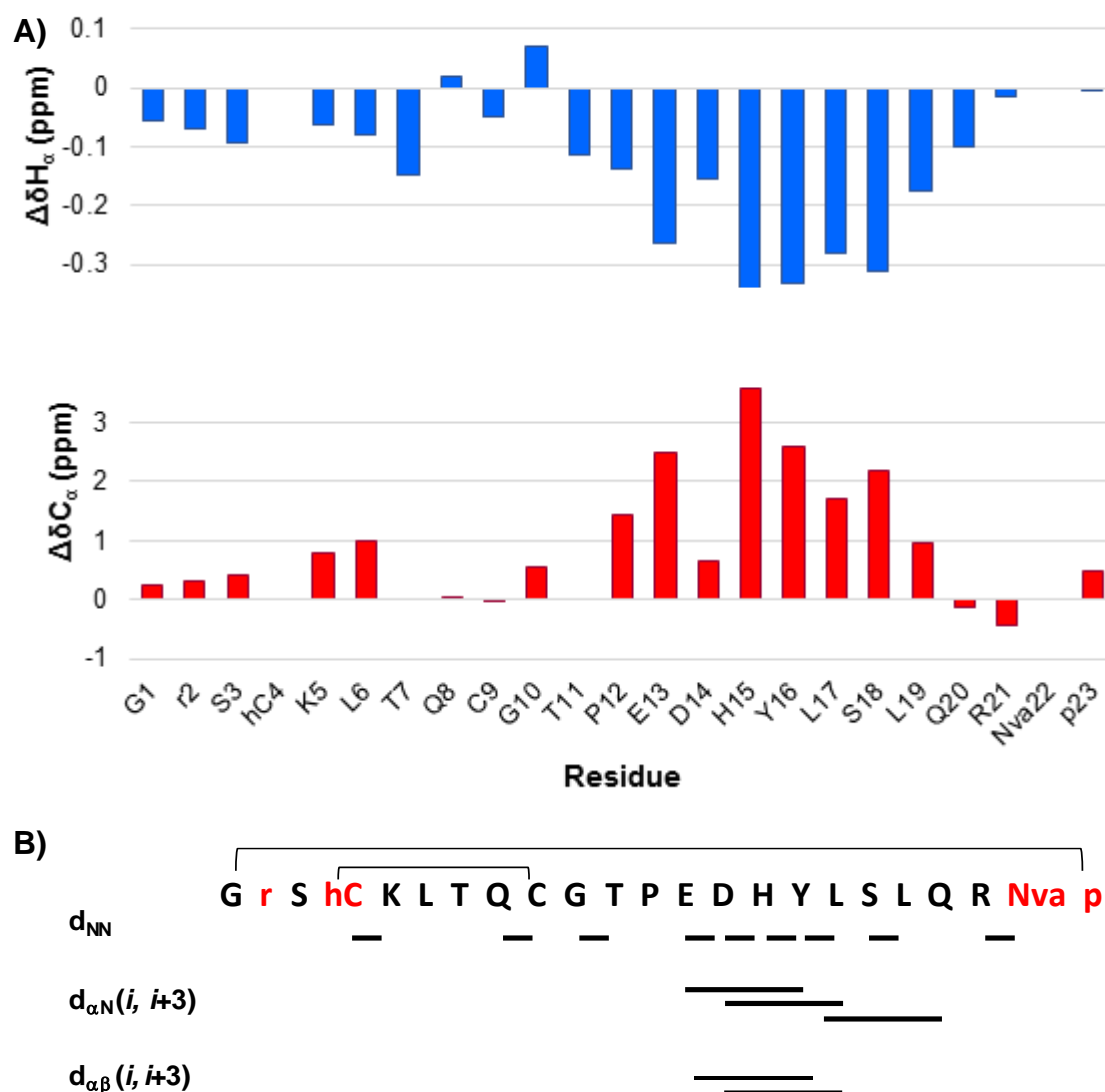


Figure 53. A) Histograms of chemical shift deviations from random coil, for ΔH_{α} (blue) and ΔC_{α} (red) in **cp23G**. Random coil values were extracted from Wishart et al.¹⁵⁸ B) Summary of NOE connectivities observed for **cp23G**. The thickness of the bar indicates the strength of the NOE (weak or strong). Non-natural residues are shown in red.

In order to obtain a 3D-representation of **cp23G**, Dr. Jesús Seco conducted a restrained molecular dynamics simulation applying the NMR-derived distance and dihedral-angle restraints. The lowest energy structure, shown in *Figure 54*, displays the main structural features that were key for the design of the bicyclic analogues (a “bioactive-like” folding of the peptide loop, which is constrained by the intramolecular disulfide bridge, as well as by the distinctive α -helical motif). Importantly, this structure shows a high similarity to the more populated cluster previously found by REMD (with an RMSD value of 1.8 Å for the backbone), which serves as a validation to our computational predictions.

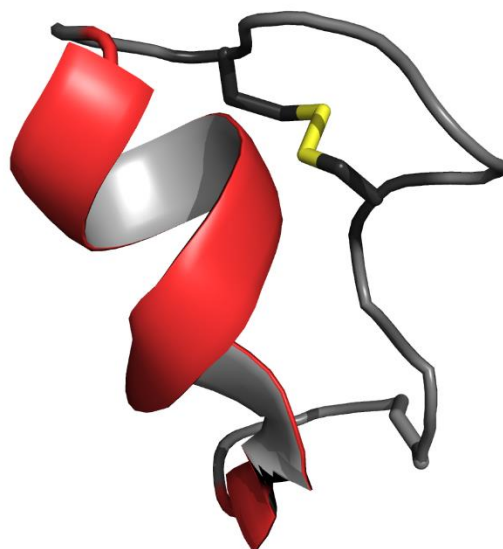


Figure 54. NMR-derived structure of peptide **cp23G**, the disulfide bond is shown in yellow.

Circular dichroism measurements at increasing temperatures were also performed to study the thermal stability of **cp28** and **cp23G**. As shown in *Figure 55*, **cp23G** was found to be exceptionally stable at the range of temperatures studied, with minimal changes observed in the CD spectra. Even at the highest temperature (84°C), the bicyclic analogue maintained a 78% of α -helical structure (measured at 222 nm), whereas this value was lower (69%) for the parent peptide. In both cases, the unfolding process was completely reversible; the peptide recovering its original structure after the heating-cooling process.

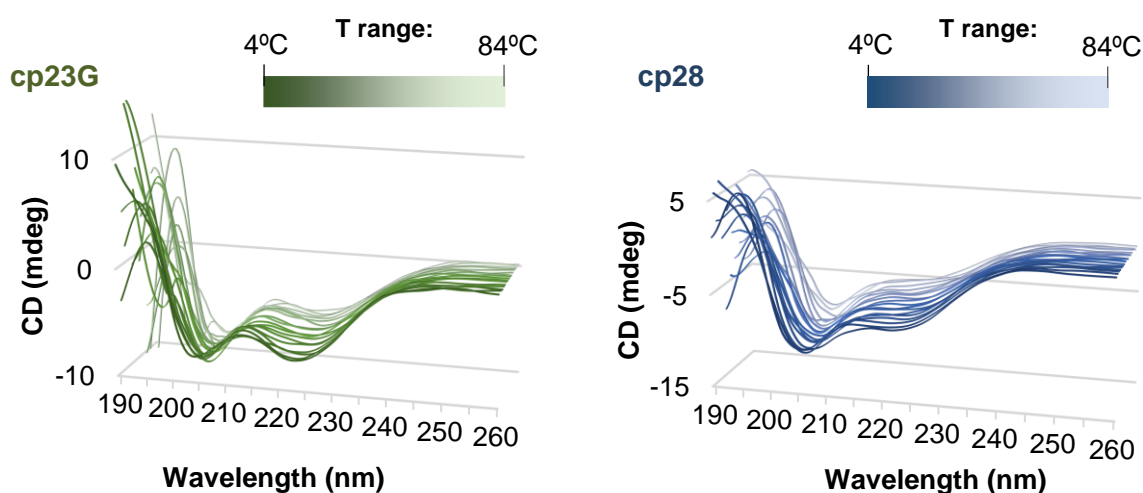


Figure 55. CD spectra of **cp23G** (green) and **cp28** (blue) at temperatures ranging from 4 to 84°C. The measurements were recorded at 90 μ M of peptide in phosphate buffer (pH = 6.8) with 10% TFE.

Discussion

The EGF-EGFR interaction is a particularly challenging PPI due to a number of factors: the large binding interface (1440 Å²) involving several discontinuous epitopes, the high affinity of the complex (~2 nM) and the positive cooperativity of EGFR dimer formation (i.e. the binding of one molecule of EGF favours the binding of a second molecule to the partner EGFR in the dimer). To evaluate if the binding of our peptides to EGF would translate into a biological effect (in other words, whether the moderate affinity of our ligands would be able to disrupt the high-affinity EGF-EGFR complex), we focused on different “cell-like” and “living cell” assays.

First, we devised a specific PPI interaction experiment in the form of a beads-supported fluorescence assay. Although the set-up was straightforward and allowed for the miniaturized screening of our ligands using low amounts of protein, the experiment presented some reproducibility issues (e.g. more consistent results are obtained when the same batch of beads and the same device are used in all the experiments). Nevertheless, all the AlphaScreen assays were able to confirm the PPI-disrupting capacity of our peptides at high micromolar concentrations. Under the same conditions, peptide **cp23G** revealed as the most potent analogue of the series.

In order to study the EGF-EGFR interaction in a biologically more relevant environment, we developed a cell-based experiment using fluorescently labelled EGF to detect the amount of protein bound to the cells. Despite that the protein was successfully derivatized with the Alexa dye (and the activity on EGFR was preserved), the fact that we measured the total fluorescence of the sample made the detection of the cell-bound EGF rather inaccurate. An effective alternative was found in the use of flow cytometry (FACS) to detect the individual fluorescence of each of the cells, thus minimizing the accumulated error during the experiment. A dissociation constant in the low nanomolar range was obtained for the binding of the Alexa-labelled EGF to EGFR(+) A-431 cells. Furthermore, the binding of the protein was found to be reversible and specific for EGFR.

Unfortunately, when we tested our peptide inhibitors, very small changes in fluorescence were detected after the cell incubation. Compared to the inhibition curves obtained in the AlphaScreen, which matched the range of affinities predicted by the biophysical assays, the poor results obtained in cells may be explained by the larger number of EGFR copies that A-431 cells

overexpress. In addition, the intrinsic recycling of the receptor on the cell surface is only partially halted by the low temperature applied to the sample during the experiment. Another possible issue is related to the derivatization of the EGF protein, which although it does not affect the recognition by EGFR, it may alter the binding epitope of our peptides, thus hindering their PPI-disrupting effect.

Nevertheless, we completed the biological characterization by further studying the activity of our peptides in front of different cell lines. As discussed earlier, some human tumours overexpress EGFR as a mechanism to drive cell growth and proliferation. In this regard, the A-431 cell line is one of the most used models for studying human EGFR(+) tumours and testing anti-EGFR therapies, since they express abnormally high levels of wild-type EGFR, and they are highly sensitive to the inhibition of this pathway. As observed in the cell proliferation assay, the IC₅₀ values obtained for the most active peptides (**cp28** and **cp23G**) were in the mid micromolar range, only one order of magnitude higher than erlotinib, an FDA-approved tyrosine-kinase inhibitor. Importantly, no inhibitory effects were observed in EGFR(-) cancer cells and in healthy cells, highlighting the EGF-specific activity of these peptides.

One of the main challenges faced by therapeutic peptides is that they typically present poor absorption, distribution, metabolism, and excretion (ADME) properties with rapid clearance, short half-life in serum and sometimes low solubility. Being an all L-peptide, **cp28** is rather labile in front of serum proteases and its disulfide bridge is known to be cleaved in reducing environments. The identification of its major metabolites, by means of LC-MS/MS, confirmed the predictions that the Arg23-Met24 loop region of the peptide is highly susceptible to proteolytic degradation. In our designed bicycles, this non-interacting loop is shortened and replaced by a turn-inducing D-Pro-Gly motif. This artificial modification, added to the highly constrained nature of the bicyclic scaffold, endowed our **cp23** analogues with an exceptional biological stability (>60% of intact peptide in serum after 24 h), a feat that was further confirmed in a liver metabolism test.

Finally, the in-depth characterization of the best **cp23** analogue (**cp23G**) by NMR has confirmed the key structural features that have been essential in guiding this rational peptide design process. As shown by the different NMR parameters, the bicyclic scaffold has proved a suitable template for stabilizing the bioactive conformation of the peptide. In aqueous solution,

the Pro12-Gln20 segment adopts a well-defined α -helical structure, and the conformational restraint induced by the disulfide bond, as well as by the two *trans* Pro residues, promotes the folding of the loop in a conformation that nicely mimics the bioactive one.

We can conclude that we have been able to identify a series of peptides able not only to interact with the desired target (EGF), but also to disrupt the interaction with its receptor in living cells. The relevance of the inhibition of this PPI has been demonstrated by the efficacy of these peptides against tumour proliferation when treating EGFR-dependent cancer cells versus EGFR-independent ones. The good tolerability shown by healthy fibroblast underlines a certain selectivity, in agreement with the principle of targeted therapy. The addiction of cancer cells to some specific signaling pathway (EGFR, in this case), if compared with healthy cells that can survive by prioritizing alternative mechanisms, explains how a moderate inhibitor (in biophysical terms), such as **cp28**, or the optimized analogue **cp23G**, shows a good *in vitro* efficacy.

Moreover, the optimization of ADME features in the third-generation bicyclic peptides gives a reasonable basis for attempting in the future the test of the best candidates in animal models.

Chapter 4:
A new class of
covalent peptides
as POP inhibitors

Bioactive peptides are typically involved in non-covalent PPI systems, performing their function by competing with an endogenous ligand (either a peptide or a protein) for binding to the target protein. However, peptides often present rapid dissociation rates that produce a transient modulation of the target protein, thus limiting the potential of these compounds to inhibit some targets, such as enzymes. In contrast, small molecules can achieve a higher potency; however, they have proved insufficient for the selective manipulation of enzymes that share a high similarity with other members of the family, such as kinases and proteases.⁸³

These issues have prompted research into the use of “modified” peptides as an alternative class of enzyme inhibitors able to encompass the following two main components: (i) better selectivity, by retaining the main features of peptides—i.e. structural diversity and selective recognition of the target; and (ii) high affinity, by transforming them into covalent inhibitors.

In this chapter, we will explore the introduction of reactive moieties into peptide scaffolds, capable of forming specific covalent bonds with the active site of the target enzyme. As a proof of concept, our efforts have been focused on inhibiting prolyl oligopeptidase (POP)—a key target in neurological disorders, which has been the object of study in our group for several years. In addition, POP inhibitors reported to date are mostly covalent peptidomimetics, thus making this enzyme a suitable target for the approach we propose.

Most POP inhibitors derive from Z-prolylprolinal (ZPP),⁹⁵ and bind to the three cavities of the catalytic site, referred to as S1, S2 and S3 according to the Schechter and Bergen notation.¹⁵⁹ Like ZPP, some compounds incorporate an electrophilic group at the C-terminus that is able to covalently bind to the nucleophilic hydroxyl group of the catalytic serine (Ser554). In this position, hydroxymethyl ketone, α -keto heterocycle, aldehyde or nitrile functionalities have been reported.⁹⁵ However, these groups form an unstable covalent bond with the enzyme, which is hydrolyzed after some time, thus regaining enzymatic activity.

Despite extensive research performed in the field, no irreversible POP inhibitors have been reported to date. We therefore focused on exploring the potential of selective covalent POP inhibitors characterized by an irreversible binding mechanism.

Design based on the structure-activity relationships of POP inhibitors

For the design of the peptidomimetic compounds, we divided the molecules into three parts (P1, P2 and P3) that target the three cavities of the POP active site (S1, S2 and S3, respectively). For the design of the inhibitors, we followed a rational strategy based on the wealth of inhibitors and SAR data that has been generated over recent years,^{95,160} In addition, we have taken advantage of our group's expertise in the discovery of POP inhibitors.^{161,162}

At position P3, POP inhibitors often present an aromatic group (e.g., Z group of ZPP) that binds to the hydrophobic S3 cavity formed by Phe173, Met235, Cys255, Ile591 and Ala594.¹⁶³ When the Z group is replaced with a phenyl ring, an aliphatic chain of three carbon atoms, as in KYP-2047, has been found to be optimal for POP inhibition.¹⁶⁴ However, the S3 pocket is big enough to accommodate even larger and aromatic heterocycles, such as 3-phenoxybenzoyl and Fmoc (*Figure 56*).^{165,166} Along this line, we recently developed a 4-(benzyloxy)-3,5-difluorobenzoyl (BdFB) moiety¹⁶² as a larger fluorinated analogue of gallic acid. When added to the P3 position of POP inhibitors, the BdFB group provided exceptional inhibitory activity, as well as enhanced membrane permeability and metabolic stability.

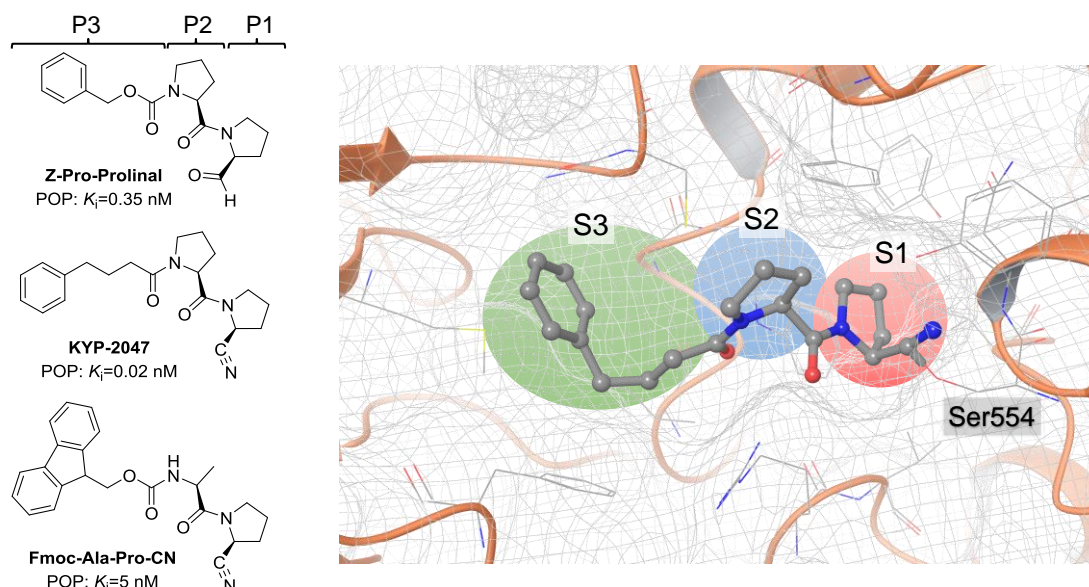


Figure 56. Some representative examples of covalent POP inhibitors (left). Crystal structure of KYP-2047 bound to the POP active site (PDB code: 4AN0). The three cavities of the catalytic site (S1, S2 and S3) accommodate the three respective positions of the inhibitor (P1, P2 and P3).

The S2 pocket of POP is structurally less demanding than its neighbor S1, accommodating proline rings and other amino acids. However, natural S stereochemistry is crucial for ligand recognition. In most reported inhibitors, P2 is in fact kept as a proline. As observed in previous work by our group, small substituents, such as F, CH₃ or CF₃, on position 4 of the pyrrolidine ring are desirable to enhance both the potency and permeability profile of the inhibitor.¹⁶²

Finally, the S1 pocket is surrounded by hydrophobic residues that tightly fit the proline ring and provide for the high substrate specificity of POP. Thus, reported inhibitors most often present a pyrrolidine ring or a very close analogue in this position.⁹⁵ Additionally, in the C-terminus (or position P1'), covalent inhibitors have an electrophilic warhead that reacts with the side chain of Ser554. As mentioned earlier, aldehyde, hydroxyacetyl, boronate and nitrile functionalities have been reported as covalent reversible groups.¹⁶⁰ In our case, we envisaged the use of a different reactive group that would allow an irreversible blockade of the enzyme activity.

For this purpose, we took advantage of our collaboration with the group of Prof. Liskamp, who has pioneered the use of alkyl sulfonyl fluorides as proteasome inhibitors.¹⁶⁷ Due to their mild electrophilic character, sulfonyl fluorides show an adequate balance of biocompatibility, aqueous stability, and protein reactivity, and have found significant utility as probes in chemical biology.¹⁶⁸

As a whole, we sought to condense all the aforementioned structure-activity data in a new peptidomimetic scaffold in which: a) three different aromatic groups (Z, phenylpropyl and BdFB), essential for tight binding to the S3 pocket, are used in P3; b) position 4 of the P2 proline ring is functionalized with small groups (F, F₂, CF₃, CH₃) in order to study their effect on enzyme inhibition and cell-membrane permeability; and c) a unique prolylsulfonyl fluoride warhead is used in P1' for irreversible inhibition of the catalytic Ser. By performing all the combinations possible, this drug discovery approach led to a series of 15 potential inhibitors (*Figure 57*).

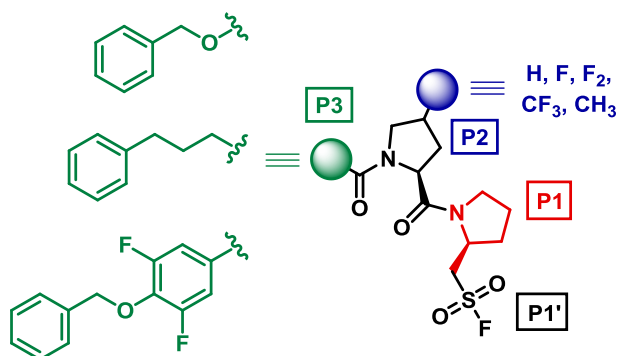
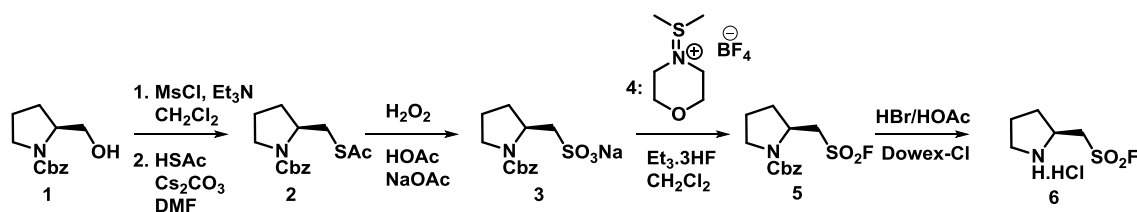


Figure 57. General structure of the family of designed POP inhibitors.

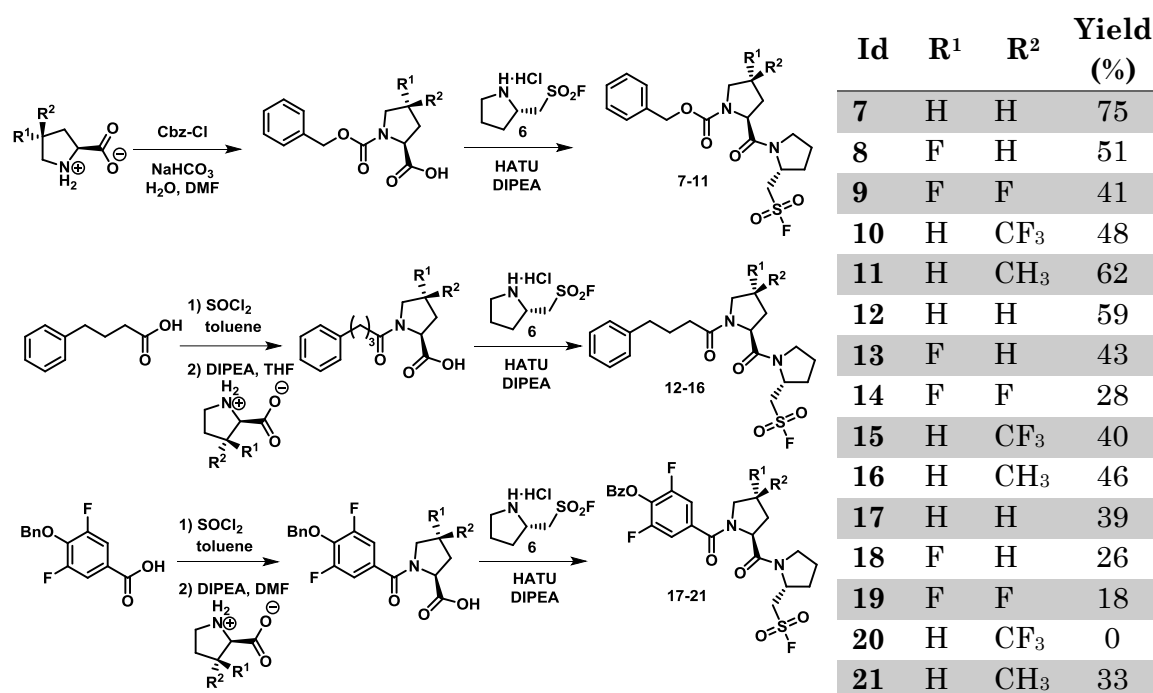
Synthesis of the potential inhibitors

As described above, the three families of designed peptidomimetics feature a prolylsulfonyl fluoride warhead, in combination with different moieties at positions P2 and P3. The reactive warhead was synthesized by the group of Prof. Liskamp according to the following procedure (*Scheme 9*). Starting from the chiral alcohol L-prolinol, the hydroxyl group was first converted into the corresponding mesylate, which was directly substituted with in situ generated cesium thioacetate in DMF. The thioacetate obtained was oxidized using hydrogen peroxide and acetic acid to afford sodium sulfonate **3**. For conversion of other amino acid-derived sodium sulfonates into sulfonyl fluorides, diethylaminosulfur trifluoride (DAST) has been previously used,¹⁶⁹ but in this case was able to produce the sulfonyl fluoride **5** in only a 13% yield. This prompted Liskamp and co-workers to examine more recently developed fluorinating agents, such as the aminodifluorosulfonium salts XtalFluor-E and XtalFluor-M (**4**).¹⁷⁰ For both reagents, substantial gains in the yield of **5** were obtained (69% and 76%, respectively), XtalFluor-M giving a cleaner crude product for easier purification. Finally, cleavage of the Cbz-group using HBr/HOAc, followed by conversion to the hydrochloride salt, afforded deprotected sulfonyl fluoride **6** in high yield.



Scheme 9. Synthesis of the proline-derived sulfonyl fluoride **6**, which was incorporated as electrophile in our designed peptidomimetics.

Then, starting from the respective P3 groups, we applied a divergent strategy, which involved two sequential coupling and deprotection reactions in solution, to achieve the desired scaffolds (*Scheme 10*). The sulfonyl fluoride moiety was coupled at the end, in order to prevent unwanted side reactions of the electrophilic sulfone. Upon completion of the synthesis, a single purification step was sufficient to yield the pure final molecules in a milligram scale. All potential inhibitors were characterized prior to biochemical testing (see Experimental section), with the exception of compound **20**, which could not be obtained due to low reaction yields and complications in the purification process.



Scheme 10. Synthetic strategy to access the three families of POP inhibitors. The structures and obtained yields (after purification) for each compound are shown on the table (right).

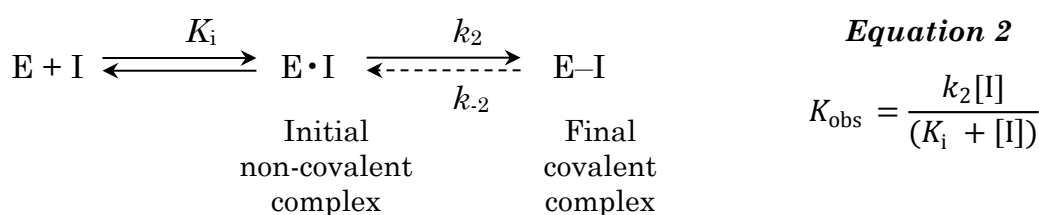
Determination of POP inhibition

For the evaluation and quantification of enzyme inhibition, a fluorometric assay was set up using human recombinant POP that was previously expressed in our laboratory. In this assay, the enzyme was pre-incubated with a range of inhibitor concentrations, followed by the addition of a single saturating concentration of a fluorogenic peptide substrate. To allow for consistent comparison of the whole set of inhibitors, the release of 7-amino-4-methylcoumarin (AMC) was measured during the linear phase of the inhibition process (*Figure 58*).

Although dependent on the pre-incubation time, IC₅₀ values are a useful metric to compare the potency of a given set of inhibitors, since their full kinetic characterization (see below) is laborious and time-consuming. As expected from the rational design, all molecules yielded excellent inhibitory potencies in the nanomolar range, with some compounds, such as **11** and **18**, achieving subnanomolar IC₅₀ values (*Figure 58* and *Table 10*). The introduction of small-sized substituents at position 4 of the P2 proline ring, meant to enhance the selectivity and membrane permeability of the inhibitors, was not detrimental for their activity. Notably, the bulkier BdFB moiety used for inhibitors **17-21** performed as well as the gold-standard P3 groups (Cbz and phenylpropyl) described in the literature for POP inhibitors.

Kinetic studies on the mode of action

From a kinetic perspective, the inhibition process of a covalent drug occurs in at least two steps (*Scheme 11*). The initial non-covalent recognition process between enzyme and inhibitor (defined by K_i) is followed by nucleophilic attack by the protein, giving rise to the inhibited complex—defined by k_2 . In those cases in which the inhibition is effectively irreversible, k_{-2} will essentially be zero, and, if allowed sufficient time, the reaction will proceed to completion rather to equilibrium. As a result of its mechanism of action, the potency and selectivity of a covalent inhibitor is either governed by the initial non-covalent binding to the enzyme (K_i), by the subsequent bonding step (k_2), or by both of them (k_2/K_i)



Scheme 11. Binding process of a target-specific covalent inhibitor (left); k_{-2} is zero in cases in which bond formation is irreversible. The observed rate constant for inhibition (k_{obs}) is determined by fitting the kinetic data to a standard exponential equation, from which K_i and k_2 can be calculated (*Equation 2*).

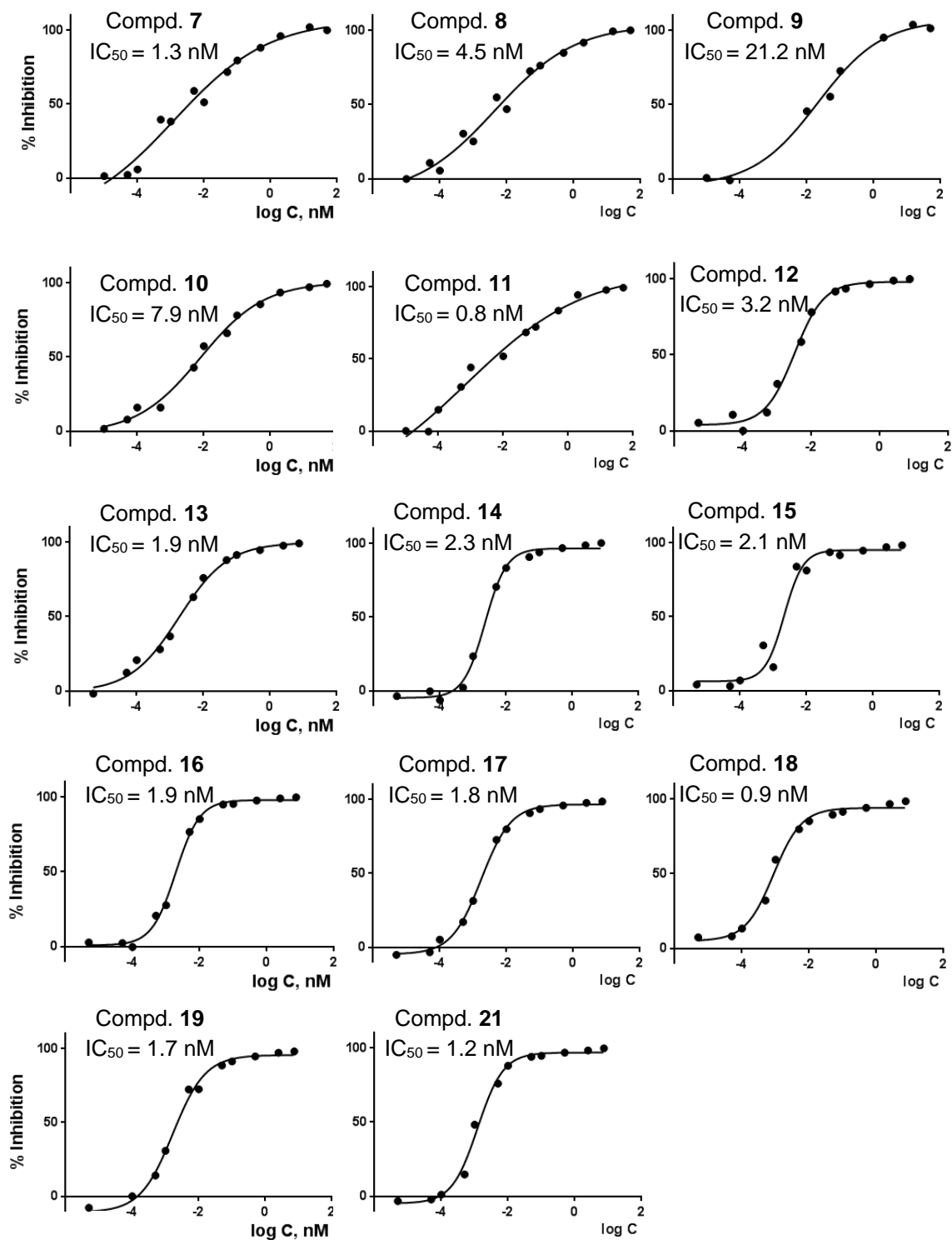


Figure 58. POP activity at different concentrations of inhibitors. IC_{50} values were obtained by fitting the data to a dose-response equation by nonlinear regression. Data represent mean values ($n = 3$).

Non-covalent binding (K_i) typically correlates with the molecular recognition of the target by the drug. By contrast, k_2 depends on the presence of an appropriate nucleophilic residue at a specific position on the target. Thus, the determination of these two orthogonal factors provides direct and quantitative information about the drug-target interaction process.⁸³

In order to confirm that the binding mechanism was irreversible, compound **18** was subjected to a complete kinetic profiling. A range of inhibitor concentrations were added to POP at several pre-incubation times, resulting in an exponential decay in enzyme activity, thereby confirming the irreversible nature of POP inhibition by **18**. By fitting the kinetic data to a standard exponential equation, the observed rate constant for inhibition (k_{obs}) can be obtained.

Then, k_{obs} , are re-plotted against inhibitor concentration, as shown in *Figure 59*, and fitted to a hyperbolic equation (*Equation 2*), from which K_i and k_2 values can be derived. The k_2/K_i ratio, equivalent to the second order rate constant for the reaction, is widely accepted as the kinetic parameter that best represents inhibitory potency.⁸⁴ In our case, a k_2/K_i ratio of $2 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ indicates a potent and fast inactivation of the enzyme—although no such parameter has been previously reported for other POP inhibitors.

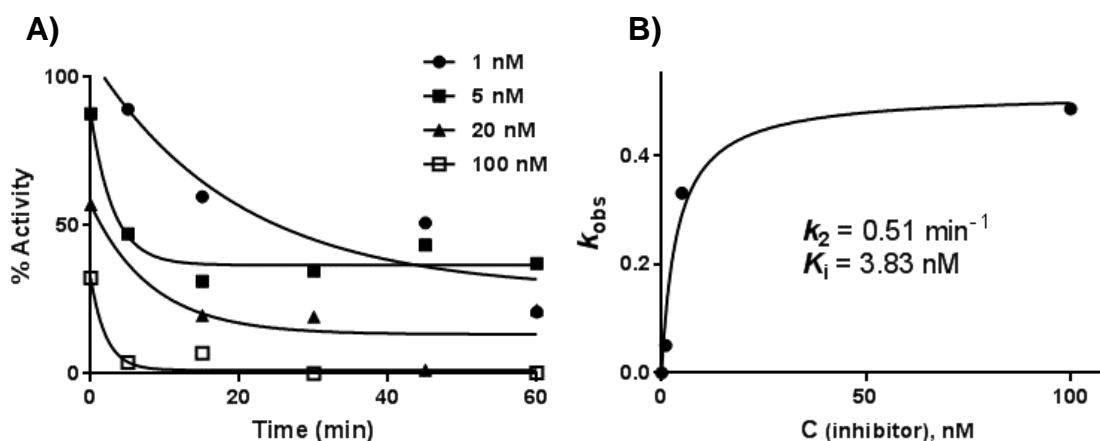


Figure 59. Kinetic characterization of inhibitor **18**. A) The exponential decay of POP activity versus the pre-incubation time, at several inhibitor concentrations. B) Representation of the first-order rate constant (k_{obs}) versus the inhibitor concentration, which yields the kinetic parameters that govern the binding of irreversible inhibitors (K_i , k_2). Data represent mean values ($n = 3$).

Docking insights on the binding mode

In order to assess the binding of our compounds to POP, and to study the key interactions within the catalytic site, we applied the docking software CovDock (which is included in the Schrodinger suite) to predict the pose of covalently bound **18** to the target. As no crystal structure of human POP has been elucidated to date, porcine POP (PDB code: 2XDW), which shares 97% identity with the human form, was used for the *in silico* experiment. After docking, a short minimization of the complex was performed, and the top 20 results were ranked.

As expected, these molecules adopt a very similar pose to that reported for peptidomimetic inhibitors ZPP and KYP-2047 (*Figure 60*). The electrophilic sulfone, despite being larger than the natural amide, is well accommodated between the back-bone of Asn555 and the side chain of His680, and it provides an optimal binding geometry towards Ser554. Two hydrogen bonds, between the backbone of the inhibitors and Arg643 and Trp595, anchor and orientate the false substrate to the active site. The main structural difference of **18**, in relation to ZPP and KYP-2047, is in the S3 pocket, where the larger BdFB moiety fills the entire S3 cavity, resulting in extended hydrophobic and aromatic (π - π) contacts.

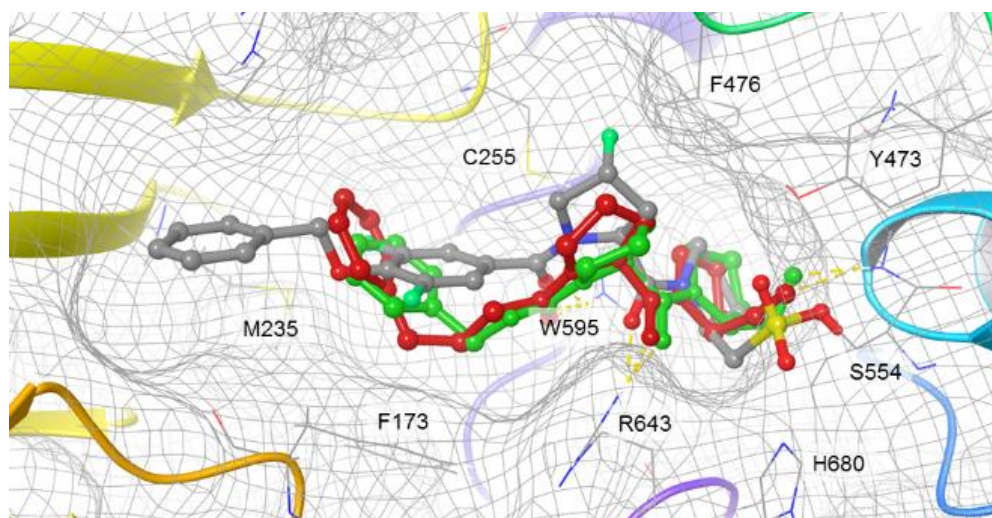


Figure 60. Predicted binding mode of inhibitor **18** to the catalytic pocket of POP, superimposed with co-crystallized ligands ZPP (in green, PDB code: 1QFS) and KYP-2047 (in red, PDB code: 4AN0). Hydrogen bonds are indicated with yellow dashes.

Selectivity versus related peptidases

Among the ~400 known human proteases, a few of them (which form the S9 family of proteases) specifically hydrolyze Pro-containing peptide hormones and neuropeptides. Examples include fibroblast activating protein (FAP) and dipeptidyl peptidase IV (DPP-IV).¹⁷¹ DPP-IV is an exopeptidase that releases X-Pro dipeptides from the N-terminus of peptides. Likewise, FAP has a molecular weight and activity comparable to that of DPP-IV, but also shares equivalent endopeptidase activity with POP. As well as being functionally related, the active sites of these three enzymes are almost identical, with conserved Arg, Phe and Tyr residues surrounding the catalytic Ser. Thus, the design of inhibitors with a good degree of selectivity between these enzymes (especially concerning FAP) has proved problematic, and it is of vital importance to avoid cross-reactivity and potential side effects.¹⁷²

Here, we screened the selectivity of three family-representative compounds (**8**, **13** and **18**) against FAP and DPP-IV, at concentrations up to 50 μM (~50,000-fold the POP IC_{50}). To determine the enzyme activity, fluorometric assays were performed analogous to that describe above for POP. As observed in *Table 9*, a remarkable level of selectivity was found for all three families. Even at the highest concentration (50 μM), the inhibition of DPP-IV was below 25% and that of FAP was barely higher. At lower concentrations—in the POP IC_{50} range—the capacity of these compounds to discriminate between these three prolyl peptidases was exceptional, thus fulfilling one of the main requirements of selective POP inhibitors.

Compd (POP IC_{50})	Enzyme inhibition percentage					
	DPP-IV			FAP		
	1 μM	5 μM	50 μM	1 μM	5 μM	50 μM
8 (4.5 nM)	5 %	6 %	10 %	0 %	7 %	26 %
13 (1.9 nM)	2 %	12 %	15 %	16 %	19 %	21 %
18 (0.9 nM)	7 %	7 %	23 %	12 %	15 %	20 %

Table 9. Inhibition of DPP-IV and FAP by compounds (**13**, **18** and **23**) at concentrations of 1, 5 and 50 μM .

Permeability through brain lipids

As we earlier introduced, even though the exact role of POP in disease is still speculative, its expression is highest in the CNS, and most hypothesis point towards its increased activity during neurodegenerative disorders. Thus, in addition to potency of action and selectivity, a crucial requirement for POP inhibitors is that they must cross the blood-brain barrier (BBB) in order to perform their action inside the CNS.

The BBB is a highly selective semipermeable membrane, formed by brain endothelial cells, which restricts the passage of ~100% of large-molecule neurotherapeutics and more than 98% of all small-molecule drugs.¹⁷³ Most reported POP inhibitors are peptidomimetics compounds with a very low BBB permeability.⁹⁵ In this study, special attention has been devoted to obtain molecules with an acceptable level of BBB permeability by passive diffusion.

To study this property, we used the Parallel Artificial Membrane Permeability Assay (PAMPA), which consists of an artificial membrane in the form of filter-supported phospholipid bilayers.¹⁷⁴ This bilayers come from porcine polar brain lipid extracts, thus they resemble the cell membranes found in the BBB. Based on the previous experience in our group, the PAMPA is a simple and accurate method to evaluate the transport of compounds by passive diffusion.¹⁷⁵

The effective permeability (P_e) of our POP inhibitors, at 200 μM concentration, was determined by quantifying the donor and acceptor compartments after 4 h of incubation (see Materials and methods for the details). In this assay, passage through the *in vitro* brain-lipid membrane was notable for the three families of compounds, achieving permeability values similar or higher than propranolol, an orally bioavailable β -blocker drug, which was used as positive control (*Figure 61*). In particular, permeability was increased upon fluorination of the P2 proline ring, a trend that we had already observed in similar molecules. In contrast, the introduction of lipophilic groups (CH_3 and CF_3) led to membrane retention and a slight decrease in permeability (*Table 10*).

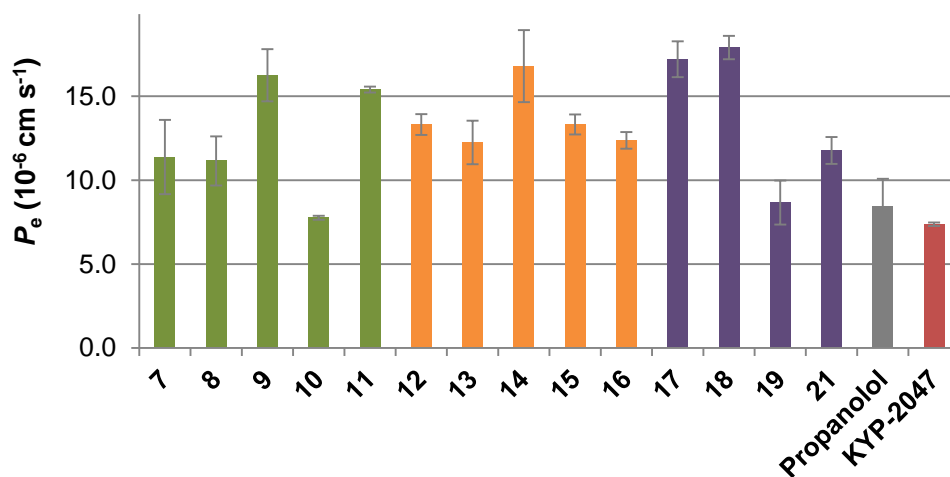


Figure 61. Permeability values obtained in the PAMPA assay for the reported inhibitors and for the covalent inhibitor KYP-2047, as well as for the reference drug propranolol. Inhibitors are color-coded according to their P3 group. Data represent mean values \pm SD ($n = 3$).

Id	R ¹	R ²	POP IC ₅₀ (nM)	Transport (%)	Retention (%)	P_e (10^{-6} cm s ⁻¹)
7	H	H	1.3	19.1	37.1	11.4
8	F	H	4.5	18.8	19.7	11.2
9	F	F	21.2	24.8	2.1	16.3
10	H	CF ₃	7.9	14.0	40.7	7.8
11	H	CH ₃	0.8	23.9	14.2	15.4
12	H	H	3.2	21.5	29.3	13.3
13	F	H	1.9	20.2	23.6	12.2
14	F	F	2.3	25.4	20.4	16.8
15	H	CF ₃	2.1	21.5	51.2	13.3
16	H	CH ₃	1.9	20.3	57.8	12.4
17	H	H	1.8	25.4	26.7	17.2
18	F	H	0.9	26.5	34.0	17.9
19	F	F	1.7	15.3	62.3	8.7
21	H	CH ₃	1.2	19.6	51.4	11.8

Table 10. Summary of the designed POP inhibitors, their structure, inhibitory activity on POP and their permeability profile (transport, retention and P_e) in the PAMPA assay.

Discussion

When targeting the active site of an enzyme, the fast dissociation rates that peptide molecules typically present represent a main obstacle that has to be overcome in order to achieve a sustained inhibition of the target protein. In recent years, the quest for longer acting and more effective drugs has produced a resurgence of covalent-acting drugs, especially those bearing mild electrophiles that enable a high target specificity. Although this strategy has been successfully applied to the field of small molecules, there are few records of peptides and peptidomimetics with this mechanism of action.

Among the different classes of proteases, prolyl oligopeptidase (POP) is a target of paramount importance in the field of CNS pathologies, due to its dual role as proteolytic enzyme and partner in PPIs. Indeed, POP dysregulation has been identified in cognitive disorders, ranging from schizophrenia to Parkinson's disease (see Introduction). Despite the huge efforts dedicated to the development of POP inhibitors, no drugs have reached the market so far—protease selectivity and lack of bioavailability being the main causes of failure.

With the aim of providing new chemical tools as POP inhibitors, we have combined the strengths of peptides (their structural diversity and modularity, which yields an exquisite recognition of protein targets) and of irreversible binders (high potency and prolonged effects) for the design of a new family of covalent-acting peptidomimetics. In our bifunctional molecules, the peptide backbone and side chains form a template that selectively binds to the POP active site, whereas the sulfonyl fluoride electrophile irreversibly reacts with the catalytic Ser554.

For the design of the inhibitors, we carefully selected different P2 and P3 moieties that would be optimal for binding activity, selectivity and BBB permeability. The synthesis of the different analogues was successfully accomplished via consecutive coupling reactions in solution. Fortunately, the designed inhibitors yielded excellent potencies in the low nanomolar range. In addition to their high affinity for POP, these compounds displayed a >1000x selectivity for POP over two closely related proteases (DPP-IV and FAP), thereby confirming the low cross-reactivity of the electrophile.

Finally, these molecules showed their potential to enter the brain by passive diffusion, achieving higher transport values in the PAMPA assay than propranolol, an orally bioavailable CNS drug. All in all, the work

disclosed in this Chapter describes the first irreversible inhibitors of a dual protease and PPI target, POP. Taking into account their unique mechanism of action, these molecules could be further developed as chemical biology tools or as potential drugs.

General discussion

Compared to small molecules, peptides have been largely overlooked as a source of novel therapeutics, owing to their intrinsic properties being diametrically opposed to what conventional medicinal chemistry has defined as “drug-like”. However, significant progress in the development of designed peptides with improved bioactivity and *in vivo* stability has been recently made. Currently, an estimated 60 peptides are approved for human use worldwide, with an additional 140 peptide therapeutics in different stages of clinical development.¹⁷⁶

This shift in paradigm has been enabled by the design and application of more sophisticated peptide scaffolds, featuring a wide range of artificial elements and structural constraints. In most cases, these “non-natural” modifications have conferred peptides with enhanced biopharmaceutical properties.

As we have seen in different parts of this thesis, designed peptides are able to display a variety of foldings and three-dimensional topologies, making them privileged structures for targeting irregular protein surfaces and challenging binding sites. In addition, the chemical diversity of functional groups that natural and/or unnatural side chains can accommodate translates into improved molecular recognition, while at the same time limiting the action of proteolytic enzymes.

Designed peptides in general, and constrained cyclic peptides in particular, have provided new chemical tools for cellular targets that were previously considered *undruggable*.

In the field of oncology, the EGF-EGFR signalling pathway has constituted, since its identification in the 1970s, a main target for chemotherapeutic intervention. Despite the undeniable success achieved in clinical practice by monoclonal antibodies and small molecule inhibitors targeting EGFR, most patients end up developing resistance to the therapy through mutations on the receptor that abrogate drug efficacy.

In the first part of this thesis, we have used peptides to explore an alternative mechanism of EGFR function inhibition, namely the disruption of the EGF-EGFR interaction. This PPI involves a small globular protein (EGF) binding with strong affinity to its receptor (EGFR), through a large and complex interface. Despite the fact that no single drug has been reported to target this PPI, similar approaches have been successfully applied to other similar systems, such as the growth factor VEGF. The recent work of Checco *et al.* is in fact a nice example of how a mini-protein

scaffold can be optimized to develop potent binders of VEGF and disrupt the VEGF-VEGFR interaction.²⁸

As an initial strategy, we relied on a docking-based *de novo* design approach, taking as template the NMR structure of the EGF protein. In spite of the difficulties to study the binding to a small protein—such as EGF—by classic biophysical techniques, we were able to adapt the setting of several methods (fluorescence, NMR, ITC, SAW) to the screening of our compounds, selecting the ones that in our hands proved to be more suitable and effective. Despite the poor correlations between the docking-predicted affinities and the experimental results, the use of ¹⁵N-labeled EGF in NMR spectroscopy, combined with SAW-derived kinetic data, provided a complete picture on the binding process, which would later become invaluable for guiding the design of novel peptides.

In order to overcome the limitations of our docking-based approach (mainly, that it does not consider the dynamic nature of EGF), the mimicry of receptor domains that are critical for the formation of the native EGF-EGFR complex offered an attractive approach for the design of peptide inhibitors. However, the synthesis of a protein sub-domain isolated from a much larger structure proved to be a challenging endeavour, and our initial attempts to reproduce an interacting 7-mer loop of EGFR, as well as a key β -sheet motif on the interface of the receptor, failed to yield positive results.

Eventually, we were more successful with the mimicry of a cyclic subdomain of EGFR, located in the domain I of the receptor. For this purpose, we chemically synthesized and tested a 28-mer cyclic peptide (**cp28**), which showed an interesting capacity to bind EGF at micromolar concentrations. Despite its modest potency, poor aqueous solubility and biochemical stability, **cp28** represented the first EGF-binding peptide, setting a valuable reference for further structure-based optimization.

The first set of analogues (**cp23A-C**) was inherently designed to reproduce the bioactive conformation of **cp28**, while introducing more rigidity and hydrophilicity into the peptide scaffold. In this series, peptide **cp23B** maintained the main structural features of the parent peptide, while the binding potency was only slightly affected.

To further restrain the bioactive conformation and provide a better interaction pattern with EGF, iterative cycles of computational simulation (MD and MM/PBSA calculations of the peptide-EGF complex, as well as REMD simulations of the peptide structure) and biophysical screening were

applied to the optimization of **cp23B**. This theoretical analysis was paralleled by the development of highly efficient peptide chemistry methods, particularly based on native chemical ligation, which was the key synthetic step to form the head-to-tail amide bond, while leaving unreacted the free thiols for the second cyclization to occur.

In a nutshell, this process led to the bicyclic peptide **cp23G**, which features (i) a second proline residue that helps to stabilize the α -helical motif, and (ii) a Thr-to-D-Arg mutation, which improves its folding, binding affinity and stability in serum. A detailed NMR characterization of **cp23G** confirmed the main structural traits depicted earlier by computational simulation.

Importantly, the binding to EGF that was reported by biophysical methods translated into the inhibition of the EGF-EGFR interaction. This effect was not only observed in a purely artificial set-up (the AlphaScreen assay) but also in human EGFR(+) cancer cells. In particular the *in vitro* assay underlined a good selectivity of our compound, which showed a low toxicity in healthy fibroblasts (expressing the EGFR at basal level) and no effect on cancer cells that do not overexpress EGFR. In addition to their biological activity, **cp23G** displayed an exceptional biological stability in aggressive biological media, such as hepatic microsomes.

The weak binding affinities of our ligands, although typically considered a handicap, is not an exception regarding PPI inhibition and may represent an interesting possibility to selectively target biomarkers (like the case of EGFR) that are overexpressed in cancer cells, but are also present on most healthy tissues, thus reducing the toxic effects of the therapy. Indeed, as a result of the harmful effects of chemotherapy, there has been a renewed interest in the use of weak binders, especially in combination with other drugs or as part of a multivalent strategy.¹⁷⁷

In this context, the insensitivity of strong binders to discriminate between two cells (healthy and cancer) presenting different densities of a specific receptor, is one of the main reasons why chemotherapy is directly harmful to the organism. Conversely, carefully designed multivalent drugs (for instance, containing anti-EGF and anti-VEGF peptides) could selectively target only cells with a cognate receptor concentration above a certain threshold value, opening up the possibility for selective chemotherapy with minimal side effects.

On another related issue, in the last chapter of this thesis we have disclosed a family of peptides with a radically different mechanism of action,

as inhibitors of an intracellular brain protease—prolyl oligopeptidase (POP). Despite the role of POP in disease has not yet been fully determined, previous work performed in our group has revealed that the inhibition of this therapeutic target can ameliorate the cognitive symptoms related to schizophrenia.¹⁶² Also recently, POP inhibition has been reported to decrease the aggregation of α -synuclein, improving the outlook of Parkinson's disease.¹⁰⁵

In the context of enzyme inhibition, to overcome the transient nature of peptide-protein interactions, we have introduced a mild electrophilic group onto the peptide scaffold, meant to covalently react with the catalytic serine of POP. Our approach led to the generation of a family of high-affinity low-molecular weight POP inhibitors that exerted sustained inhibition of the target. In addition to their potency, the unique template formed by the peptide backbone and side chains accounted for the >1000x selectivity of these compounds for POP over two closely related proteases (DPP-IV and FAP) in *in vitro* enzymatic assays.

From a pharmacodynamics perspective, these molecules showed their potential to enter the brain by passive diffusion, as reported by the high permeability values achieved in the PAMPA assay. As a whole, these results describe the first irreversible inhibitors of a dual protease and PPI target (POP), and may open up new avenues in the therapeutic scene.

In a nutshell, our findings show how two completely different classes of peptides, bicycle-constrained miniproteins and covalent-acting peptidomimetics, can be efficiently designed to target either large and complex protein surfaces (such as the case of EGF-EGFR) or specific catalytic clefts (such as the case of POP). With protein-protein interactions and challenging binding sites becoming the focus of current drug discovery projects, designed peptides have the necessary qualities to deliver major breakthroughs in terms of novel drugs for the treatment of unmet medical needs.

CONCLUSIONS

In view of our initial objectives, and according to the results presented in the previous chapters, this PhD thesis concludes as following:

1. Using computational tools, we have modeled a series of small-sized peptides that are able to bind *in silico* to EGF, a protein up to now elusive to conventional drugs. Docking simulations have enabled the design of a family of tripeptides that target a hydrophobic cleft present on the NMR structure of EGF. However, the experimental affinities of our peptides did not agree with the *in silico* predictions, a fact that can be explained by the intrinsic limitations of the docking algorithms. In particular, **EGF is not a good system to be treated by current docking methods**, which do not take the flexibility of the protein into account.

Different biophysical methods have been explored to characterize the interactions of our designed ligands with EGF. The small size of the protein, added to the weak affinities displayed by our tripeptide ligands, had consequences on the biophysical techniques that could be applied for the screening. Nevertheless, we managed to overcome these limitations with two approaches:

- (i) **NMR spectroscopy has been used to track ligand-induced chemical-shift changes on the protein ^1H - ^{15}N HSQC spectrum**, an experiment which provides invaluable information on the binding mode of our peptides. For this, we first accomplished the recombinant expression in *E. coli* of uniformly labelled ^{15}N -EGF.
- (ii) In parallel, we have taken advantage of a recently developed methodology to set up a **low-cost, label-free, and highly sensitive SAW experiment able to quantify the interaction of our peptides with EGF**. During the rest of this thesis, the combined use of the NMR and SAW techniques has played a leading role in the discovery of new EGF ligands.

2. To overcome the limitations of docking methods, a diversity of **constrained cyclic peptides** have been developed to **mimic different interacting regions of EGFR**. These are **larger and structurally more complex peptides than those designed by docking, and have proved more efficient in targeting EGF**.

In particular, we have discovered a 28-mer mini-protein (**cp28**) that binds EGF at micromolar concentrations, an interaction that has been thoroughly documented by NMR, SAW and ITC. In addition, CD and REMD experiments have shown that the structure of **cp28** is retained outside its membrane receptor environment. Taking this result as starting point, we have designed of a family of improved **cp28** analogues:

2.1. As a major guide of the optimization process, we have used *in silico* tools (MD and MM/PBSA calculations) to assess the effect of systematic mutations on the theoretical binding affinity of the resulting analogues.

2.2. A carefully designed bicyclic scaffold was shown to reproduce the main structural features of peptide **cp28**, while presenting a more compact size with better solubility and a more restricted conformation.

2.3. The chemical synthesis of these complex bicyclic analogues was facilitated by an intramolecular version of the NCL reaction.

2.4. The iterative screening and optimization process resulted in three generations of bicyclic peptides, from which **cp23B**, **cp23F** and **cp23G** emerged as the hit compounds.

2.5. In peptide **cp23G**, the folding of the backbone was constrained to mimic the bioactive conformation, while several non-natural amino acids were introduced to stabilize the interaction with EGF. As a result, **cp23G** is able to reproduce the binding of **cp28** to EGF, achieving a similar affinity as the parent peptide, albeit with reduced size and hydrophobicity.

3. As well as binding to EGF, **our best peptide candidates are able to disrupt the EGF-EGFR** interaction, an effect that has been tested in several “cell-like” and “living cell” assays. Moreover, **the bicyclic analogues display privileged stability in biological media**. In particular:

3.1. We have implemented a miniaturized experiment, based on fluorescence transfer and emission, which measures the relative PPI-disrupting capacity of our peptides. Under these conditions, **cp23G** displayed the highest activity of the series.

3.2. Using flow cytometry, we have been able to quantify the amount of EGF bound to human EGFR(+) cancer cells. The robustness of the experiment has been tested under a variety of experimental conditions; however, our peptides have produced only minor reductions in the levels of cell-bound EGF.

3.3. The study of cancer cell proliferation, in particular of cells that overexpress EGFR (and are thus highly sensitive to the inhibition of this pathway) is a direct way of measuring the biological efficacy of anti-EGFR therapies. In this setting, our best candidates (**cp28** and **cp23G**) achieved IC_{50} values in the mid micromolar range, a potency only one order of magnitude lower than that of the FDA-approved drug erlotinib.

3.4. The ADME properties of our peptides were characterized in serum and in hepatic microsomes. Being an all L-peptide, **cp28** was found to be rather labile in front of proteases, which cleave the Arg23-Met24 loop region of the peptide. In contrast, the bicyclic analogues exhibit an exceptional level of biological stability (more than half of peptide is still detected in serum at 24 h). This striking stability was further confirmed in a more aggressive hepatic metabolism test.

3.5. Finally, the structure in solution of the best peptide hit (**cp23G**) has been characterized by NMR, providing a better understanding of the bioactive conformation of the peptide. This information can be highly valuable for the design of future analogues.

4. In the last chapter of the thesis, we have designed a series of bioactive peptides with a radically different mechanism of action. In particular, **three families of covalent peptidomimetics have been developed to target the POP active site**. In these bifunctional molecules, the peptide backbone and side chains form a template that selectively binds to the POP active site, whereas a new type of sulfonyl fluoride electrophile is optimally positioned to covalently react with the catalytic Ser of the enzyme.

4.1. **These compounds are highly potent**, being able to inactivate POP at low nM concentrations. A complete kinetic characterization of one of the hit compounds (inhibitor **18**) has confirmed its mechanism of action as an irreversible inhibitor. This is the first example of irreversible peptide inhibitors of human POP.

4.2. In addition to their potency, these peptidomimetics display a **>1,000-fold selectivity for POP** over two closely related proteases (DPP-IV and FAP), reinforcing their low cross-reactivity.

4.3. The capacity of these molecules to **enter the brain by passive diffusion** was assessed in the PAMPA assay, in which they achieved higher transport values than propranolol, an orally bioavailable CNS drug.

EXPERIMENTAL SECTION

Materials and methods

Solvents and reagents

All the reagents employed were purchased to Sigma-Aldrich unless otherwise specified. Dichloromethane, dimethylformamide, methanol, *tert*-butylmethylether were purchased to Schralau; while acetone, acetonitrile and toluene were obtained from SDS. All amino acids used for peptide synthesis, TBTU and Dawson Dbz AM resin were purchased to Iris Biotech. ChemMatrix® resin was purchased to PCAS BioMatrix Inc. Isotopically labeled reagents were purchased to Cortecnet.

Peptide synthesis and characterization

All peptides were synthesized using Fmoc/*t*Bu solid phase peptide synthesis (SPPS). Peptide elongation and other manipulations were done in polypropylene syringes (Schralau) with a polypropylene porous disk. Solvents and soluble reagents were removed by suction. Washings between synthetic steps were done with DMF (5 x 30 s) and DCM (5 x 30 s) using 10 mL of solvent/g resin each time. During couplings the mixture was allowed to react with intermittent manual stirring.

Protected amino acids used were the following unless otherwise specified: Fmoc-Ala-OH, Fmoc-Cys(Trt)-OH, Fmoc-Asp(*t*Bu)-OH, Fmoc-Glu(*t*Bu)-OH, Fmoc-Phe-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Lys(Boc)-OH, Fmoc-Leu-OH, Fmoc-Met-OH, Fmoc-Asn(Trt)-OH, Fmoc-Pro-OH, Fmoc-Gln(Trt)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Ser(*t*Bu)-OH, Fmoc-Thr(*t*Bu)-OH, Fmoc-Val-OH, Fmoc-Tyr(*t*Bu)-OH, Fmoc-Trp(Boc)-OH, Fmoc-Norvaline-OH, Fmoc-Homocysteine(Trt)-OH,

Solid-phase peptide synthesis

Tests to assess coupling completion

Colorimetric tests were used in the manual synthesis after each deprotection step to verify Fmoc removal and after each coupling to know if the new protected amino acid had been successfully coupled. These tests are based on the detection of primary and/or secondary amino groups.

Kaiser or ninhydrin test

Ninhydrin test detects primary amino groups as the ninhydrin chromophore is generated through a Schiff-base intermediate. In this assay, several dried but DCM solvated beads of resin were transferred into a glass

tube. Subsequently, 6 drops of solution A and 2 of solution B were added and heated at 110°C for 3 min. A deep blue or purple solution indicated the presence of free primary amines (positive test), whereas a yellow or pale pink/orange colour indicated the absence of amines (negative test).

Solution A was prepared by adding 40 g of phenol to 10 mL of ethanol and heated until they were dissolved. Another solution of 65 mg of KCN in 100 mL of water was prepared and mixed with 100 mL of pyridine (freshly distilled over ninhydrin). Both solutions were stirred for 45 min with 4 g of Amberlite MB-3 resin and finally filtered and mixed. Solution B was prepared by dissolving 205 mg of ninhydrin in 50 mL of ethanol. This solution was protected from the light.

Chloranil test

The chloranil test is a qualitative test for the detection of free secondary amines on the resin. This test is based on the formation of dialkylaminovinyl derivatives of chloranil. In this assay a few washed resin beads are transferred to a glass tube containing 200 μ L of acetone. Subsequently, 20 μ L of a stock solution of chloranil is added. After shaking briefly, the mixture is left at room temperature for 5 min and the beads are inspected. Blue to green beads means free primary amines (positive test), whereas colourless or yellowish beads indicates the absence of amines (negative test). Chloranil stock solution is prepared by dissolving the chloranil compound in DMF at 2% (w/v) concentration. This stock solution should be kept in the fridge for one month maximum, protected from light.

Initial conditioning of the resin

All peptides with an N-terminal amide were synthesized on ChemMatrix® resin with a substitution of 0.4-0.6 mmol/g. All peptides with an N-terminal carboxylate were synthesized on 2-chlorotriyl chloride resin. In both cases, prior to the first amino acid coupling, the resin was swelled in DCM for 30 minutes and subsequently washed with DMF (5 x 30 s).

All peptides bearing a C-terminal Dbz group were synthesized on Dawson Dbz AM resin with a substitution of 0.4-0.6 mmol/g. In this case, the resin was swelled in DCM for 30 minutes and subsequently an Fmoc removal treatment (see below) was performed.

Loading of the first amino acid

Depending on the type of resin used, the following procedures were followed:

	2-chlorotrityl chloride	ChemMatrix RinkAmide	Dawson Dbz AM
Coupling	1 eq Fmoc-AA-OH, 10 eq DIPEA, DCM (40 min)	3 eq Fmoc-AA-OH, 3 eq TBTU, 6 eq DIPEA, DMF (30 min)	3 eq Fmoc-AA-OH, 3 eq HATU, 3 eq DIPEA, DMF (30 min)
Capping	MeOH (0.8 mL/g resin), 10 eq DIPEA, DCM (15 min)	10 eq Ac ₂ O, 15 eq DIPEA, DCM (15 min)	Not needed

Fmoc group removal

The Fmoc group was removed before each coupling by treating the resin with 20% (v/v) piperidine in DMF (10 mL/g resin, 1 x 1 min and 2 x 10 min). After deprotection, the resin was washed with DMF (5 x 30 s) and DCM (5 x 30 s).

Quantification of resin loading capacity

In order to measure the loading capacity of the resin, after coupling the first amino acid, piperidine washes were collected and measured by UV spectroscopy. Loading was calculated using Equation 3:

$$X = \frac{A \cdot V}{\varepsilon \cdot m \cdot b} \quad \text{Equation 3}$$

where X is the loading of the resin, A is Fmoc absorbance at 290 nm, V is the dilution volume of solvent (in mL), ε is the molar extinction coefficient of Fmoc at 290 nm ($5800 \text{ M}^{-1} \text{ cm}^{-1}$), m is the mass of resin (in mg) and b is the optical path (typically, 1 cm).

Chain elongation in manual synthesis

Method 1 was used as a default for all Fmoc-Aa-OH couplings. When the first coupling was not complete, a second coupling was performed. When the recoupling was unsuccessful, method 2 was used.

Method 1		Method 2	
Treatment	Reagents	Treatment	Reagents
Wash	DMF (5 x 30 s)	Wash	DMF (5 x 30 s)
Coupling	3 eq Fmoc-AA-OH, 3 eq TBTU, 6 eq DIPEA, DMF (35 min)	Coupling	3 eq Fmoc-AA-OH, 3 eq HATU, 6 eq DIPEA, DMF (35 min)
Wash	DMF (5 x 30 s) DCM (5 x 30 s)	Wash	DMF (5 x 30 s) DCM (5 x 30 s)

Chain elongation in automated microwave synthesis

A CEM Liberty Blue microwave peptide synthesizer was used. Drain washings were performed with DMF (2 x 5 mL) with nitrogen gas agitation. Fmoc deprotection was carried out using 10% (w/v) piperazine and 0.1M oxyma in a 9:1 mixture of NMP and EtOH. The same protected amino acids as for the manual synthesis were used. Five equivalents of 0.2 M solutions of Fmoc-AA-OH in DMF were added with oxyma (5 eq, 1 M in DMF), DIC (5 eq, 0.5 M in DMF). The mixtures were stirred for 3 min at 90°C, except for cysteines, histidines and arginines, which were coupled at 50°C for 10 min. After completion of the automatic synthesis following the final Fmoc-deprotection cycle, the peptidyl-resin was washed twice with 10 mL of DMF before being removed from the reactor.

C-terminal activation for native chemical ligation

For peptides bearing a C-terminal Dbz group, a specific treatment was performed prior to cleavage from the resin. The N-terminal residue was introduced using a Boc-amino acid. Then, the resin was washed extensively with DMF and of 1,2-dichloroethane. 4-nitrophenyl chloroformate was dissolved in the least amount of 1,2-dichloroethane, added to the resin and left to gently agitate for 1 hour. The resin was washed with DCM and DMF, and a 0.5 M solution of DIPEA in anhydrous DMF was added. The resin was agitated under nitrogen for 30 min. Then, the peptide was cleaved following the standard procedure (see below).

Cleavage and deprotection of side chains

Peptides were cleaved with concomitant removal of the side-chain protecting groups, using TFA, H₂O and TIS (92.5:5:2.5). When cysteine or homocysteine was present in the sequence the following cocktail was used instead: TFA, DODT, H₂O and TIS (92.5:2.5:2.5:2.5).¹⁷⁸

Work-up

After cleavage of the peptides, the solvent was evaporated applying a current of N₂. The residue was washed 3 times by suspension in TBME and subsequent centrifugation. After that, the cleaved peptides were dissolved in H₂O/ACN (1:1) with 0.1% TFA and freeze-dried.

Native chemical ligation (NCL)

The ligation buffer (6M guanidinium hydrochloride and 200 mM NaP, pH=7) was freshly prepared. Nitrogen was bubbled through the buffer and TCEP 20 mM and 100 mM 4-mercaptophenol were added. The pH was adjusted to 7 by adding a small volume of 12 M NaOH solution. The peptide was dissolved at a 2-3 mM concentration, and the solution was stirred at room temperature during 4 hours. Then, the reaction was acidified, extracted with TBME (2 x 50 mL) and loaded on a PoraPak™ C18 Cartridge for desalting. The guanidinium salts were washed with buffer, while the peptide was eluted at the end in H₂O/ACN (1:1) and freeze-dried.

Thiol cyclization in solution

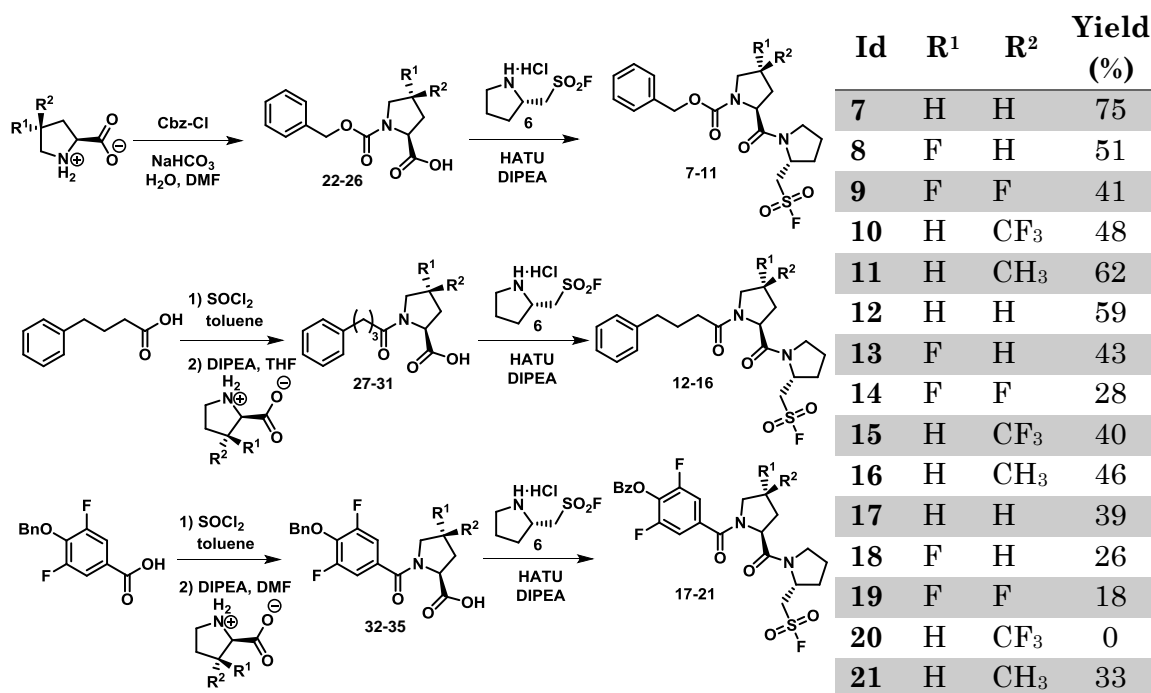
Disulfide bond formation was performed at room temperature with intense stirring under air oxygen during 24 hours. Reaction were followed by HPLC and MALDI. Typically, a 1 L round-bottom flask with a large aperture was used to allow a better oxygenation. High dilution (20-40 μM) was required to minimize the formation of intermolecular dimers and polymers. The pH was adjusted to 8 by addition of ammonium bicarbonate. After 24 h, the reaction was acidified with TFA and lyophilized.

Purification

Crudes were dissolved in H₂O /ACN (keeping ACN to a minimum), filtered through a 0.45 μm filter and purified by semi-preparative RP-HPLC. Fractions corresponding to the peak of interest were analysed by HPLC, pooled and lyophilized. The peptides were purified in a Waters system with a 2545 binary gradient module, a 2767 manager collector and a 2998 photodiode array detector with MassLynx 4.1 software. The column used was a Sunfire C18 (150 x 10 mm x 3.5 μm, 100 Å, Waters), solvents ACN (0.1% TFA) and H₂O (0.1% TFA), flow rate: 6.6 mL/min. Purity was checked by reverse-phase HPLC.

Peptide synthesis in solution

Synthesis of peptidomimetics targeting POP



General procedure for the synthesis of intermediates 22-26. Benzylchloroformate (1.1 eq., 0.13 mmol) was added dropwise to a stirred solution of the corresponding proline analogue (1 eq., 0.12 mmol) and NaHCO₃ (2.5 eq., 0.3 mmol) in water/THF (1:1, 1 mL) at 0°C. The reaction was stirred for 30 min at 0°C and for 5 h at rt. The solution was acidified with 1M HCl aqueous solution and extracted with AcOEt (3 x 10 mL). The combined organic layer was washed with brine (20 mL), dried over MgSO₄, filtered, and evaporated. Crude products were used in the next step without further purification.

General procedure for the synthesis of intermediates 27-35. 4-phenylbutanoic acid or 4-(benzyloxy)-3,5-difluorobenzoic acid (1 eq., 0.12 mmol) were stirred at rt in anhydrous toluene (1 mL) for 10 min. Oxalyl chloride (1.5 eq., 0.18 mmol) was added, and the reaction was stirred at 50°C for 1.5 h. The solvent was removed *in vacuo* and the resulting crude product was redissolved in anhydrous THF. DIPEA (2 eq., 41.8 μL) and the corresponding proline (1 eq., 0.12 mmol) was added neat at 0°C. The reaction was stirred at 0°C for 1 h and at rt for 3 h. The mixture was concentrated *in vacuo*, redissolved in water (3 mL), and acidified with 1M HCl aqueous solution. The aqueous solution was extracted with AcOEt (3 x 10 mL); the combined organic layer

was washed with brine (20 mL), dried over MgSO₄, filtered, and evaporated. Crude products were used in the next step without further purification.

General procedure for L-prolylmethanesulfonyl fluoride coupling. HATU (1.1 eq., 0.13 mmol) and DIPEA (2 eq., 0.24 mmol) were added to a solution of the corresponding intermediate **22-35** (1 eq., 0.12 mmol) in DCM (1 mL). After 5 min, neat (S)-pyrrolidin-2-ylmethanesulfonyl fluoride was added, and the mixture was stirred for 3 h at rt. The reaction mixture was diluted with DCM (10 mL), and 5% KHCO₃ aqueous solution was added (10 mL). The organic layer was then dried over MgSO₄, filtered, and evaporated *in vacuo*. The crude product was redissolved in 2:8 water/acetonitrile and purified by semi-preparative HPLC.

Purity assessment

UPLC

UPLC chromatograms were obtained on a Waters Acquity system (PDA eλ detector, sample manager FNT and Quaternary solvent manager) using a column Acquity BEH C18 (50 x 2 mm x 1.7 μm). The flow rate was 0.6 mL/min using ACN (0.036% TFA) and H₂O (0.045% TFA). 2-min linear gradients were used in all cases.

HPLC analytical scale

HPLC chromatograms were obtained on a Waters Alliance 2695 with an automatic injector and a photodiode array detector 2998 Waters (Waters, Milford, MA) using a Sunfire C18 column (100 x 4.6 mm x 5 μm, 100 Å, Waters) and software EmpowerPro 2. The flow rate was 1 mL/min using ACN (0.036% TFA) and H₂O (0.045% TFA). 8-min linear gradients were used in all cases.

Identification

UPLC-ESI MS

For some peptides, chromatograms and MS spectra were obtained on a Waters Acquity system (PDA eλ detector, sample manager FNT and Quaternary solvent manager) and ESI-MS model SQ Detector2. Column Acquity BEH C18 (50 x 2 mm x 1.7 μm), solvents ACN (0.07% formic acid) and H₂O (0.1% formic acid), flow rate 0.6 mL/min. 2-min linear gradients were used in all cases.

HPLC-ESI MS

For some peptides, chromatograms and MS spectra were obtained on a Waters system Alliance 2695, photodiode detector Waters 2998, ESI-MS model Micromass ZQ and Masslynx 4.1 software (Waters). Column Sunfire C18 (100 x 2.1 mm x 3.5 μm , 100 Å, Waters), solvents ACN (0.07% formic acid) and H₂O (0.1% formic acid), flow rate 1 mL/min. 8-min linear gradients were used in all cases.

MALDI-TOF MS

The determination of the molecular weight of some peptides was performed using a MALDI-TOF/TOF Applied Biosystem 4700. 1 μL of peptide solution (0.5-2 mg/mL) mixed with 1 μL of α -cyano-4-hydroxycinnamic acid (ACH) matrix were seeded on the MALDI plate and air-dried. The matrix was prepared as follows: a 10 mg/mL solution of ACH was prepared in ACN/ H₂O 1:1 (v/v) containing 0.1% TFA.

LTQ-FT MS

Some of the peptides were analysed using a high resolution mass spectrometer to obtain their exact mass. Samples were dissolved in 200 μL of H₂O:ACN and diluted in H₂O:ACN 1% formic acid for MS analysis. The analysis was performed in a LTQ-FT Ultra (ThermoScientific) and the sample was introduced by automated nanoelectrospray. A NanoMate (Advion BioSciences, Ithaca, NY, USA) infused the samples through the ESI Chip, which consists of 400 nozzles in a 20 x 20 array. Spray voltage was 1.4 kV and delivery pressure was 0.3 psi. Data was acquired with Xcalibur software, vs.2.0 (ThermoScientific).

Quantification by amino acid analysis

The content and ratio of amino acids present in a peptide sample were determined by ion exchange chromatographic analysis after acid hydrolysis. The hydrolysis is performed with 6M HCl at 110°C for 16 h. After that time, the sample was evaporated to dryness at reduced pressure. The residue was dissolved in 20 mM aqueous HCl, derivatized using the AccQ-Tag protocol from Waters, which uses 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate as a derivatization reagent, and finally analyzed by ion exchange HPLC.

Structural analysis

Circular dichroism

Circular dichroism spectra were recorded using a Jasco 810 UV-Vis spectropolarimetric, equipped with a CDF 426S/426L peltier. Peptide samples were dissolved in NMR buffer (20 mM sodium phosphate, 50 mM NaCl, 0.1% NaN₃, pH 6.8), and spectra were recorded at concentrations of 20 and 100 μM . A third reading was done after adding 10% TFE to the sample. The following parameters were used: sensitivity (standard, 100 mdeg), start (260 nm), end (190 nm), data pitch, (0.5 nm), scanning mode (continuous), scanning speed (200 nm/min), response (1 s), band width (1.0 nm), and accumulation (3). A blank spectrum of the buffer was subtracted from all recordings, and molar ellipticity values were calculated from experimental ellipticity using the following equation:

$$\theta = \frac{\theta_{exp} \cdot 10^6}{b \times C \times n} \quad \text{Equation 4}$$

where θ is molar ellipticity in $\text{deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$, θ_{exp} is measured ellipticity in mdeg, b is the optical path in mm, C is the peptide concentration in μM and n is the number of residues in the peptide.

After unit conversion, the spectra was smoothed using the Savitzky-Golay method (convolution width = 21) and taken to zero at the far-UV region ($\lambda = 260 \text{ nm}$).

Computational experiments

Protein preparation and docking studies

The NMR structure coordinates of human EGF were extracted from the Protein Data Bank (PDB ID: 2KV4). The structure was then processed with the Schrödinger Suite 2012 Protein Preparation Wizard tool,¹⁷⁹ which automatically adds hydrogen atoms to the protein, creates disulfide bonds, assigns the correct bond orders, and determines histidine protonation states. Finally, a restrained minimization was performed with the OPLS-2005 force field to relieve steric clashes and minimize hydrogen atoms.

Docking studies were performed using Glide v6.5. The protein structure, prepared as described above, was used to build the energy grid. The enclosing box was centred on residues 3, 13 and 29 of EGF, with sizes of 10

Å³ for the inner box (ligand centroid bounding) and 28 Å³ for the outer grid box. The XP docking protocol was used. Ligands were prepared using LigPrep and were flexibly docked. Sampling of ring conformations was included and nonplanar amide conformations were penalized. The best 1000 poses per ligand were saved for restrained energy minimization. All other parameters were set to their default value. Docking results were sorted by GlideScore and top-ranked compounds were visually inspected.

Conformational sampling

The coordinates for peptides **Lp1-16** were extracted from the crystal structure of EGFR in complex with EGF, residues 12 to 18 (PDB code: 1IVO). Then, Asn12 and Leu17 were manually mutated to Lys and Glu (or shorter analogues, such as Orn and Asp), respectively, and the peptide was manually cyclized through the side chains of these residues. 60 minimization steps of the newly introduced linker were performed to relieve the local clashes. The resulting structures were used as input for the Macrocycle Conformational Sampling module of Schrodinger, and conformations were generated using the OPLS_2005 force field and GBSA for electrostatic treatment. Conformations were kept when energies were below 10 kcal/mol, and redundant conformations (RMSD > 0.75 Å) were removed. For each peptide, 5000 simulation cycles were employed with 5000 LLMOD search steps.

Molecular dynamics

The dynamics of EGF, as well as the interaction of all the designed peptides with the protein, were studied by performing molecular dynamics simulation. Coordinates for the protein and the peptide (residues 7 to 34) were extracted from PDB 1IVO. When necessary, the peptide structure was manually modified and the local clashes were relieved by performing a short minimization. The Amber Parm99SB force field was used, while the non-canonical residue parameters of **d1** ligand (2-naphthylalanine) were taken from the FFNCAA force field. The Leap module of the AMBER package was used to immerse the EGF-peptide complex in a pre-equilibrated octahedral box of TIP3P water molecules. Chlorine or sodium ions were added to obtain an electrostatically neutral system. The initial complex structure was first subjected to a minimization protocol consisting of 1000 steps of steepest decent method followed by 500 steps of conjugate gradient method. Thermalization of the system was performed in the NVT ensemble during 200 ps, using a time step of 1fs and increasing the temperature from 100 to

298 K, where a force constant of $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ was applied to protein backbone atoms. Prior to the production run, a short MD simulation (100 ps) in the NPT ensemble was done in order to equilibrate the system density to 1 atm and 298K. The simulations were carried out at constant pressure (1 atm) and temperature (298 K) for 20 ns, except for some peptides that were extended to 50 ns. Low harmonic constraints ($2 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) were used to reduce the protein mobility. Constant temperature was achieved using the Langevin thermostat with a collision frequency of 2 ps^{-1} . The SHAKE algorithm was used to keep bonds involving hydrogen atoms at their equilibrium length. The particle-mesh Ewald summation method was used to deal with long range electrostatic interactions and a cut-off of 10 \AA was applied for non-bonded interactions. The simulations were conducted with the PMEMD module of the AMBER 14 program. The trajectories were analysed using frames collected every 2 ps during the production runs.

MM/PBSA and MM/GBSA calculations

To evaluate the free energy of binding (ΔG) between the peptide ligands and EGF along the molecular dynamics simulation, the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) and Generalized Born (MM/GBSA) methods were used. The MM/PBSA approach¹⁸⁰ estimates the free energy of a system as the sum of molecular mechanics terms as described by the force-field, the solvation free energy using a continuum model and the solute entropy, which can be estimated by quasi harmonic analysis or other means. We have used a numerical solution of the Poisson-Boltzmann equation, as implemented in AMBER, to calculate the polar part of the solvation free energy, using a low ($\epsilon = 1$) and high-dielectric medium ($\epsilon = 80$) for the solute and solvent, respectively. The solute-solvent interface was defined by the solvent-excluded molecular surface, where a solvent probe radius of 1.4 \AA was employed. The non-polar contribution to the solvation was estimated from the solvent-accessible surface area¹⁸¹ using the default scaling factor in Amber ($\gamma = 0.0072 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$). The less computationally demanding MM/GBSA approach was also computed using analogous parameters.¹⁸² The reported values of MM/GBSA and MM/PBSA were computed for a subset of 500 equally spaced snapshots along the converged part of each simulation, when both ligand and protein reached a stable interaction state.

Replica exchange molecular dynamics (REMD)

To assess the structure of the peptides in solution, replica exchange molecular dynamics (REMD) was performed. The initial structure (identical to the one used in the MD simulation described above) was first subjected to minimization protocol consisting of 1000 steps of steep decent method followed by 500 steps of conjugate gradient method. The optimized structure was gradually heated to 300 K in 200 ps. The final structure was chosen as the initial structure for all the 16 replicas in REMD simulations. Temperatures were set in a range from 300 to 600 K with an exchange probability of 30%.¹⁸³ Generalized Born model with an effective salt concentration of 0.2 M was deployed to mimic the solvation effect. Nonpolar solvation term was approximately represented by surface area term.¹⁸⁴ Integral time step was set to 1 fs. Temperature was regulated using Berendsen thermostat with a coupling time constant of 1 ps. SHAKE algorithm was used to constrain all the covalent bonds involving hydrogen atoms. Swaps were attempted every 2 ps and MD simulations were extended to 200 ns for each replica.

Recombinant human EGF expression

The nucleotide sequence encoding hEGF was cloned into a pOPIN vector¹⁸⁵ as a thioredoxin fusion protein, which also includes a N-terminal His-tag and a HRV 3C protease cleavage site. SHuffle T7 3030 *E. coli* cells were used to express the recombinant (His)₆-Trx-EGF fusion protein. Transformed cells were grown at 30°C in LB (Luria-Bertani) media supplemented with ampicillin (100 µg mL⁻¹) until an OD₆₀₀ of 0.8 was reached. Overexpression of the fusion protein was induced overnight at 20°C by the addition of IPTG (isopropyl β-D-thiogalactoside) (1 mM final concentration). The cell pellets were harvested by centrifugation at 3500 g for 30 min at 4°C, re-suspended in 20 mM Tris, 500 mM NaCl, 5 mM imidazole and 0.2% (v/v) Tween20, pH 8.0, and frozen. After thawing, cells were disrupted by intermittent sonication on ice using a regime of 10 s pulse and 10 s break (10 cycles). The lysate was centrifuged at 35000 g for 30 min at 4°C, and the supernatant was then incubated with Ni-NTA (Ni²⁺-nitrilotriacetate)-agarose (Qiagen). The resin was washed extensively with the same buffer, and the fusion protein was eluted using 400 mM imidazole. The fusion protein was further purified on a Superdex 75 column in 20 mM sodium phosphate, 250 mM NaCl, 0.2 mM EDTA and 0.01% (w/v) NaN₃, pH 7.4, and cleaved overnight at 4°C by HRV3C protease (1 mg per 50 mg fusion protein). After cleavage, the supernatant was treated with Ni-NTA

resin to remove the (His)₆-Trx tag. EGF was recovered in the flow-through fraction and was further purified by size- exclusion chromatography on a Superdex 75 column in NMR buffer: 20 mM sodium phosphate, 50 mM NaCl and 0.01% (w/v) NaN₃, pH 6.8. Compared with protein molecular-mass standards, EGF showed an apparent molecular mass consistent with a monomer. Electrospray analysis of the purified protein yielded an average molecular weight of 6216.0 Da, consistent with the formation of the three expected disulphide bridges. Fractions containing purified EGF were combined and quantified by UV. Aliquots were prepared, immediately frozen in liquid nitrogen and stored at -80 °C.

Uniform ¹⁵N-labeling was achieved by growing the bacteria in M9 minimal medium containing ¹⁵NH₄Cl as the sole nitrogen source.

Fluorescence emission titrations

For the fluorescence titration of peptides **d1-DMAP** and **d2-DMAP** with EGF, a concentration of 5 µM of peptide in PBS was titrated against a range of EGF concentrations (0-200 µM). Samples were incubated for 1 h. Spectra were recorded in 1-cm path length quartz cells (450 µL nominal volume) at 25°C. The spectrum of PBS was recorded as the blank. The DMAP fluorophore was excited at 421 nm, and the spectra were recorded between 432 nm and 730 nm. The following parameters were used: excitation slit width (2 nm), emission slit width (5 nm), 0.5 nm increments and 0.1 s integration time). Intensities at the wavelength of maximal emission were represented in front of the concentrations of EGF. Binding curves were obtained with the GraphPad Prism 6.03 software, using a non-linear fit and variable slope.

SAW experiments

Affinity analyses were performed with a Sam5 Blue biosensor (SAW-Instruments, Bonn, Germany), consisting of a biosensor unit, an autosampler, and a microchip module with a gold layer sensing surface on a quartz chip. The chip surface was prepared and cleaned before binding and affinity determinations by 45 min washing with a 1:1 mixture (v/v) of concentrated sulphuric acid and hydrogen peroxide (30%). Immobilization of EGF was performed by covalent binding of a flat 16-mercaptohexadecanoic acid self-assembled monolayer (COOH-SAM), as previously described.¹⁸⁶

EGF was immobilized on the SAM by carboxyl-group activation with a 1:1 mixture (v/v) of 200 mM (1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) and 50 mM *N*-hydroxysuccinimide (NHS).¹⁸⁷ A 10–20 μM solution of EGF in 150 μL PBS was used for coupling, followed by capping of unreacted carboxyl groups with 1 M ethanolamine, pH 8.5. Peptide affinities were determined at 25°C by injecting a range of peptide concentrations (micro- to millimolar) selected to cover the equilibrium concentration. For the regeneration of the chip between injections, affinity bound ligands were eluted with 0.1 M HCl aqueous solution (pH 1).

Binding curves were analysed using the Origin Pro 7.5 software, and K_D values were calculated determined using the FitMaster software. The signal at the association equilibrium (A_{eq}) was represented for each ligand concentration (c), and non-linear fitting of the data to *Equation 5* yielded the K_D values for the interaction.

$$A_{eq} = \frac{c}{c + K_D} \quad \text{Equation 5}$$

ITC experiments

ITC experiments were performed at 25°C using a low volume nano ITC calorimeter (TA instruments). Peptide and protein samples were dissolved in the same buffer (20 mM NaP, 50 mM NaCl, 0.1% NaN₃, pH 6.8), centrifuged, and degassed prior to the ITC experiments.

For each titration, a concentrated peptide solution was injected into a cell containing 190 μL of protein solution at a concentration of 15-20 μM , measured in a NanoDrop™ 2000 spectrophotometer. We performed 16 injections of 3 μL per titration with a 4-min delay after each injection. Binding isotherms were analysed using the software provided by TA instruments, assuming a single binding site for the independent domains. Baseline controls were acquired with buffer and pure peptide solutions.

NMR spectroscopy

¹⁵N-¹H HSQC spectra

¹⁵N-labelled EGF was prepared in NMR buffer (20 mM sodium phosphate, 50 mM NaCl, 0.1% NaN₃, pH 6.8) supplemented with 10% D₂O for a final sample volume of 180 μL , and fixed concentrations of protein (50 μM) and peptide (typically 1 mM).

^{15}N - ^1H HSQC spectra were recorded at 25°C using a Bruker 600 MHz NMR spectrometer equipped with a cryogenic probe. Spectra were acquired with 128x2048 complex points with a total of 8 transients per increment. The x-carrier frequency was referenced to internal DSS-d6. Data processing was performed using TopSpin3.0. Briefly the F2 domain of the datasets was increased to a factor of two by linear prediction and then the F1 and F2 dimensions were zero-filled by a factor of two. A q-sine function was used for line broadening.

The quality of the protein sample was first evaluated by performing a ^{15}N - ^1H HSQC spectrum of EGF, and comparing the results with the reported data.⁶⁹ For the determination of CSP perturbations, combined NH chemical shift perturbations ($\text{CSP} = [(\delta\text{N}/5)^2 + (\delta\text{H})^2]^{1/2}$) were plotted for each protein residue; only CSPs > (mean + SD) were considered significant.

Structural analysis of cp23G

The peptide **cp23G** was dissolved in NMR buffer (20 mM sodium phosphate, 50 mM NaCl, 0.1% NaN_3 , pH 6.8) supplemented with 10% trifluoroethanol and 10% D_2O for a final sample volume of 350 μL . The spectra were recorded at 4°C using a Bruker 600 MHz NMR spectrometer equipped with a cryogenic probe. Suppression of the water signal was achieved by excitation sculpting.¹⁸⁸ The x-carrier frequency was referenced to internal DSS-d6.

Residue specific assignments were obtained from 2D total correlated spectroscopy (TOCSY),¹⁸⁹ while 2D nuclear Overhauser effect spectroscopy (NOESY)¹⁹⁰ permitted sequence specific assignments. ^{13}C resonances were assigned from 2D ^{13}C - ^1H HSQC spectra. The TOCSY and NOESY mixing times were 70 and 200 ms, respectively. The D1 relaxation delay was 1.5 s.

PPI inhibition assay

All the AlphaScreen assays were performed in a final volume of 50 μL in 96-well Optiplates (PerkinElmer). 10 μL of EGF (His-tag functionalized, 3 nM final concentration in PBS pH 7.4 buffer, 0.1% BSA, 0.1% Tween-20) and 10 μL of EGFR-Fc (3 nM final concentration in the same buffer) were added to each well and incubated at room temperature for 15 min. Next, 10 μL of Protein A 3835 Donor beads (20 $\mu\text{g}/\text{mL}$ final concentration) were added and incubated in the dark for 60 min. 10 μL of anti-His acceptor beads (20 $\mu\text{g}/\text{mL}$ final concentration) were added to each well and incubated for 30 min in the dark before performing the fluorescence reading (emission at 615

nm) in a EnVision® Multilabel Reader. Dose-response curves were obtained with the GraphPad Prism 6.03 software, using a non-linear fit and variable slope, from which IC₅₀ and Hill slope values were calculated.

Binding of EGF to cells

EGF was derivatized with 5 eq of Alexa Fluor® 488 NHS Ester (Thermo Scientific). The labelled EGF species were purified by HPLC, and the purity was assessed by UPLC-MS.

The receptor-binding properties of AF488-labeled EGF were evaluated in a direct binding assay using A-431 human cancer cells (2.5×10^6 EGFR/cell).¹⁹¹ Briefly, various concentrations of AF488-labeled EGF (0.1-300 nM) in 120 μ L of PBS were incubated with 10^6 cells in 1.5 mL microtubes for 2 h at 4°C. Cells were washed by centrifugation at 1,000 rpm for 5 min, and the supernatant was discarded. Cells were washed once more with 120 μ L of PBS, then centrifuged and resuspended in 500 μ L of PBS. Propidium iodide (Sigma), at a final concentration of 1-2 μ g/mL, was used to distinguish live cells from dead cells during cell analysis. Cell fluorescence was analysed using a Beckman Coulter Gallios (Beckman Coulter). The number of events was stopped at 10,000 cells.

For the competition assays, various concentrations of peptides (5-1000 μ M) or cetuximab (0.1-1000 nM) in 120 μ L of PBS were added to AF488-labeled EGF (20 nM) and incubated with 10^6 cells in 1.5 mL microtubes for 2 h at 4°C. Cell bound fluorescence was measured as described above.

Cell viability assay

A-431 human cancer cells (2.5×10^6 EGFR/cell)¹⁹¹ were seeded in a 96-well Corning microplates (2×10^3 cells/well). After 24 h of incubation at 37°C, cells were treated with the peptides dissolved in Dulbecco's Modified Eagle's Medium - high glucose medium (concentration range 0.5-150 μ M, well volume 100 μ L) for 72 h. Afterwards, the peptide solution was removed and substituted with fresh medium. 50 μ L of activated-XTT solution (0.1 mL of the activation reagent in 5.0 mL of the XTT reagent, mixed 15 min before) was added to each well and incubate for 4 h at 37°C. The absorbance at 450 nm was measured in a PowerWave X reader (Bio-Tek, INC), where the positive control was cells treated with DMSO and the negative control was untreated cells. Each experiment was performed in triplicate, and the IC₅₀

was calculated by linear regression fitting (concentration vs. response curve) using GraphPad Prism.

The same conditions were used for the XTT experiments with MCF7 human cancer cells and J2P mouse fibroblasts.

Stability in human serum

Peptides at a final concentration of 150 μ M were dissolved in HBSS and incubated at 37°C in the presence of 90% human serum during 24 h. The 50 μ L aliquots were extracted at different incubation times and treated with 200 μ L of MeOH to precipitate serum proteins. After 30 min centrifugation at 4°C, the supernatant was filtered and analysed by RP-HPLC to calculate the percentage of intact peptide in the sample.

Stability in human hepatic microsomes

The compound was incubated at 1 μ M concentration as recommended by the microsomes supplier (Corning) for metabolic stability determinations. This is consistent with tissue concentration levels reached after oral dosing, but is not so high as to saturate the metabolizing enzymes. For the assay, the compound was dissolved at 4°C in water (713 μ L) and 200 μ L 0.5 M potassium phosphate pH 7.4 (100 μ M final concentration), 50 μ L NADPH regenerating system solution A, and 10 μ L NADPH regenerating system solution B were added.

This mixture was warmed to 37°C for 5 minutes. After which, 25 μ L (0.5 mg, stock solution was 20 mg/mL) of human liver microsomes were added. The mixture was immediately vortexed and incubated at 37°C with orbital agitation (100 rpm). After 0, 5, 10, 20, 30, 40, 50 and 60 minutes, 100 μ L of the sample were withdrawn and 100 μ L of acetonitrile were added to the extracted sample, mixing and placing it on wet ice afterwards.

After finishing the experiment, the extracted samples were centrifuged at 13000 rpm (4°C) for 20 minutes. The supernatant was withdrawn from the protein pellet and analysed by UPLC-MS. The experiment was performed in triplicate.

POP activity assay

POP was obtained by expression in *E. coli* and affinity purification as reported previously.¹⁹² POP activity was determined following the method

described by Toide *et al.*¹⁹³ The reactions were performed in 96-well microplates, thereby allowing the simultaneous monitoring of multiple reactions. For each reaction, the activity buffer (137 μL , 100 mM of Na/K phosphate buffer, pH 8.0) was pre-incubated for 15 min at 37°C with POP (5 nM) and with the corresponding inhibitor solution (3 μL). Stock solutions at a range of inhibitor concentrations (typically 1 mM to 5 mM) were prepared in DMSO. After pre-incubation, ZGP-AMC (N- benzyloxycarbonyl- Gly-Pro-methylcoumarinyl-7-amide, 10 μL , 450 μM in 40% of 1,4-dioxane) was added, and the reaction was monitored at the fluorimeter at 37°C . The formation of AMC was measured fluorimetrically (excitation and emission wavelengths were 360/40 and 485/20 nm, respectively). For each sample, a negative control containing 3 μL of the sample in DMSO, 10 μL of ZGP-AMC solution and 137 μL of activity buffer were added into the well plate. The positive control of the reaction consisted of 3 μL of DMSO, 10 μL of ZGP-AMC solution and 137 μL of 5 nM POP in activity buffer. Triplicates for each condition were included. Fluorescence values versus time were plotted and the lineal time was determined. Afterwards, the percentage of inhibition at each concentration was calculated and the IC₅₀ value was obtained. Data were analyzed using GraphPad Prism 6 software. For the kinetic determination of compound **18**, the same methodology was applied using a range of inhibitor concentrations (1-100 nM) and several preincubation times (0-60 min). Then, fluorogenic substrate ZGP-AMC was added and fluorescence was read after 30 minutes.

FAP activity assay

Recombinant fibroblast activation protein (FAP) was purchased from Sigma. The reactions were performed in 96-well microplates. For each reaction, the activity buffer (129.5 μL , 100 mM of Na/K phosphate buffer, pH 8.0) was pre-incubated for 15 min at 37°C with recombinant human FAP (7.5 μL , 0.57 nM) and with the corresponding inhibitor solution (3 μL). Stock solutions at a range of inhibitor concentrations were prepared in DMSO. After pre-incubation, ZGP-AMC (N-benzyloxycarbonyl-Gly-Pro-methylcoumarinyl-7-amide, 10 μL , 1.5 mM in 40% of 1,4-dioxane) was added, and the reaction was incubated for 60 min at 37°C in a fluorimeter. The formation of AMC was monitored fluorimetrically (excitation and emission wavelengths were 360/40 and 485/20 nm, respectively). For each sample, a negative control containing 3 μL of the sample in DMSO, 10 μL of ZGP-AMC solution, and 129.5 μL of activity buffer were added to the well plate. The positive control of the reaction consisted of 3 μL of DMSO, 10 μL

of ZGP-AMC solution, and 137 μL of 0.57 nM FAP in activity buffer. Triplicates for each condition were included. Fluorescence values versus time were represented and the lineal time was determined. Afterwards, the percentage of FAP inhibition was calculated at point concentrations of inhibitors (50, 10, 5 and 1 μM) by comparing the fluorescence intensity of the sample with that of the blank.

DPP-IV activity assay

The ectodomain (residues 39-766) of dipeptidyl peptidase-IV (DPP-IV) was obtained by expression in Sf9 insect cells using the Baculovirus technique. DPP-IV activity was determined following the method described by Checler *et al.*¹⁹⁴ The reactions were performed in 96-well microplates. For each reaction, the activity buffer (132 μL , 50 mM Tris, 1 M NaCl, 1 mg/mL BSA, pH 7.5) was pre-incubated for 15 min at 37°C with recombinant human DPP-IV (5 μL , 12.3 $\mu\text{g}/\text{mL}$) and with the corresponding inhibitor solution (3 μL). Stock solutions at a range of inhibitor concentrations were prepared in DMSO. After pre-incubation, HGP-AMC (N-Gly-Pro-methylcoumarinyl-7-amide, 10 μL , 0.75 mM in 40% of 1,4-dioxane) was added, and the reaction was incubated for 20 min at 37°C. The formation of AMC was monitored fluorimetrically every 5 min (excitation and emission wavelengths were 360/40 and 485/20 nm, respectively). For each sample, a negative control containing 3 μL of the sample in DMSO, 10 μL of HGP-AMC solution and 132 μL of activity buffer were added to the well plate. The positive control of the reaction consisted of 3 μL of DMSO, 10 μL of HGP-AMC solution and 132 μL of 12.3 $\mu\text{g}/\text{mL}$ DPP-IV in activity buffer. Triplicates for each condition were included. Fluorescence values versus time were represented and the lineal time was determined. Afterwards, the percentage of DPP-IV inhibition was calculated at point concentrations of inhibitors (50, 10, 5 and 1 μM) by comparing the fluorescence intensity of the sample with that of the blank.

Parallel Artificial Membrane Permeability Assay (PAMPA)

The PAMPA was used to determine the capacity of the compounds to cross the BBB by passive diffusion.¹⁷⁴ The effective permeability of the compounds was measured at an initial concentration of 200 μM . The PAMPA buffer solution was prepared from a concentrated one (commercialized by pION Inc.) and the pH was adjusted to 6.8. 1-propanol (20%) was added as a

cosolvent to the buffer solution. Stock solutions of all compounds at 20 mM concentration were prepared in DMSO, and 1:100 dilution with PAMPA buffer was performed in each case. For the assays, the PAMPA sandwich was disabled, and the donor wells were filled with 200 μ L of the compound solution of interest. Before this, the polycarbonate filter of the membrane was coated with 4 μ L of the commercial mixture of phospholipids. The acceptor compartments were filled with PAMPA buffer containing 1% DMSO (200 μ L). After this, both plates were reassembled. The PAMPA sandwich was incubated in humidity-saturated chamber for 4 h with orbital agitation at 100 rpm. After this time, the donor and acceptor wells were analyzed and quantified by HPLC at $\lambda=220$ nm. Propranolol was used as a positive control. Experiments were performed in triplicate. The phospholipid mixture used was a porcine polar brain lipid extract. Composition: phosphatidylcholine (PC) 12.6%, phosphatidyl-ethanolamine (PE) 33.1%, phosphatidylserine (PS) 18.5%, phosphatidylinositol (PI) 4.1%, phosphatidic acid 0.8%, and 30.9% of other compounds. After 4 h, the effective permeability (P_e), the standard parameter that quantifies transport independently of time and concentrations, was calculated using *Equation 6*:

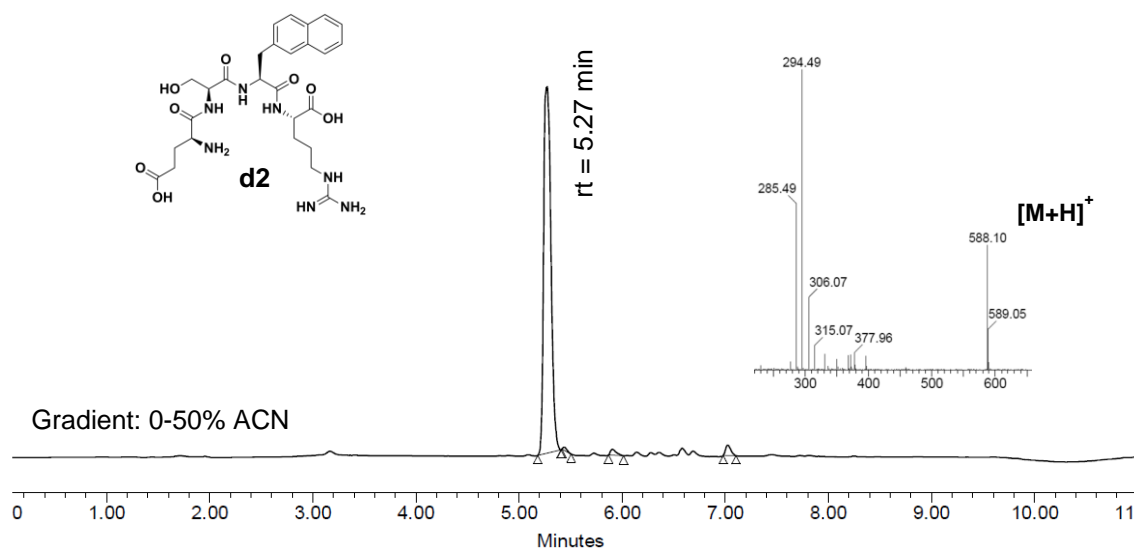
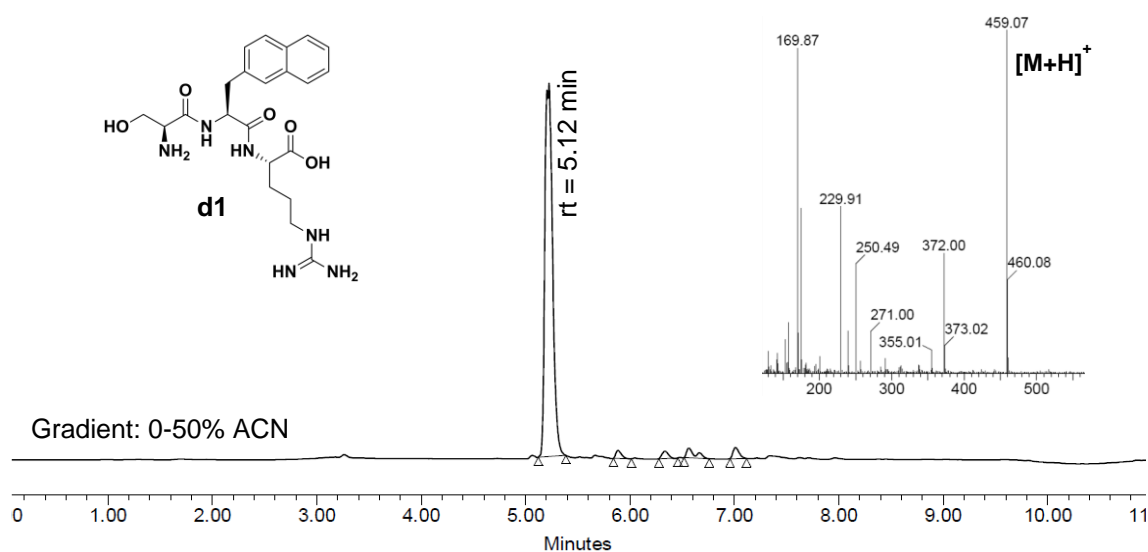
$$P_e = \frac{-218.3}{t} \times \log \left[1 - \frac{2C_A(t)}{C_D(t_0)} \right] \times 10^{-6} \text{ cm/s} \quad \text{Equation 6}$$

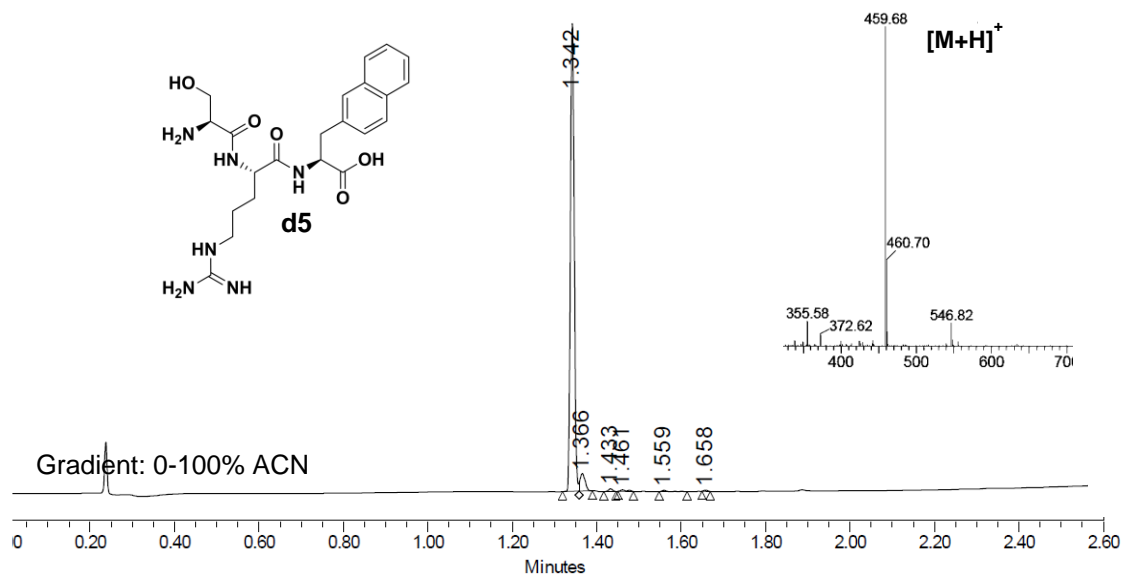
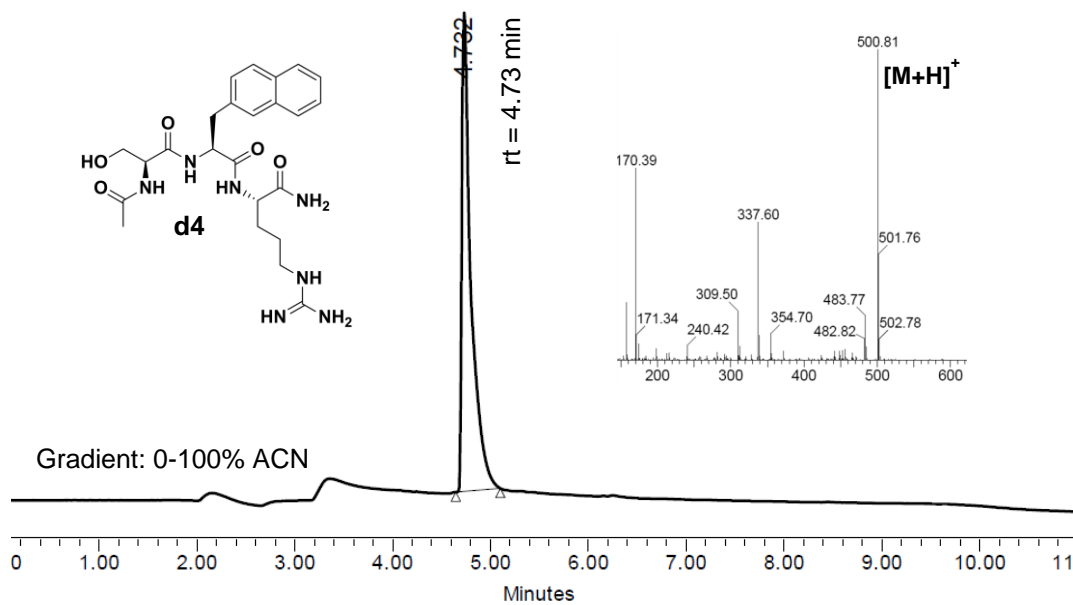
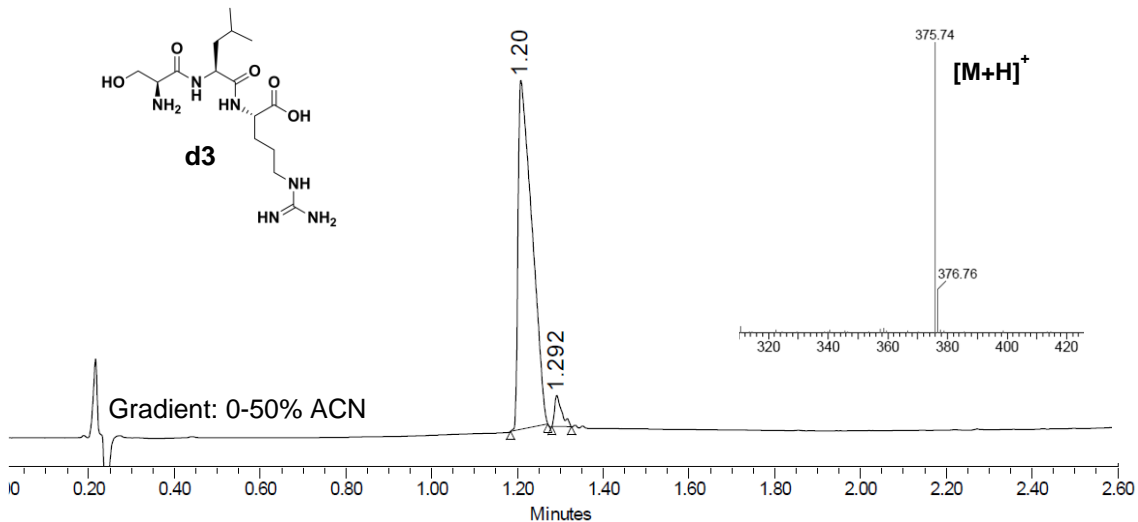
where t is the running time (4 h), $C_A(t)$ is the concentration of compound in the acceptor well at time t and $C_D(t_0)$ is the concentration of compound in the donor well at $t = 0$ h. Transport (%) values were obtained by dividing the amount in the acceptor well at time t , $C_A(t)$, by the amount in the donor well at time zero, $C_D(t_0)$, multiplied by 100. Permeability was considered excellent when values were $>4.0 \times 10^{-6}$ cm/s, uncertain between 2.0×10^{-6} and 4.0×10^{-6} cm/s, and poor below 2.0×10^{-6} cm/s.¹⁷⁴

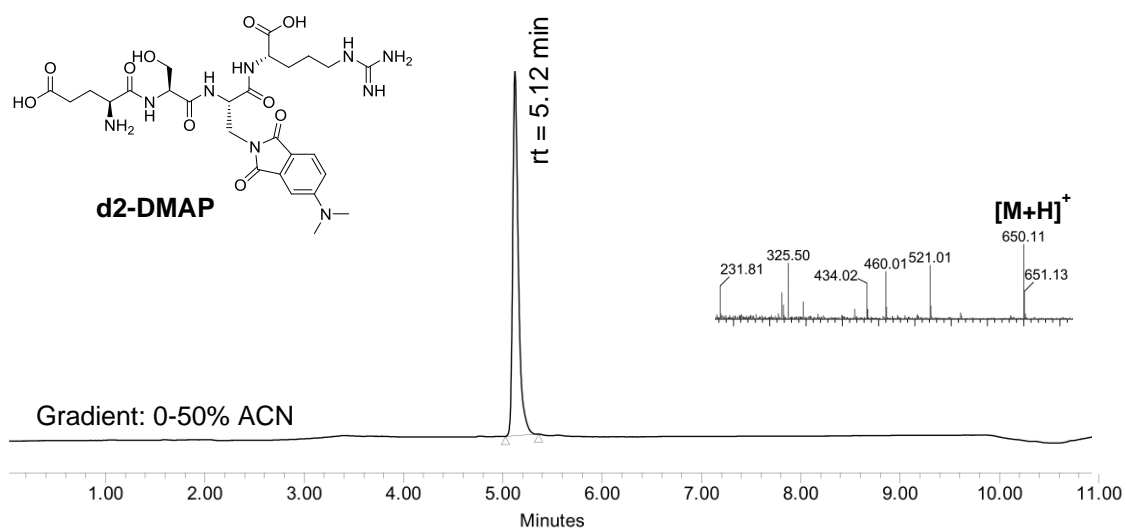
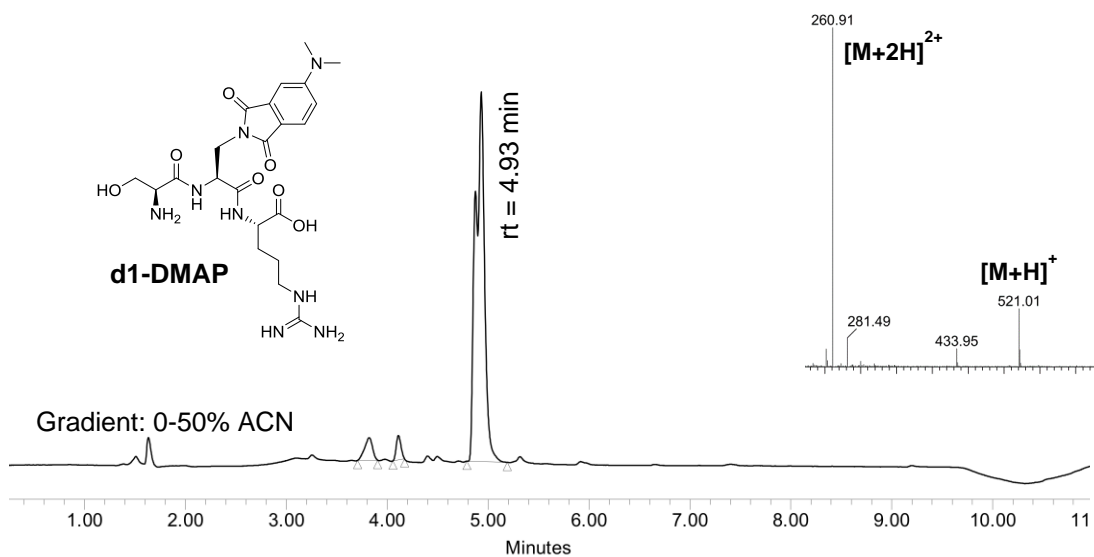
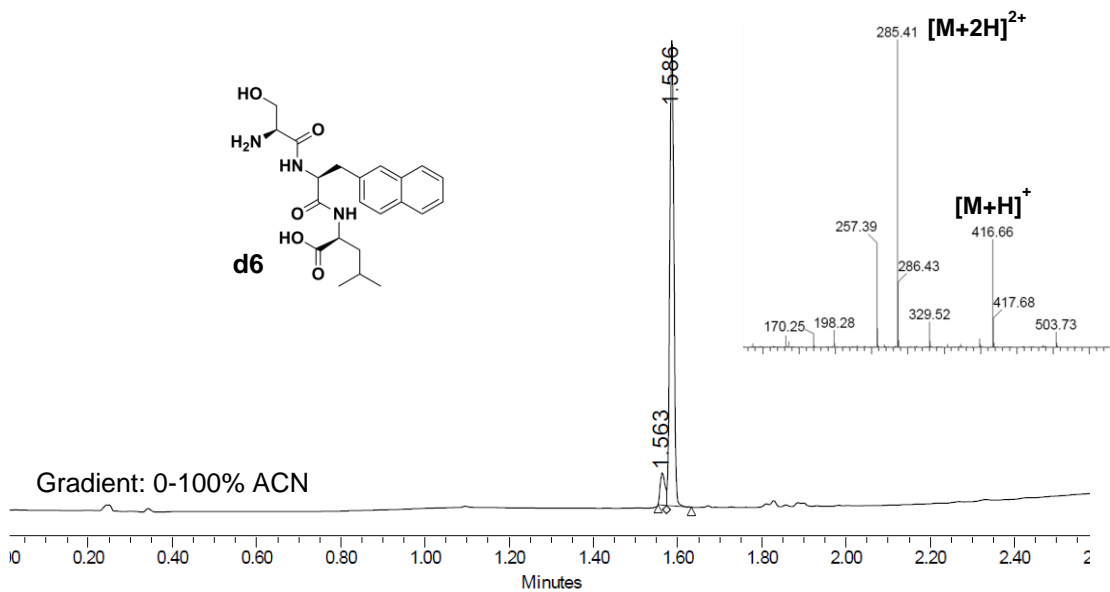
Product characterization

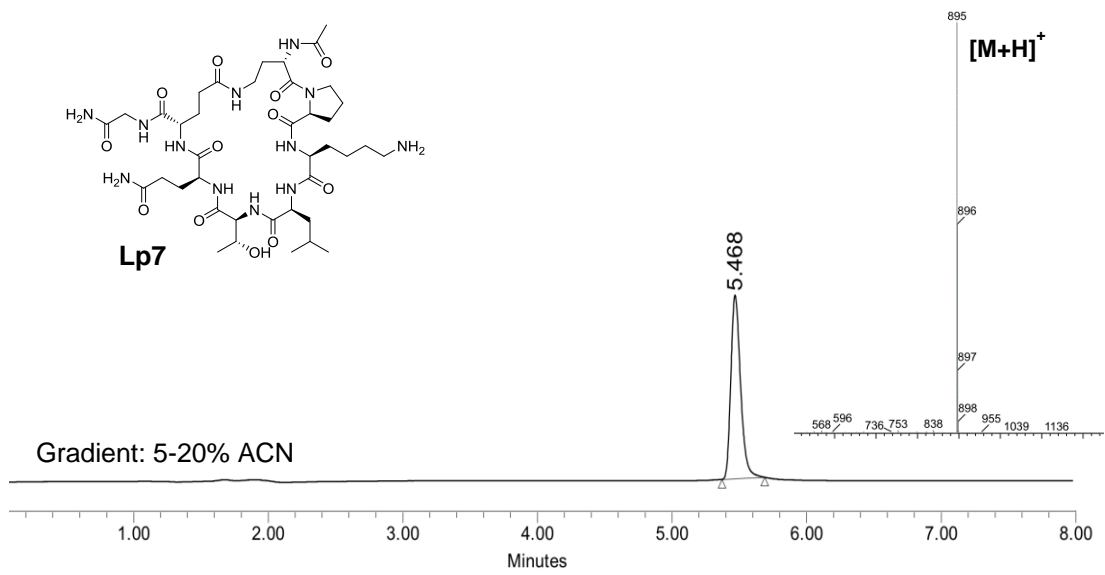
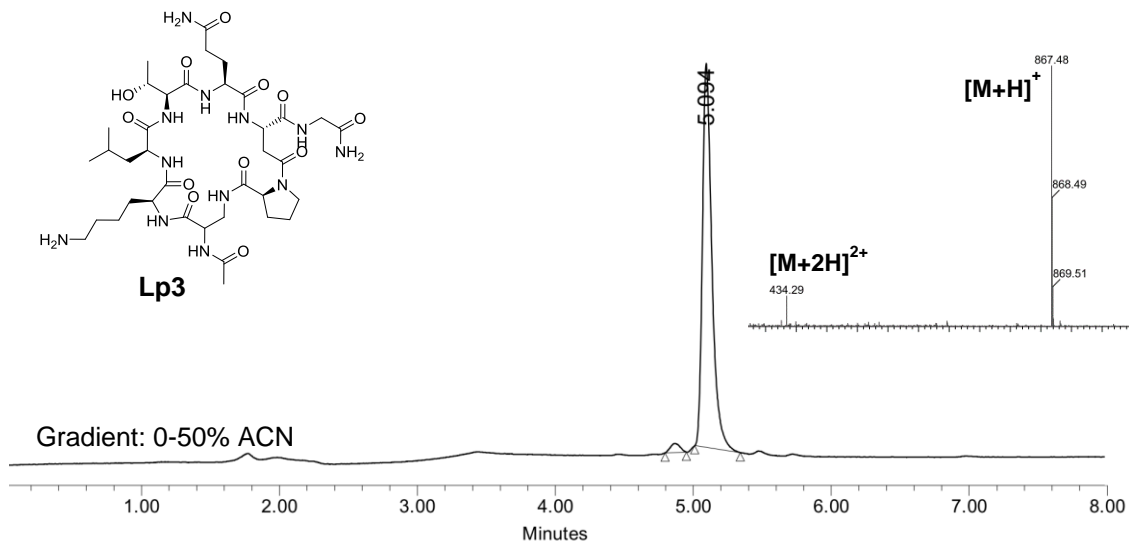
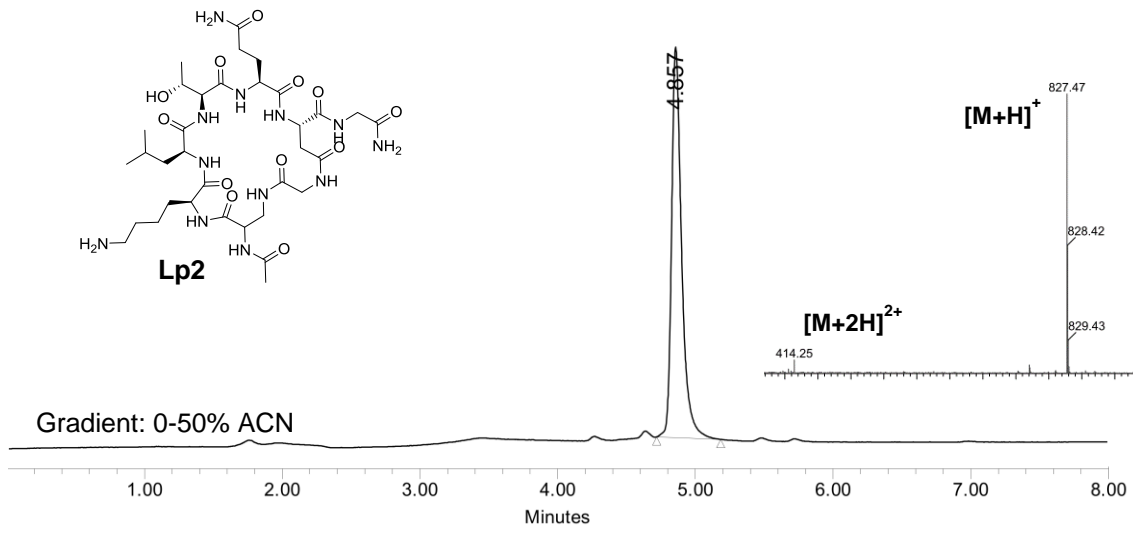
Peptides targeting EGF

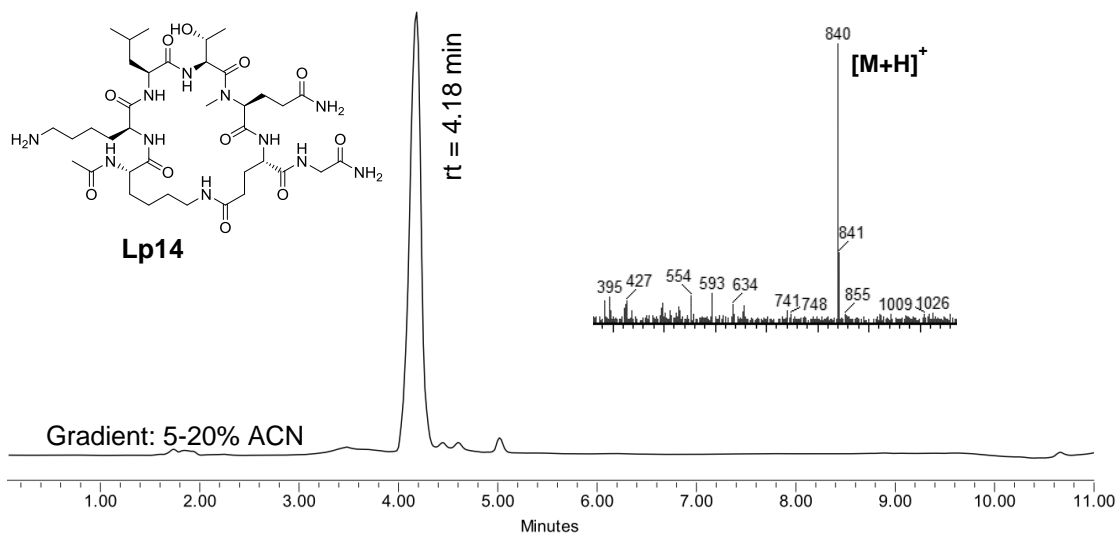
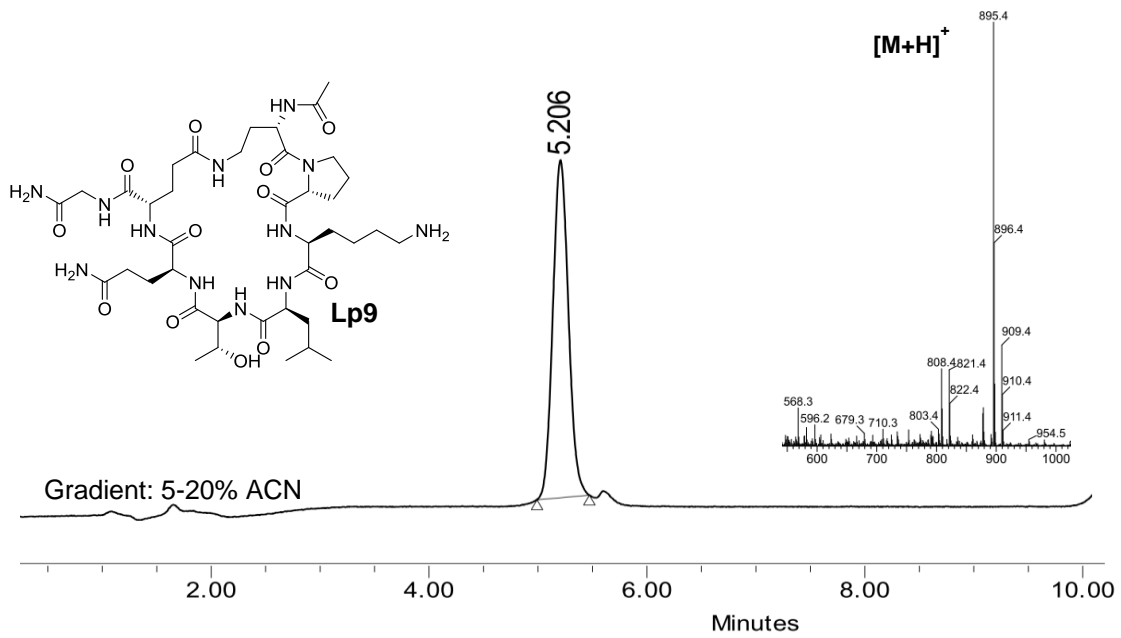
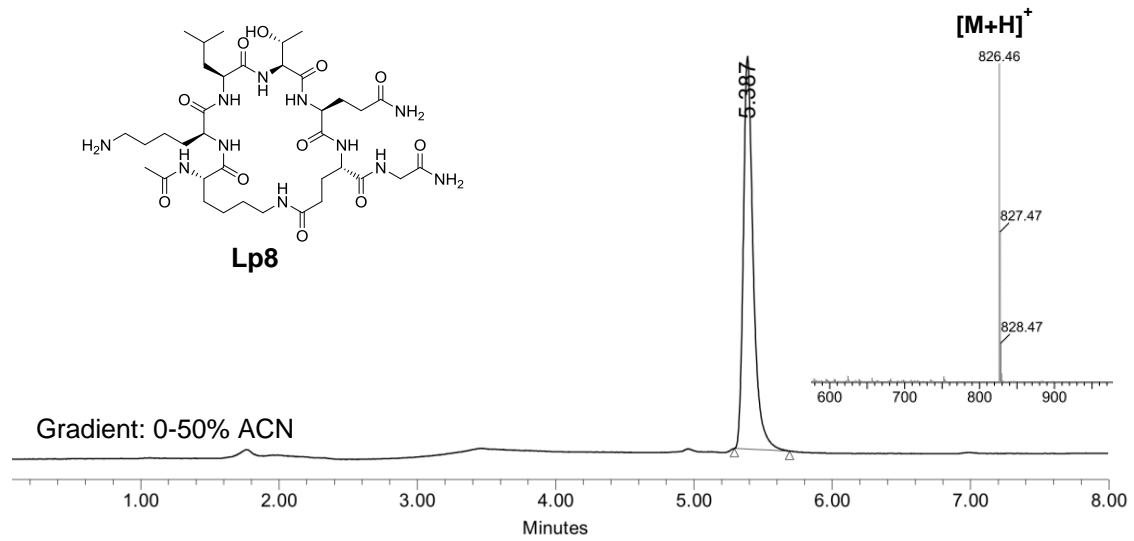
HPLC or UPLC chromatograms and UPLC-MS or LC-ESI-Q mass spectra are shown for each synthesized peptides. All HPLC chromatograms were recorded at 220 nm in an 8-min linear gradient, using as solvents ACN (0.036% TFA) and H₂O (0.045% TFA). All UPLC chromatograms were recorded at 220 nm in a 2-min linear gradient, using as solvents ACN (0.036% TFA) and H₂O (0.045% TFA). The observed [M+H]⁺ values are shown on the mass spectra.



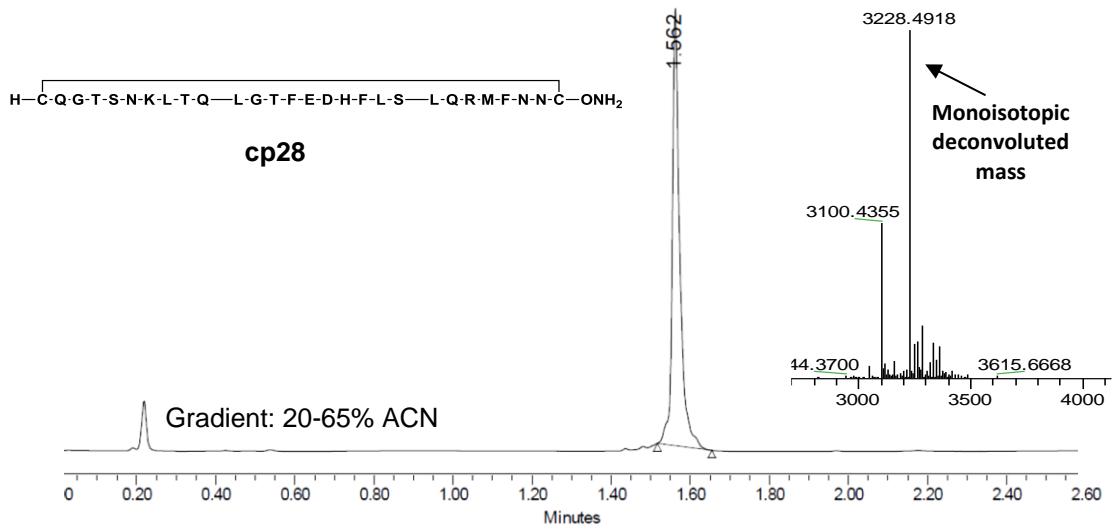
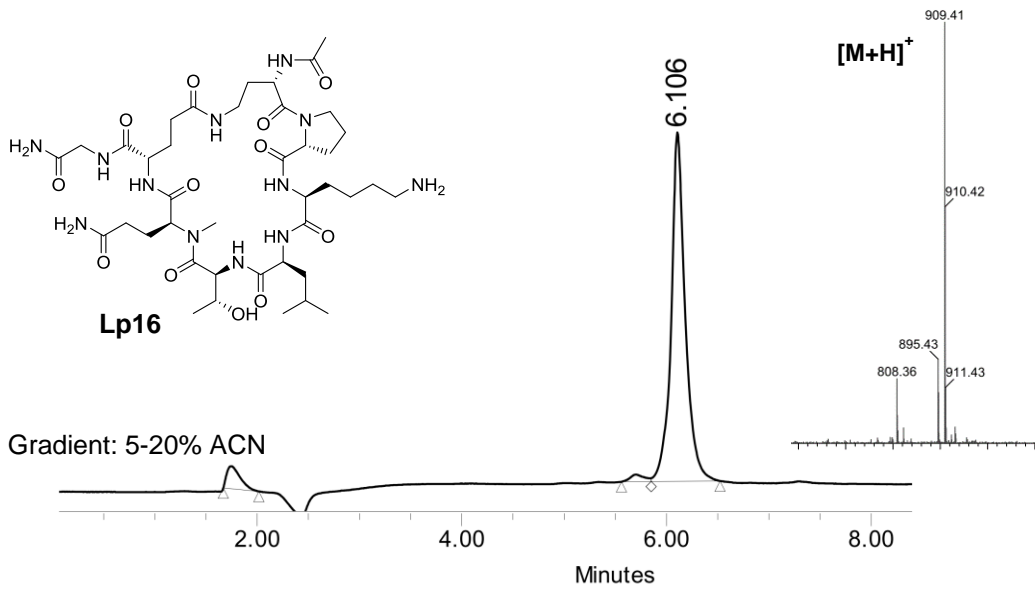
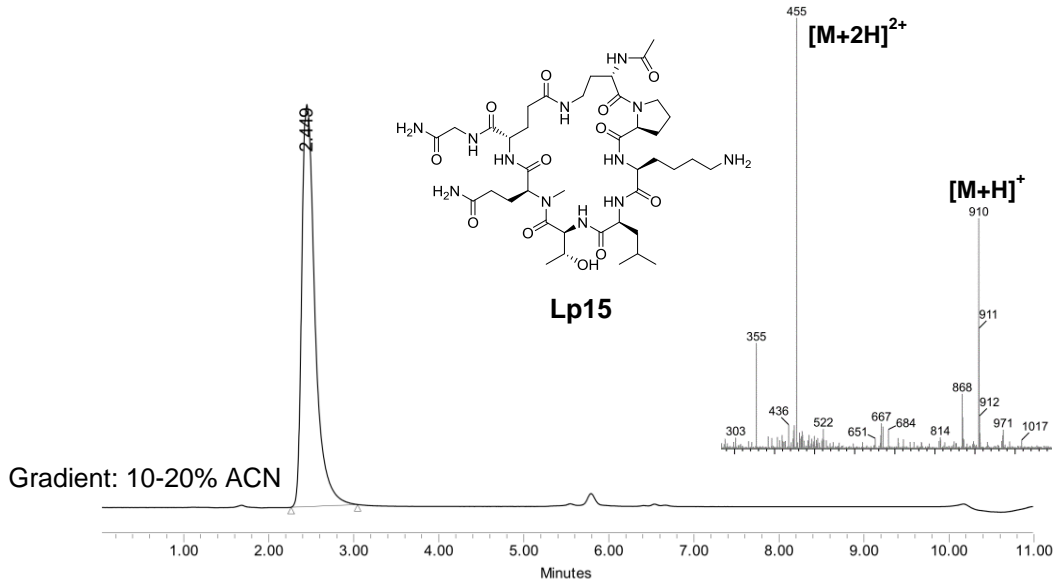


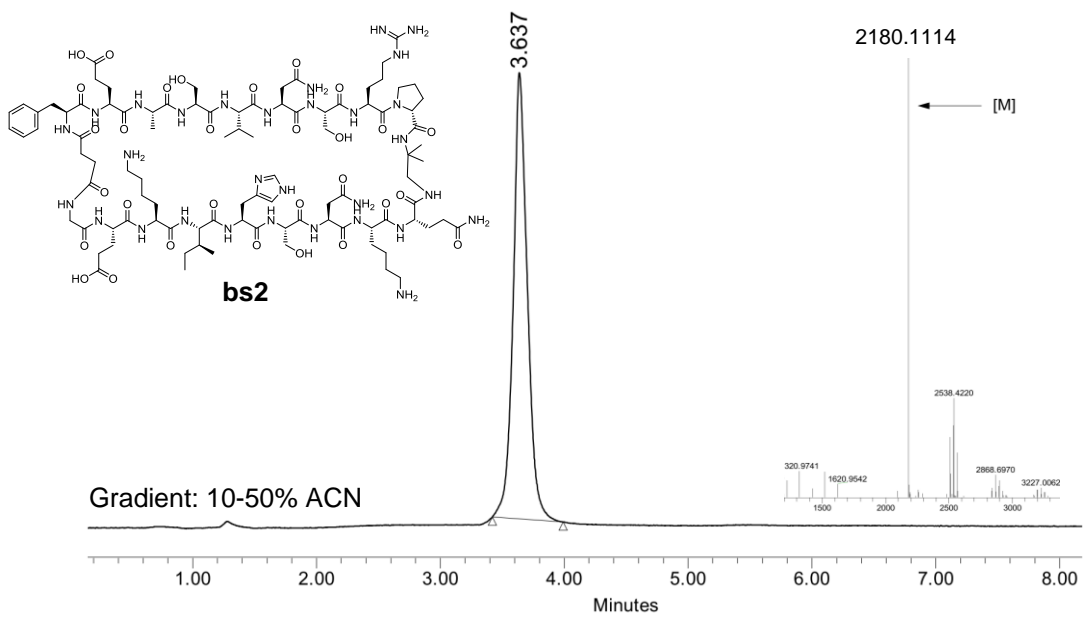
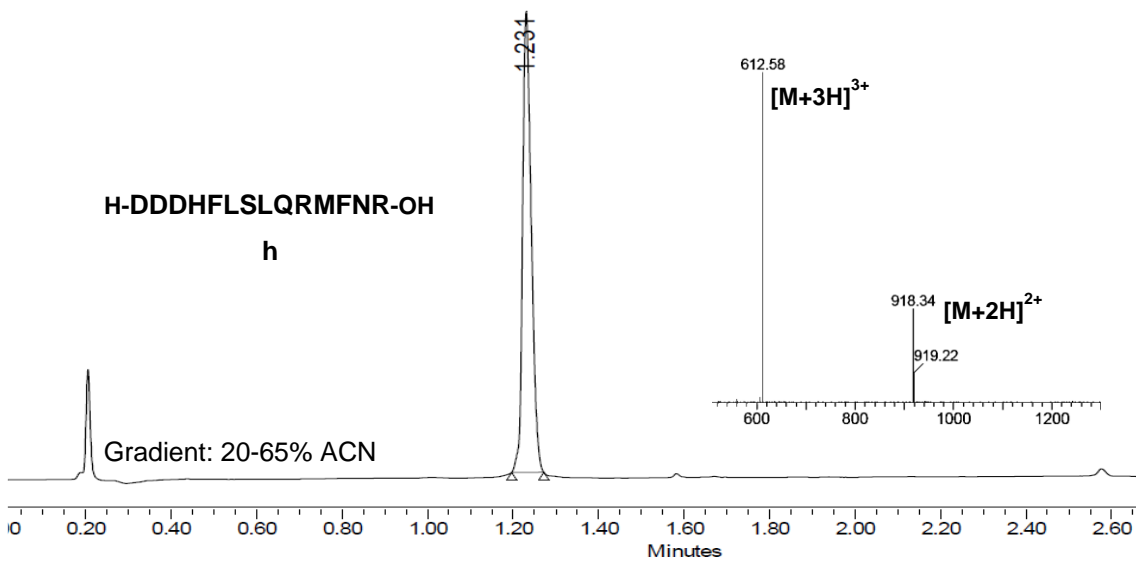
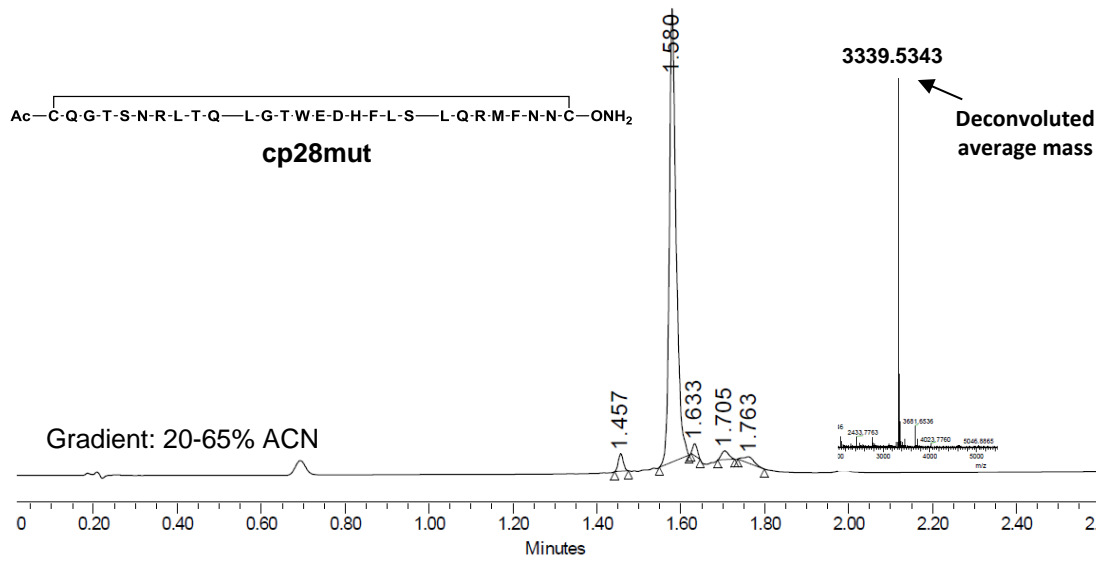


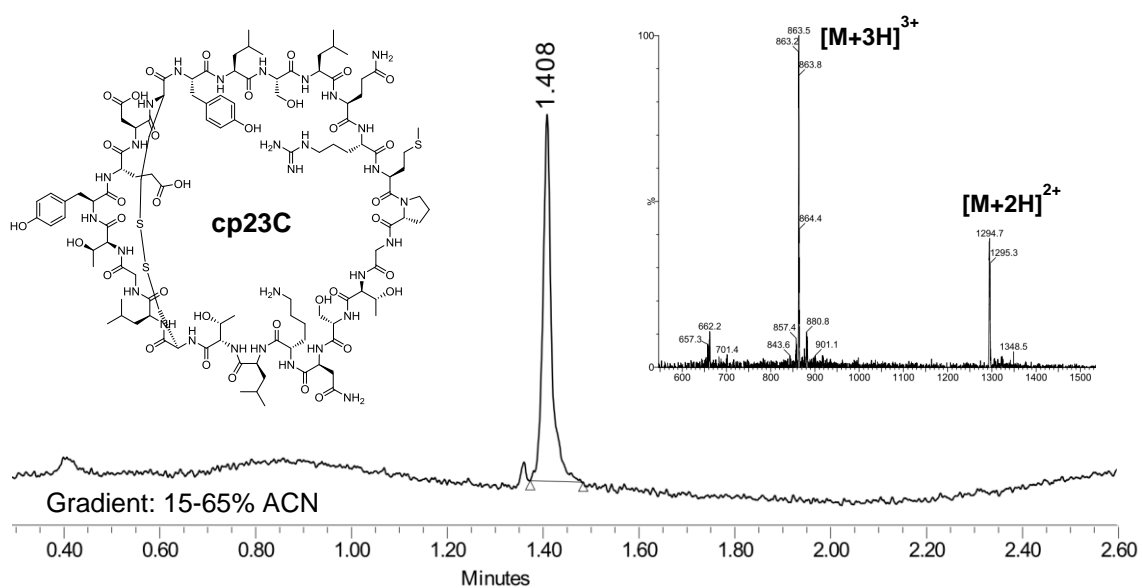
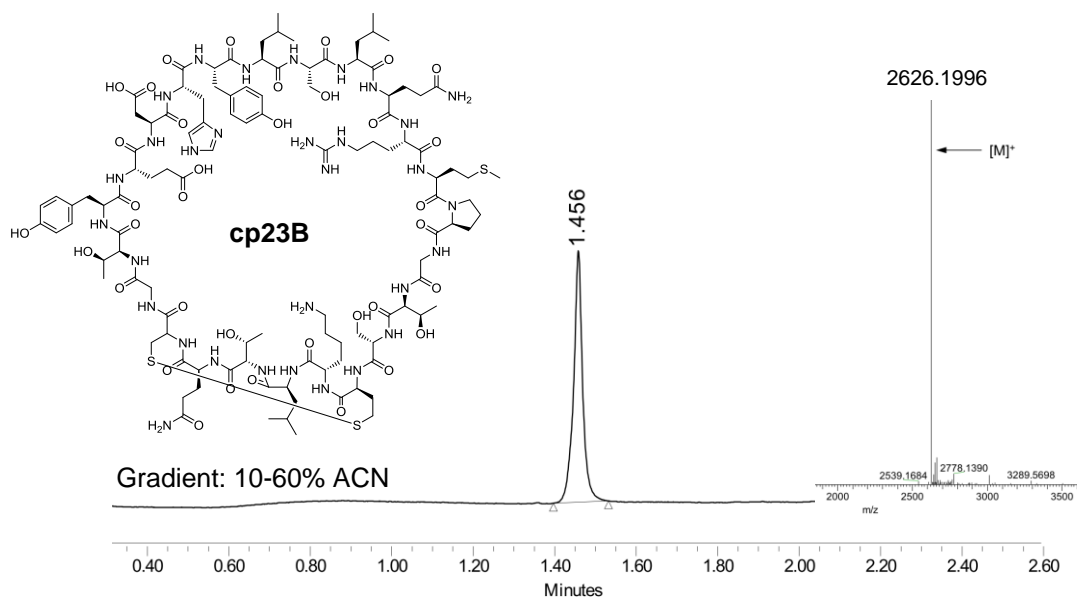
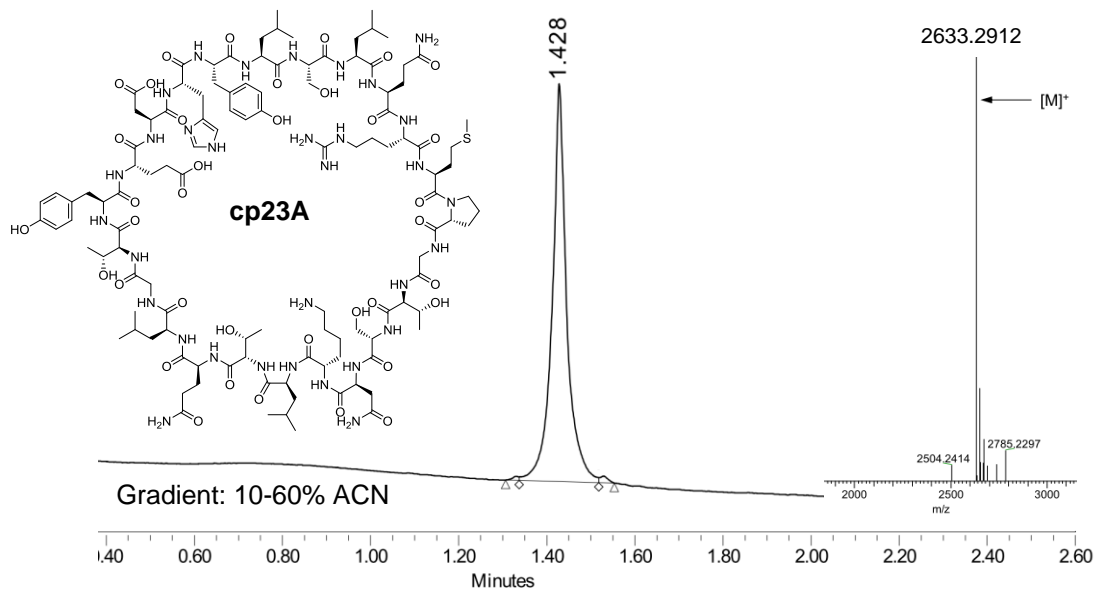


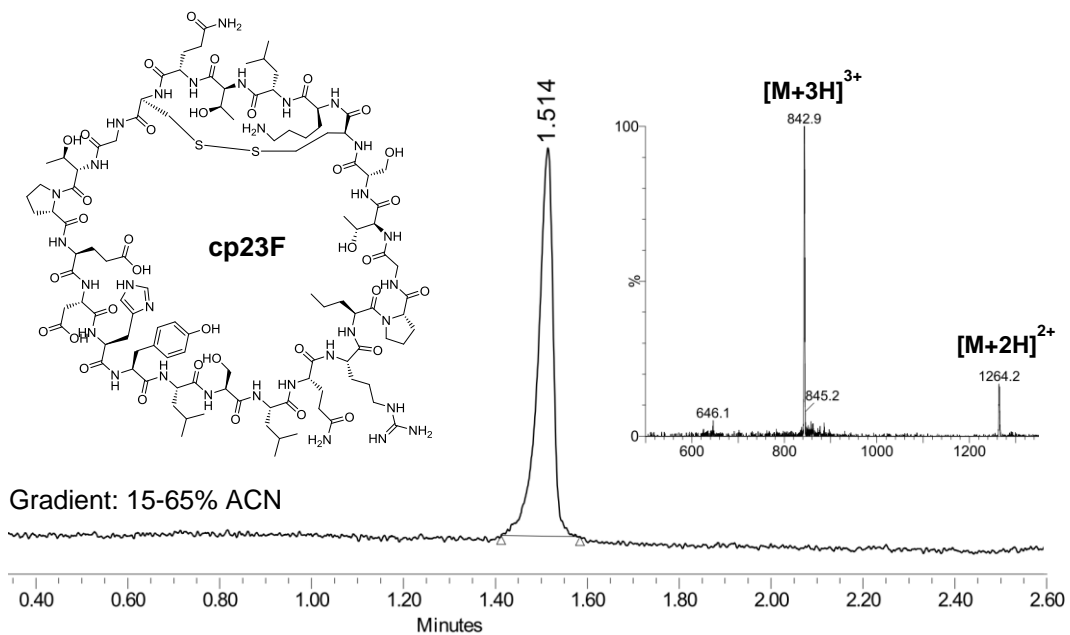
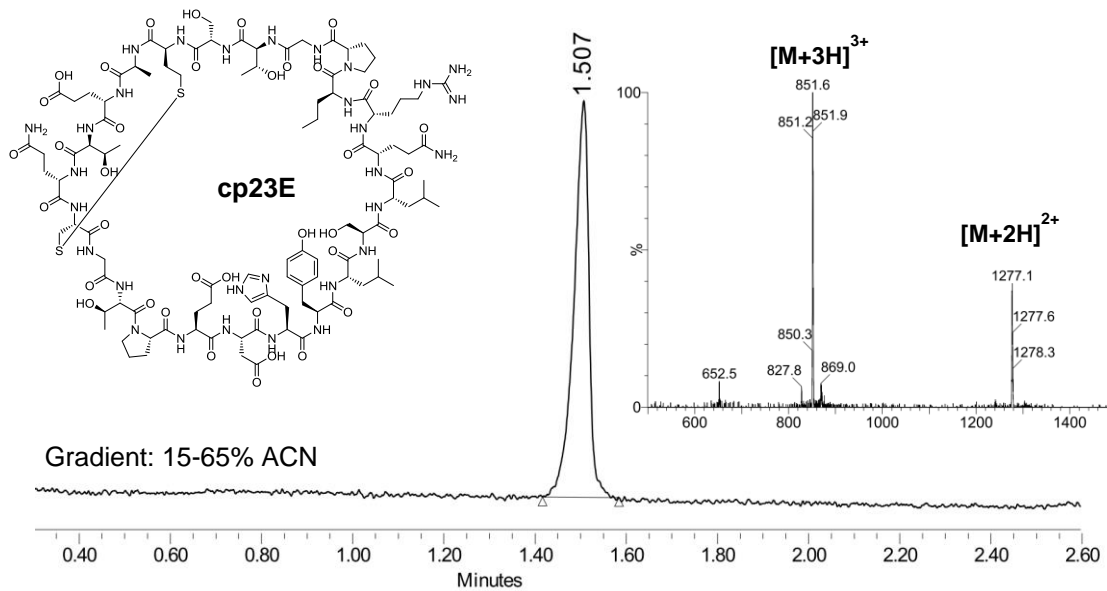
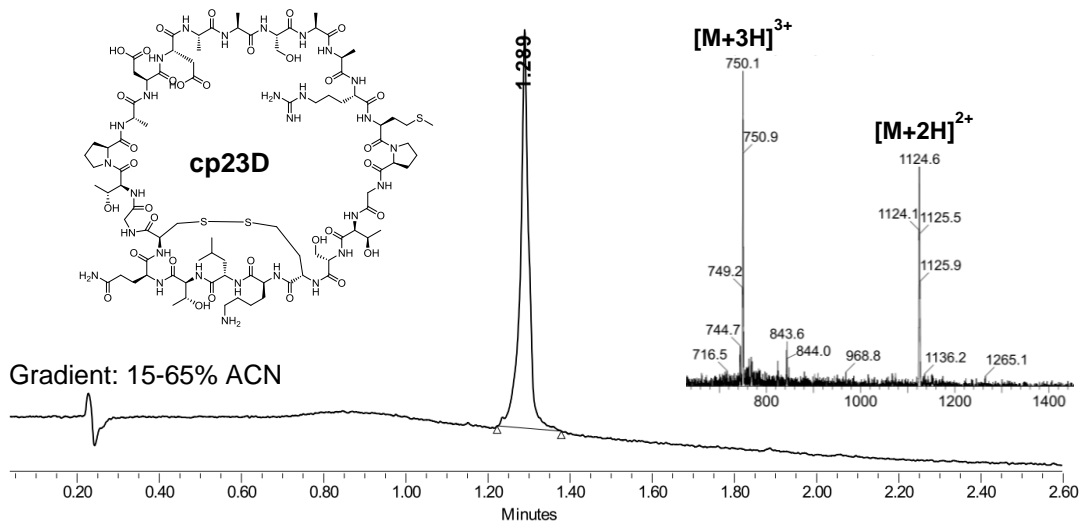


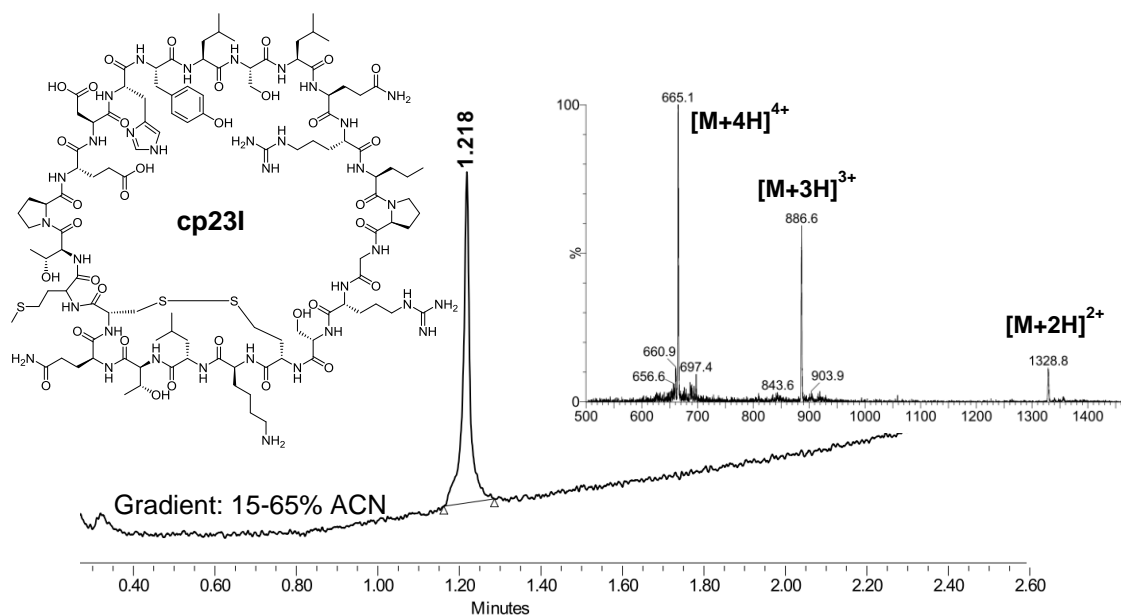
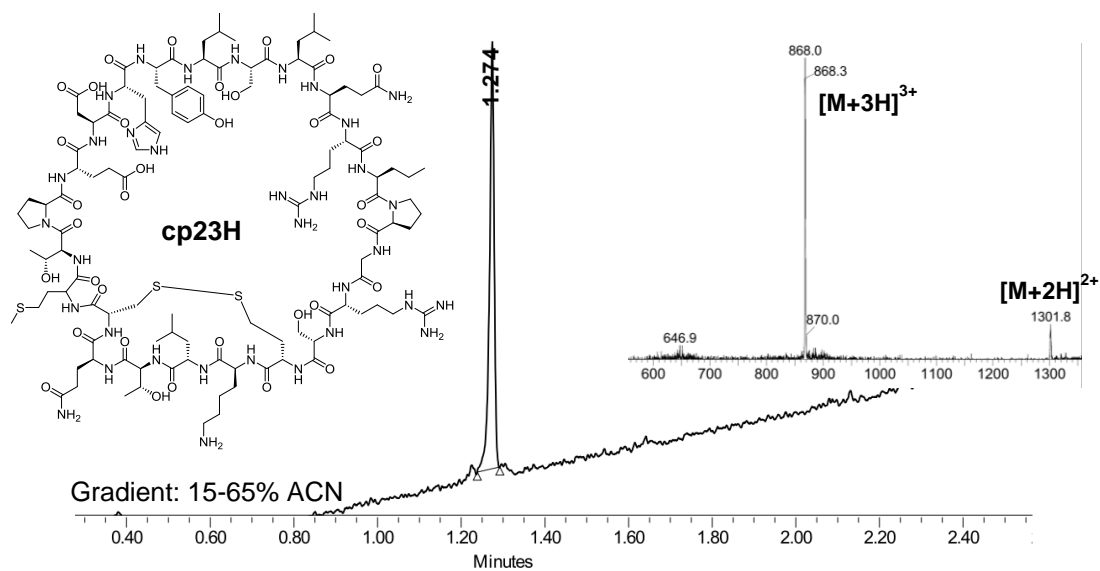
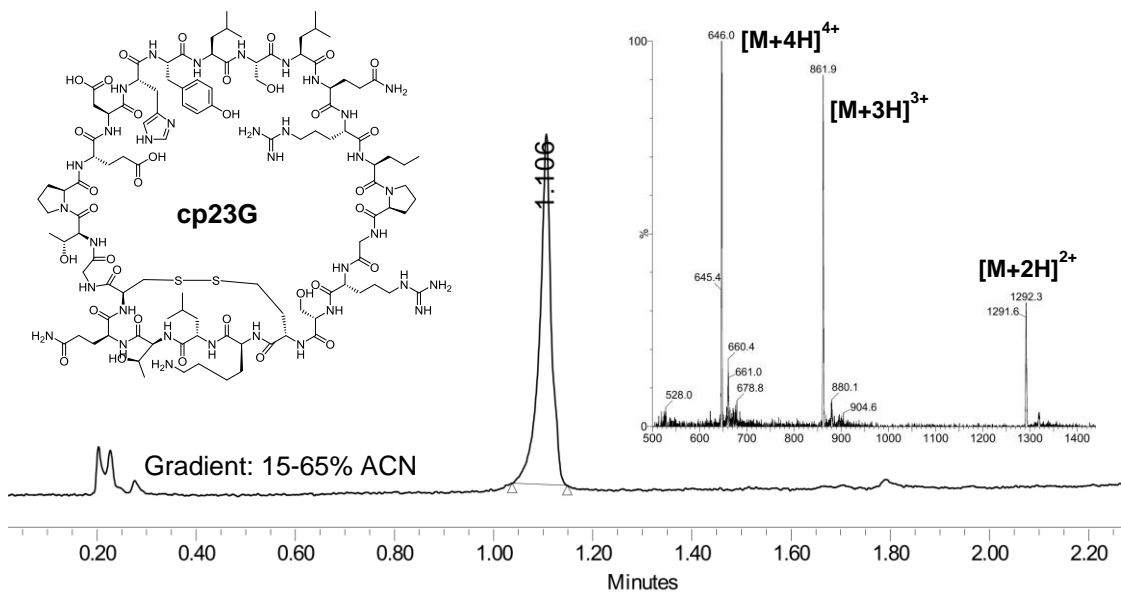
<











Peptidomimetics targeting POP

NMR spectra were recorded on Varian Mercury 400, Agilent 400, or Varian Inova 500 MHz spectrometers, using CDCl₃ as solvent at 25°C. Chemical shifts are given in parts per million (ppm) (δ relative to residual solvent peak for ¹H and ¹³C. Most of the ¹H NMR and ¹³C NMR spectra contained additional peaks, due to the presence of rotamers and F atoms. Exact mass determination was performed on a LTQ-FT spectrometer (Thermo Scientific) by direct infusion of sample.

(S)-benzyl 2-((S)-2-((fluorosulfonyl)methyl)pyrrolidine-1-carbonyl)pyrrolidine-1-carboxylate (7). White solid. Yield: 75%. ¹H NMR (500 MHz, CDCl₃) δ 7.40 – 7.27 (m, 5H), 5.21 – 5.11 (m, 1H), 5.09 – 4.86 (m, 1H), 4.64 – 4.57 (m, 1H), 4.55 – 4.37 (m, 1H), 4.31 – 4.23 (m, 1H), 3.89 – 3.78 (m, 2H), 3.70 – 3.56 (m, 2H), 3.55 – 3.48 (m, 1H), 3.38 – 3.32 (m, 1H), 2.29 – 1.84 (m, 6H), 1.83 – 1.48 (m, 1H). ¹³C NMR (126 MHz, CDCl₃) δ 172.54 (d, J = 24.2 Hz), 155.38, 136.40 (d, J = 16.5 Hz), 128.61, 128.18, 127.95, 67.65 (d, J = 36.0 Hz), 58.30 (d, J = 78.2 Hz), 53.25 (d, J = 26.6 Hz), 51.84 (t, J = 13.1 Hz), 47.35 (d, J = 15.1 Hz), 46.96 (d, J = 15.4 Hz), 30.64, 29.61, 28.56 (d, J = 27.9 Hz), 24.45, 23.97 (d, J = 19.9 Hz). HRMS (ESI+) for C₁₈H₂₄O₅N₂F₁S₁ (M + H) calcd., 399.13845; found, 399.13851.

(2S,4R)-benzyl 4-fluoro-2-((S)-2-((fluorosulfonyl)methyl)pyrrolidine-1-carbonyl)pyrrolidine-1-carboxylate (8). White solid. Yield: 51%. ¹H NMR (400 MHz, CDCl₃) δ 7.39 – 7.28 (m, 5H), 5.34 (t, J = 3.2 Hz, 1H), 5.23 – 5.18 (m, 1H), 5.18 – 4.89 (m, 1H), 4.75 – 4.54 (m, 1H), 4.24 – 4.16 (m, 1H), 4.08 – 3.51 (m, 5H), 3.36 (t, J = 6.8 Hz, 1H), 2.57 – 2.41 (m, 1H), 2.37 – 2.04 (m, 3H), 2.03 – 1.75 (m, 1H), 1.72 – 1.30 (m, 1H). ¹³C NMR (126 MHz, CDCl₃) δ 171.80 (d, J = 6.4 Hz), 154.98, 153.95, 135.86 (d, J = 19.5 Hz), 128.79, 128.60, 128.52, 128.17, 127.81, 91.72 (dd, J = 178.5, 93.6 Hz), 67.85 (d, J = 59.4 Hz), 56.37 (d, J = 69.3 Hz), 53.66 (dd, J = 51.7, 22.5 Hz), 53.19 (d, J = 28.7 Hz), 51.68 (d, J = 13.4 Hz), 47.34 (d, J = 53.6 Hz), 37.58 (d, J = 22.0 Hz), 36.58 (d, J = 22.1 Hz), 28.49 (d, J = 52.7 Hz), 23.95 (d, J = 76.6 Hz). HRMS (ESI+) for C₁₈H₂₃O₅N₂F₂S₁ (M + H) calcd., 417.12903; found, 417.12931.

(S)-benzyl 4,4-difluoro-2-((S)-2-((fluorosulfonyl)methyl)pyrrolidine-1-carbonyl)pyrrolidine-1-carboxylate (9). White solid. Yield: 41%. ¹H NMR (500 MHz, CDCl₃) δ 7.40 – 7.27 (m, 5H), 5.21 – 5.05 (m, 1H), 4.69 (dd, J = 8.3, 7.3 Hz, 1H), 4.65 – 4.54 (m, 1H), 4.08 – 3.82 (m, 4H), 3.80 – 3.65 (m, 1H), 3.56 – 3.46 (m, 1H), 3.32 (t, J = 6.9 Hz, 1H), 2.71 – 2.57 (m, 1H), 2.55 – 2.38 (m, 1H), 2.34 – 1.95 (m, 3H), 1.92 – 1.37 (m, 1H). ¹³C NMR (126 MHz,

CDCl₃) δ 170.15, 154.65, 135.59, 130.10 – 126.05 (m), 68.08 (d, J = 38.1 Hz), 55.89 (dd, J = 49.0, 4.2 Hz), 53.25 (d, J = 28.1 Hz), 51.58 (d, J = 13.5 Hz), 47.21 (d, J = 45.7 Hz), 39.34 – 37.02 (m), 28.46 (d, J = 48.6 Hz), 24.02 (d, J = 71.6 Hz). HRMS (ESI+) for C₁₈H₂₂O₅N₂F₃S₁ (M + H) calcd., 435.11960; found, 435.11957.

(2S,4S)-benzyl 2-((S)-2-((fluorosulfonyl)methyl)pyrrolidine-1-carbonyl)-4-(trifluoromethyl)pyrrolidine-1-carboxylate (10). White solid. Yield: 48%. ¹H NMR (400 MHz, CDCl₃) δ 7.41 – 7.26 (m, 5H), 5.22 – 4.88 (m, 2H), 4.66 – 4.41 (m, 2H), 4.08 – 3.41 (m, 5H), 3.32 (t, J = 6.8 Hz, 1H), 3.05 – 2.88 (m, 1H), 2.56 – 2.40 (m, 1H), 2.37 – 1.92 (m, 4H), 1.80 – 1.40 (m, 1H). ¹³C NMR (101 MHz, CDCl₃) δ 170.80, 136.13, 128.74 (d, J = 2.8 Hz), 128.42, 128.17, 67.75, 57.75, 57.34, 53.30 (d, J = 16.4 Hz), 51.93 (d, J = 13.7 Hz), 47.21 (d, J = 36.3 Hz), 42.39 (q, J = 29.7 Hz), 29.29, 28.87, 28.50, 24.29 (d, J = 52.1 Hz). HRMS (ESI+) for C₁₉H₂₃O₅N₂F₄S₁ (M + H) calcd., 467.12583; found, 467.12547.

(2S,4S)-benzyl 2-((S)-2-((fluorosulfonyl)methyl)pyrrolidine-1-carbonyl)-4-methylpyrrolidine-1-carboxylate (11). White solid. Yield: 62%. ¹H NMR (400 MHz, CDCl₃) δ 7.41 – 7.27 (m, 5H), 5.20 – 5.02 (m, 1H), 4.68 – 4.58 (m, 1H), 4.52 – 4.32 (m, 1H), 3.94 – 3.72 (m, 3H), 3.60 – 3.43 (m, 1H), 3.32 (t, J = 6.8 Hz, 1H), 3.14 – 3.02 (m, 1H), 2.40 – 1.40 (m, 8H), 1.08 (dd, J = 10.1, 6.4 Hz, 3H). ¹³C NMR (101 MHz, CDCl₃) δ 172.67, 155.14, 136.38, 128.82, 128.69, 128.63, 128.23, 128.00, 67.76 (d, J = 37.2 Hz), 58.89 (d, J = 34.1 Hz), 54.04 (d, J = 32.8 Hz), 53.33 (d, J = 20.8 Hz), 51.77, 47.22 (d, J = 33.7 Hz), 38.22 (d, J = 100.9 Hz), 33.45 (d, J = 57.5 Hz), 28.55 (d, J = 29.3 Hz), 24.16 (d, J = 48.0 Hz), 16.63 (d, J = 11.9 Hz). HRMS (ESI+) for C₁₉H₂₆O₅N₂F₁S₁ (M + H) calcd., 413.15410; found, 413.15409.

((S)-1-((S)-1-(4-phenylbutanoyl)pyrrolidine-2-carbonyl)pyrrolidin-2-yl)methanesulfonyl fluoride (12). White solid. Yield: 59%. ¹H NMR (400 MHz, CDCl₃) δ 7.30 – 7.23 (m, 2H), 7.20 – 7.14 (m, 3H), 4.61 (dd, J = 8.1, 4.1 Hz, 1H), 4.58 – 4.43 (m, 1H), 3.92 (dt, J = 9.7, 7.1 Hz, 1H), 3.85 – 3.74 (m, 2H), 3.64 – 3.56 (m, 1H), 3.55 – 3.38 (m, 2H), 2.67 (t, J = 7.5 Hz, 2H), 2.42 – 1.89 (m, 13H). ¹³C NMR (101 MHz, CDCl₃) δ 172.80, 171.90, 141.49, 128.60, 128.47, 126.07, 58.22, 53.15, 52.05, 51.92, 47.61, 47.37, 35.20, 33.55, 28.91, 28.73, 26.13, 24.85, 24.39. HRMS (ESI+) for C₂₀H₂₈O₄N₂F₁S₁ (M + H) calcd., 411.17483; found, 411.17508.

((S)-1-((2S,4R)-4-fluoro-1-(4-phenylbutanoyl)pyrrolidine-2-carbonyl)pyrrolidin-2-yl)methanesulfonyl fluoride (13). White solid. Yield: 43%. ¹H NMR (400 MHz, CDCl₃) δ 7.33 – 7.23 (m, 2H), 7.24 – 7.13 (m,

3H), 5.41 – 5.35 (m, 1H), 5.28 – 5.22 (m, 0H), 4.78 – 4.69 (m, 1H), 4.64 – 4.54 (m, 1H), 4.06 (dt, $J = 9.7, 7.0$ Hz, 1H), 3.91 – 3.81 (m, 1H), 3.81 – 3.69 (m, 3H), 3.55 (dt, $J = 9.6, 6.7$ Hz, 1H), 2.66 (t, $J = 7.4$ Hz, 2H), 2.57 – 2.42 (m, 1H), 2.36 – 2.25 (m, 3H), 2.23 – 2.04 (m, 4H), 2.03 – 1.91 (m, 2H). ^{13}C NMR (101 MHz, CDCl_3) δ 173.04, 171.54, 141.08, 128.43 (d, $J = 2.7$ Hz), 126.04, 92.10 (d, $J = 179.4$ Hz), 56.64, 53.96 (d, $J = 22.6$ Hz), 53.21, 51.73 (d, $J = 13.4$ Hz), 47.70, 35.82 (d, $J = 21.8$ Hz), 34.90, 33.40, 28.71, 25.94, 24.18. HRMS (ESI+) for $\text{C}_{20}\text{H}_{27}\text{O}_4\text{N}_2\text{F}_2\text{S}_1$ ($\text{M} + \text{H}$) calcd., 429.16541; found, 429.16524.

((S)-1-((S)-4,4-difluoro-1-(4-phenylbutanoyl)pyrrolidine-2-carbonyl)pyrrolidin-2-yl)methanesulfonyl fluoride (14). White solid. Yield: 28%. ^1H NMR (400 MHz, CDCl_3) δ 7.32 – 7.26 (m, 2H), 7.22 – 7.15 (m, 3H), 4.74 (t, $J = 8.0$ Hz, 1H), 4.62 – 4.53 (m, 1H), 4.05 – 3.84 (m, 2H), 3.85 – 3.70 (m, 3H), 3.55 – 3.45 (m, 1H), 2.71 – 2.64 (m, 2H), 2.64 – 2.55 (m, 1H), 2.55 – 2.38 (m, 1H), 2.37 – 2.23 (m, 3H), 2.22 – 2.03 (m, 3H), 2.03 – 1.91 (m, 2H). ^{13}C NMR (101 MHz, CDCl_3) δ 172.51, 170.12, 141.03, 128.44, 126.09, 55.76 (d, $J = 4.8$ Hz), 53.83 (dd, $J = 33.6, 29.9$ Hz), 51.65 (d, $J = 13.5$ Hz), 47.53, 36.99 (t), 34.82, 33.19, 28.68, 25.75, 24.26. HRMS (ESI+) for $\text{C}_{20}\text{H}_{26}\text{O}_4\text{N}_2\text{F}_3\text{S}_1$ ($\text{M} + \text{H}$) calcd., 447.15599; found, 447.15608.

((S)-1-((2S,4S)-1-(4-phenylbutanoyl)-4-(trifluoromethyl)pyrrolidine-2-carbonyl)pyrrolidin-2-yl)methanesulfonyl fluoride (15). White solid. Yield: 40%. ^1H NMR (400 MHz, CDCl_3) δ 7.31 – 7.26 (m, 2H), 7.22 – 7.16 (m, 3H), 4.65 – 4.47 (m, 1H), 4.01 – 3.91 (m, 2H), 3.86 – 3.78 (m, 2H), 3.77 – 3.58 (m, 2H), 3.56 – 3.44 (m, 1H), 3.05 – 2.89 (m, 1H), 2.75 – 2.59 (m, 2H), 2.53 – 2.38 (m, 1H), 2.37 – 1.89 (m, 9H). ^{13}C NMR (101 MHz, CDCl_3) δ 172.67, 170.68, 141.02, 128.45, 128.42, 126.09, 57.56, 53.24, 51.68 (d, $J = 13.4$ Hz), 47.51, 46.73, 42.65 (d, $J = 29.5$ Hz), 34.83, 33.26, 28.62, 25.79, 24.25. HRMS (ESI+) for $\text{C}_{21}\text{H}_{27}\text{O}_4\text{N}_2\text{F}_4\text{S}_1$ ($\text{M} + \text{H}$) calcd., 479.16222; found, 479.16238.

((S)-1-((2S,4S)-4-methyl-1-(4-phenylbutanoyl)pyrrolidine-2-carbonyl)pyrrolidin-2-yl)methanesulfonyl fluoride (16). White solid. Yield: 46%. ^1H NMR (400 MHz, CDCl_3) δ 7.31 – 7.24 (m, 2H), 7.22 – 7.15 (m, 3H), 4.64 – 4.47 (m, 2H), 4.01 – 3.64 (m, 2H), 3.64 – 3.55 (m, 1H), 3.55 – 3.33 (m, 2H), 3.12 (t, $J = 10.1$ Hz, 1H), 2.78 – 2.58 (m, 3H), 2.40 – 1.89 (m, 9H), 1.67 – 1.52 (m, 1H), 1.07 (dd, $J = 18.2, 6.5$ Hz, 3H). ^{13}C NMR (101 MHz, CDCl_3) δ 172.53, 172.14, 141.45, 128.62, 128.49, 126.10, 58.89, 54.69, 53.21, 52.04, 47.53, 47.18, 37.08, 35.17, 34.17, 33.57, 28.78, 26.07 (d, $J = 22.1$ Hz), 24.38, 16.59. HRMS (ESI+) for $\text{C}_{21}\text{H}_{30}\text{O}_4\text{N}_2\text{F}_4\text{S}_1$ ($\text{M} + \text{H}$) calcd., 425.19048; found, 425.19089.

((S)-1-((S)-1-(4-(benzyloxy)-3,5-difluorobenzoyl)pyrrolidine-2-carbonyl)pyrrolidin-2-yl)methanesulfonyl fluoride (17). Clear oil. Yield: 39%. ¹H NMR (400 MHz, CDCl₃) δ 7.45 – 7.40 (m, 2H), 7.39 – 7.32 (m, 3H), 7.19 – 7.11 (m, 2H), 5.23 (s, 2H), 4.76 (dd, *J* = 8.0, 6.3 Hz, 1H), 4.66 – 4.56 (m, 1H), 4.09 – 3.97 (m, 1H), 3.88 – 3.81 (m, 2H), 3.76 – 3.65 (m, 1H), 3.63 – 3.51 (m, 2H), 2.39 – 1.83 (m, 8H). ¹³C NMR (126 MHz, CDCl₃) δ 171.61, 167.01, 156.45, 154.46, 135.98, 129.38 – 127.59 (m), 112.92 – 111.11 (m), 75.96, 58.80, 53.14, 51.84 (d, *J* = 13.2 Hz), 50.39, 47.34, 28.79 (d, *J* = 27.3 Hz), 25.64, 24.35. HRMS (ESI+) for C₂₄H₂₆O₅N₂F₃S₁ (M + H) calcd., 511.15090; found, 511.15110.

((S)-1-((2S,4R)-1-(4-(benzyloxy)-3,5-difluorobenzoyl)-4-fluoropyrrolidine-2-carbonyl)pyrrolidin-2-yl)methanesulfonyl fluoride (18). Clear oil. Yield: 26%. ¹H NMR (400 MHz, CDCl₃) δ 7.45 – 7.40 (m, 2H), 7.39 – 7.32 (m, 3H), 7.19 – 7.10 (m, 2H), 5.35 (t, *J* = 3.4 Hz, 1H), 5.24 (s, 2H), 4.97 (dd, *J* = 9.4, 7.6 Hz, 1H), 4.65 – 4.54 (m, 1H), 4.24 – 4.14 (m, 1H), 4.01 – 3.75 (m, 4H), 3.64 – 3.55 (m, 1H), 2.63 – 2.50 (m, 1H), 2.39 – 2.05 (m, 5H). ¹³C NMR (101 MHz, CDCl₃) δ 171.67, 129.13, 129.03, 128.73, 112.87 – 112.36 (m), 93.41, 91.61, 76.44, 57.55, 56.89 (d, *J* = 22.5 Hz), 53.76, 52.21 (d, *J* = 13.5 Hz), 48.09, 30.17, 29.29, 28.65, 24.88, 24.80. HRMS (ESI+) for C₂₄H₂₅O₅N₂F₄S₁ (M + H) calcd., 529.14148; found, 529.14152.

((S)-1-((S)-1-(4-(benzyloxy)-3,5-difluorobenzoyl)-4,4-difluoropyrrolidine-2-carbonyl)pyrrolidin-2-yl)methanesulfonyl fluoride (19). Clear oil. Yield: 18%. ¹H NMR (400 MHz, CDCl₃) δ 7.45 – 7.40 (m, 2H), 7.40 – 7.30 (m, 3H), 7.16 – 7.07 (m, 2H), 5.24 (d, *J* = 3.8 Hz, 2H), 4.96 (t, *J* = 8.4 Hz, 1H), 4.65 – 4.53 (m, 1H), 4.18 – 4.02 (m, 2H), 3.94 – 3.75 (m, 2H), 3.61 – 3.53 (m, 1H), 3.50 (q, *J* = 7.0 Hz, 1H), 2.76 – 2.47 (m, 1H), 2.32 – 2.08 (m, 5H). ¹³C NMR (101 MHz, CDCl₃) δ 156.70, 128.54, 128.41, 128.09, 112.02, 75.81, 65.72, 55.99, 53.13, 51.54, 51.40, 47.35, 28.63, 24.19, 15.01. HRMS (ESI+) for C₂₄H₂₄O₅N₂F₅S₁ (M + H) calcd., 547.13206; found, 547.13219.

((S)-1-((2S,4S)-1-(4-(benzyloxy)-3,5-difluorobenzoyl)-4-methylpyrrolidine-2-carbonyl)pyrrolidin-2-yl)methanesulfonyl fluoride (21). Clear oil. Yield: 33%. ¹H NMR (500 MHz, CDCl₃) δ 7.51 – 7.40 (m, 2H), 7.40 – 7.29 (m, 3H), 7.23 – 7.10 (m, 2H), 4.74 (dd, 1H), 4.64 – 4.54 (m, 1H), 4.23 – 4.02 (m, 1H), 3.92 – 3.75 (m, 2H), 3.68 – 3.60 (m, 1H), 3.59 – 3.52 (m, 1H), 3.49 – 3.38 (m, 1H), 3.34 (t, *J* = 10.6 Hz, 1H), 2.44 – 2.05 (m, 6H), 1.73 – 1.62 (m, 2H), 1.09 (d, *J* = 6.5 Hz, 3H). ¹³C NMR (126 MHz, CDCl₃) δ 171.79, 128.77, 128.69, 128.42, 112.39 (d), 112.20 (d), 76.12, 59.38, 57.48, 53.27,

52.06, 51.96, 47.45, 37.05, 34.84, 29.84, 28.90, 24.53, 16.31.HRMS (ESI+) for $C_{25}H_{28}O_5N_2F_3S_1$ (M + H) calcd., 525.16655; found, 525.16661.

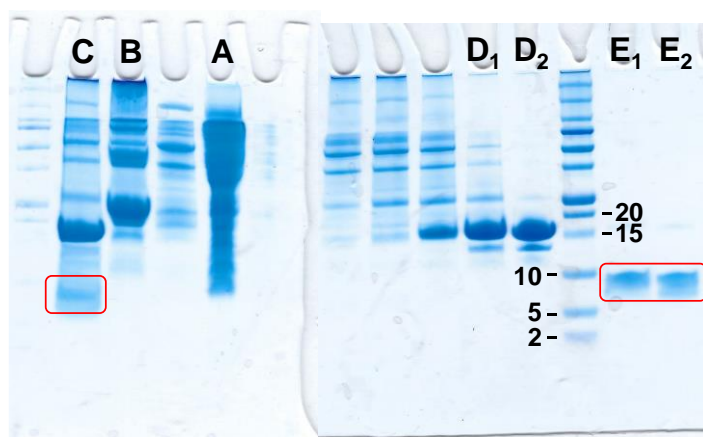
Recombinant EGF protein characterization

Sequence

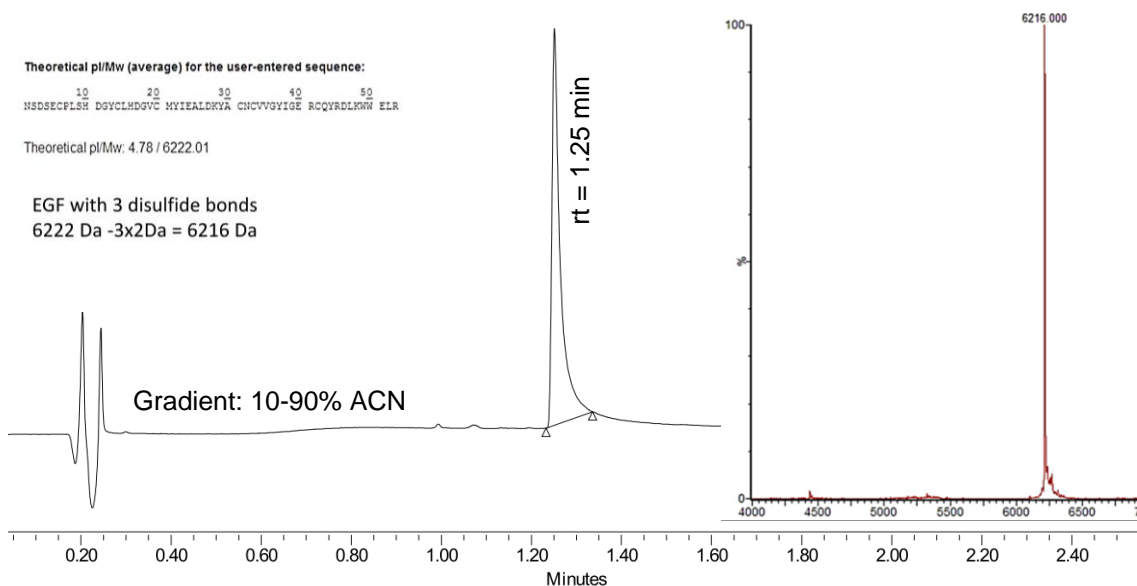
NSDSECLSHDGYCLHDGVCMYIEALDKYACNCVVG YIGERCQYRDLK
WWELR

SDS-PAGE of the different fractions obtained in the purification

A: flow-through NiNTA; B: elution of the NiNTA-bound proteins; C: cleavage of the SUMO-EGF fusion protein; D: SUMO fractions; E: EGF fractions. EGF bands are highlighted:

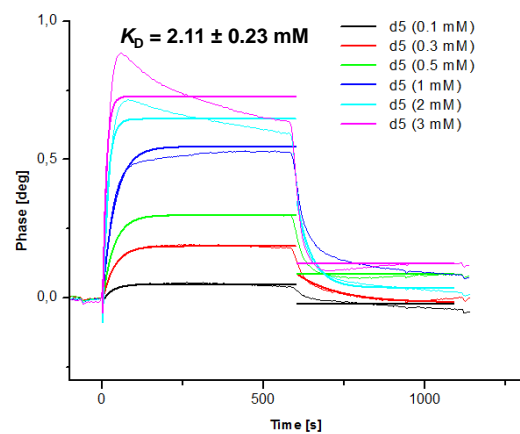
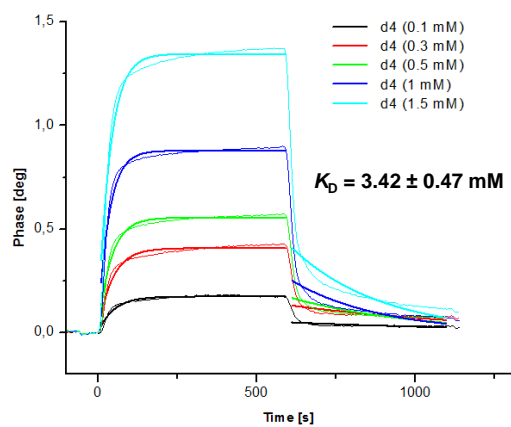
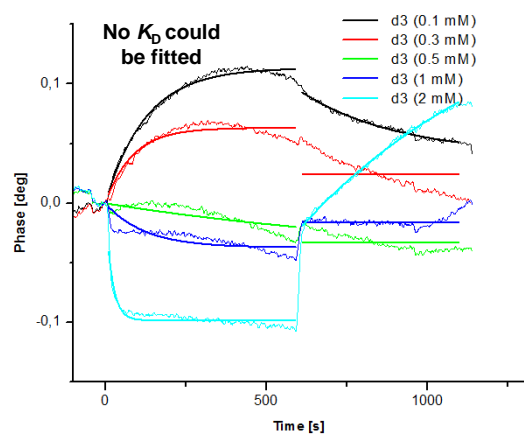
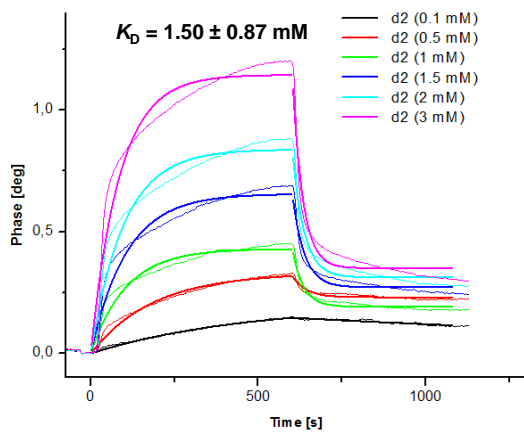
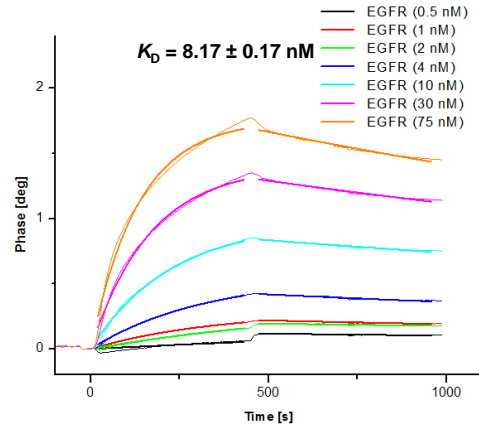
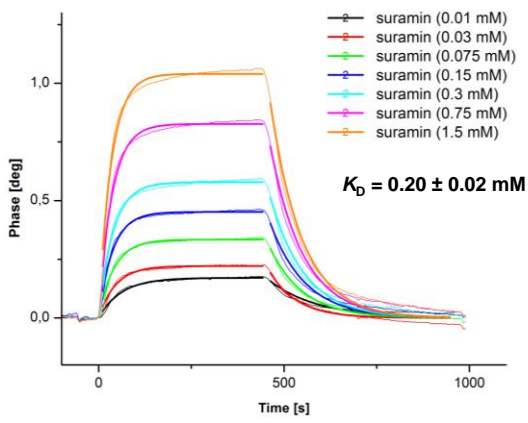


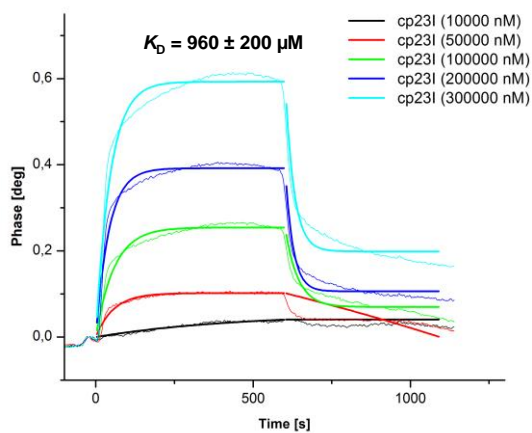
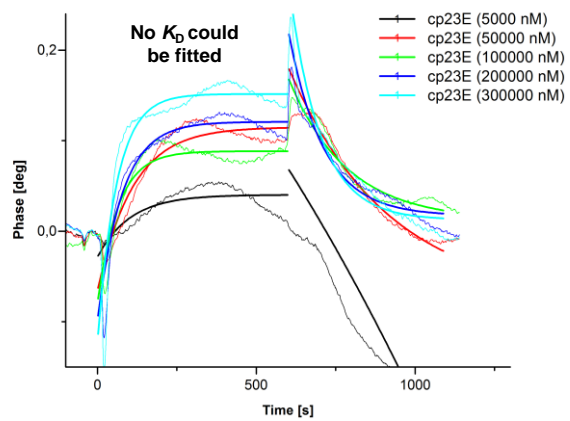
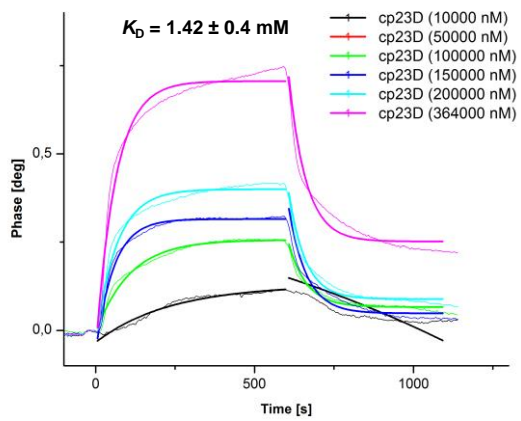
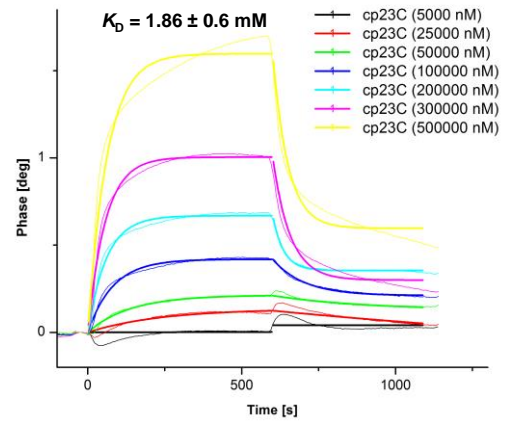
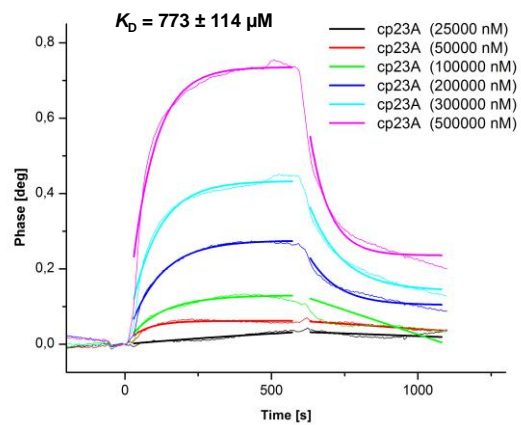
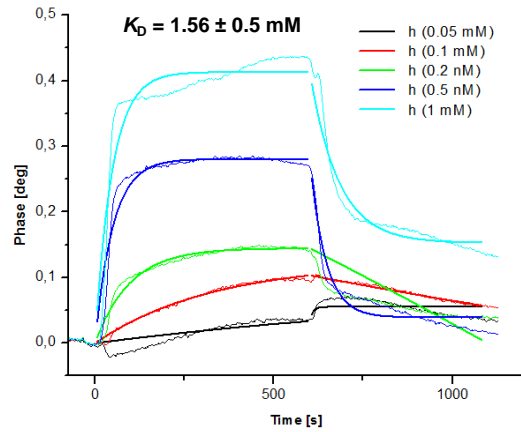
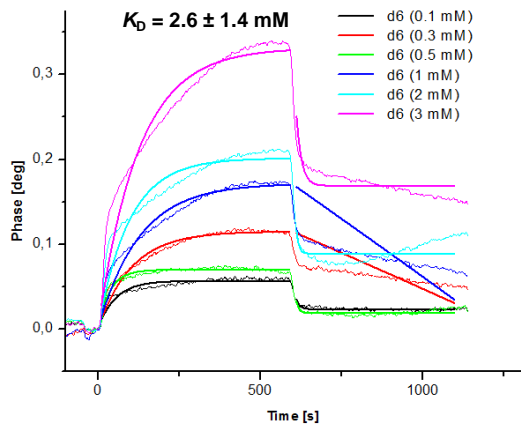
UPLC-MS of the recombinantly expressed EGF



SAW sensorgrams

The SAW sensorgrams not contained in the main body of the thesis are shown here:





REFERENCES

1. Venkatesan, K.; Rual, J.-F.; Vazquez, A.; et al. An Empirical Framework for Binary Interactome Mapping. *Nat. Methods* **2009**, *6* (1), 83–90.
2. Rolland, T.; Taşan, M.; Charlotiaux, B.; et al. A Proteome-Scale Map of the Human Interactome Network. *Cell* **2014**, *159* (5), 1212–1226.
3. Wells, J. A.; McClendon, C. L. Reaching for High-Hanging Fruit in Drug Discovery at Protein-Protein Interfaces. *Nature* **2007**, *450* (7172), 1001–1009.
4. Morelli, X.; Bourgeas, R.; Roche, P. Chemical and Structural Lessons from Recent Successes in Protein-Protein Interaction Inhibition (2P2I). *Curr. Opin. Chem. Biol.* **2011**, *15* (4), 475–481.
5. Tse, C.; Shoemaker, A. R.; Adickes, J.; et al. ABT-263: A Potent and Orally Bioavailable Bcl-2 Family Inhibitor. *Cancer Res.* **2008**, *68* (9), 3421–3428.
6. Kim, W.; Bird, G. H.; Neff, T.; et al. Targeted Disruption of the EZH2–EED Complex Inhibits EZH2-Dependent Cancer. *Nat. Chem. Biol.* **2013**, *9* (10), 643–650.
7. Murugan, R. N.; Park, J. E.; Lim, D.; et al. Development of Cyclic Peptomer Inhibitors Targeting the Polo-Box Domain of Polo-like Kinase 1. *Bioorganic Med. Chem.* **2013**, *21* (9), 2623–2634.
8. Nevola, L.; Giralt, E. Modulating Protein-Protein Interactions: The Potential of Peptides. *Chem. Commun. (Camb)*. **2015**, *51* (16), 3302–3315.
9. Bogan, A. A.; Thorn, K. S. Anatomy of Hot Spots in Protein Interfaces. *J. Mol. Biol.* **1998**, *280* (1), 1–9.
10. Guney, E.; Tuncbag, N.; Keskin, O.; et al. HotSprint: Database of Computational Hot Spots in Protein Interfaces. *Nucleic Acids Res.* **2008**, *36* (Database issue), D662–6.
11. Jones, S.; Thornton, J. M. Principles of Protein-Protein Interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (1), 13–20.
12. Scott, D. E.; Bayly, A. R.; Abell, C.; et al. Small Molecules, Big Targets: Drug Discovery Faces the Protein–protein Interaction Challenge. *Nat. Rev. Drug Discov.* **2016**, *15* (8), 533–550.
13. Hill, T. A.; Shepherd, N. E.; Diness, F.; et al. Constraining Cyclic Peptides To Mimic Protein Structure Motifs. *Angew. Chemie Int. Ed.* **2014**, *53* (48), 13020–13041.
14. Bautista, A. D.; Appelbaum, J. S.; Craig, C. J.; et al. Bridged 3-Peptide Inhibitors of p53-hDM2 Complexation: Correlation between Affinity and Cell Permeability. *J. Am. Chem. Soc.* **2010**, *132* (9), 2904–2906.

15. Wójcik, P.; Berlicki, Ł. Peptide-Based Inhibitors of Protein–protein Interactions. *Bioorg. Med. Chem. Lett.* **2016**, *26* (3), 707–713.
16. Villar, E. A.; Beglov, D.; Chennamadhavuni, S.; et al. How Proteins Bind Macrocycles. *Nat. Chem. Biol.* **2014**, *10* (9), 723–731.
17. London, N.; Raveh, B.; Schueler-Furman, O. Druggable Protein-Protein Interactions - from Hot Spots to Hot Segments. *Curr. Opin. Chem. Biol.* **2013**, *17* (6), 952–959.
18. Jochim, A. L.; Arora, P. S. Systematic Analysis of Helical Protein Interfaces Reveals Targets for Synthetic Inhibitors. *ACS Chem. Biol.* **2010**, *5* (10), 919–923.
19. Cardote, T. A. F.; Ciulli, A. Cyclic and Macrocyclic Peptides as Chemical Tools To Recognise Protein Surfaces and Probe Protein-Protein Interactions. *ChemMedChem* **2016**, *11* (8), 787–794.
20. Check, C. F.; Verma, C. S.; Baselga, J.; et al. Translating p53 into the Clinic. *Nat. Rev. Clin. Oncol.* **2011**, *8* (1), 25–37.
21. Burgess, A.; Chia, K. M.; Haupt, S.; et al. Clinical Overview of MDM2/X-Targeted Therapies. *Front. Oncol.* **2016**, *6*.
22. Hu, B.; Gilkes, D. M.; Chen, J. Efficient p53 Activation and Apoptosis by Simultaneous Disruption of Binding to MDM2 and MDMX. *Cancer Res.* **2007**, *67* (18), 8810–8817.
23. Chang, Y. S.; Graves, B.; Guerlavais, V.; et al. Stapled α -Helical Peptide Drug Development: A Potent Dual Inhibitor of MDM2 and MDMX for p53-Dependent Cancer Therapy. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (36), E3445-54.
24. Checco, J. W.; Gellman, S. H. Iterative Non-Proteinogenic Residue Incorporation Yields A/ β -Peptides with a Helix-Loop-Helix Tertiary Structure and High Affinity for VEGF. *ChemBioChem* **2016**.
25. Stricher, F.; Martin, L.; Vita, C. Design of Miniproteins by the Transfer of Active Sites Onto Small-Size Scaffolds. In *Protein Design*; Humana Press: New Jersey, 2006; Vol. 340, pp 113–150.
26. Olsson, A.-K.; Dimberg, A.; Kreuger, J.; et al. VEGF Receptor Signalling — in Control of Vascular Function. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (5), 359–371.
27. Fedorova, A.; Zobel, K.; Gill, H. S.; et al. The Development of Peptide-Based Tools for the Analysis of Angiogenesis. *Chem. Biol.* **2011**, *18* (7), 839–845.
28. Checco, J. W.; Kreitler, D. F.; Thomas, N. C.; et al. Targeting Diverse Protein-Protein Interaction Interfaces with A/ β -Peptides Derived from the Z-Domain Scaffold. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (15),

4552–4557.

29. Cohen, S. Isolation of a Mouse Submaxillary Gland Protein Accelerating Incisor Eruption and Eyelid Opening in the New-Born Animal. *J. Biol. Chem.* **1962**, *237* (5), 1555–1562.
30. Savage, C. R.; Hash, J. H.; Cohen, S. Epidermal Growth Factor: Location of Disulfide Bonds. *J. Biol. Chem.* **1973**, *248* (22), 7669–7672.
31. Carpenter, G.; King, L.; Cohen, S. Epidermal Growth Factor Stimulates Phosphorylation in Membrane Preparations in Vitro. *Nature* **1978**, *276* (5686), 409–410.
32. Downward, J.; Yarden, Y.; Mayes, E.; et al. Close Similarity of Epidermal Growth Factor Receptor and v-Erb-B Oncogene Protein Sequences. *Nature* *307* (5951), 521–527.
33. Wong, A. J.; Bigner, S. H.; Bigner, D. D.; et al. Increased Expression of the Epidermal Growth Factor Receptor Gene in Malignant Gliomas Is Invariably Associated with Gene Amplification. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (19), 6899–6903.
34. Troiani, T.; Napolitano, S.; Della Corte, C. M.; et al. Therapeutic Value of EGFR Inhibition in CRC and NSCLC: 15 Years of Clinical Evidence. *ESMO Open* **2016**, *1* (5).
35. Sibilias, M.; Steinbach, J. P.; Stingl, L.; et al. A Strain-Independent Postnatal Neurodegeneration in Mice Lacking the EGF Receptor. *EMBO J.* **1998**, *17* (3), 719–731.
36. Arteaga, C. L.; Engelman, J. A. ERBB Receptors: From Oncogene Discovery to Basic Science to Mechanism-Based Cancer Therapeutics. *Cancer Cell* **2014**, *25* (3), 282–303.
37. Vecchione, L.; Jacobs, B.; Normanno, N.; et al. EGFR-Targeted Therapy. *Exp. Cell Res.* **2011**, *317* (19), 2765–2771.
38. Chong, C. R.; Jänne, P. A. The Quest to Overcome Resistance to EGFR-Targeted Therapies in Cancer. *Nat. Med.* **2013**, *19* (11), 1389–1400.
39. Arteaga, C. Targeting HER1/EGFR: A Molecular Approach to Cancer Therapy. *Semin. Oncol.* **2003**, *30* (3 Suppl 7), 3–14.
40. Yarden, Y.; Sliwkowski, M. X. Untangling the ErbB Signalling Network. *Nat. Rev. Mol. Cell Biol.* **2001**, *2* (2), 127–137.
41. Cho, H.-S.; Mason, K.; Ramyar, K. X.; et al. Structure of the Extracellular Region of HER2 Alone and in Complex with the Herceptin Fab. *Nature* **2003**, *421* (6924), 756–760.
42. Tebbutt, N.; Pedersen, M. W.; Johns, T. G. Targeting the ERBB

- Family in Cancer: Couples Therapy. *Nat. Rev. Cancer* **2013**, *13* (9), 663–673.
43. Lemmon, M. A. Ligand-Induced ErbB Receptor Dimerization. *Exp. Cell Res.* **2009**, *315* (4), 638–648.
 44. Wieduwilt, M. J.; Moasser, M. M. The Epidermal Growth Factor Receptor Family: Biology Driving Targeted Therapeutics. *Cell. Mol. Life Sci.* **2008**, *65* (10), 1566–1584.
 45. Harris, R. C.; Chung, E.; Coffey, R. J. EGF Receptor Ligands. *Exp. Cell Res.* **2003**, *284* (1), 2–13.
 46. Jones, J. T.; Akita, R. W.; Sliwkowski, M. X. Binding Specificities and Affinities of Egf Domains for ErbB Receptors. *FEBS Lett.* **1999**, *447* (2–3), 227–231.
 47. Van Zoelen, E. J.; Stortelers, C.; Lenferink, a E.; et al. The EGF Domain: Requirements for Binding to Receptors of the ErbB Family. *Vitam. Horm.* **2000**, *59*, 99–131.
 48. Jones, J. T.; Akita, R. W.; Sliwkowski, M. X. Binding Specificities and Affinities of Egf Domains for ErbB Receptors. *FEBS Lett.* **1999**, *447* (2–3), 227–231.
 49. Yewale, C.; Baradia, D.; Vhora, I.; et al. Epidermal Growth Factor Receptor Targeting in Cancer: A Review of Trends and Strategies. *Biomaterials* **2013**, *34* (34), 8690–8707.
 50. Lemmon, M. A.; Bu, Z.; Ladbury, J. E.; et al. Two EGF Molecules Contribute Additively to Stabilization of the EGFR Dimer. *EMBO J.* **1997**, *16* (2), 281–294.
 51. Ferguson, K. M.; Berger, M. B.; Mendrola, J. M.; et al. EGF Activates Its Receptor by Removing Interactions That Autoinhibit Ectodomain Dimerization. *Mol. Cell* **2003**, *11* (2), 507–517.
 52. Ogiso, H.; Ishitani, R.; Nureki, O.; et al. Crystal Structure of the Complex of Human Epidermal Growth Factor and Receptor Extracellular Domains. *Cell* **2002**, *110* (6), 775–787.
 53. Kovacs, E.; Zorn, J. A.; Huang, Y.; et al. A Structural Perspective on the Regulation of the Epidermal Growth Factor Receptor. *Annu. Rev. Biochem.* **2015**, *84* (1), 739–764.
 54. Kaplan, M.; Narasimhan, S.; de Heus, C.; et al. EGFR Dynamics Change during Activation in Native Membranes as Revealed by NMR. *Cell* **2016**, *167* (5), 1241–1251.e11.
 55. Dawson, J. P.; Bu, Z.; Lemmon, M. A. Ligand-Induced Structural Transitions in ErbB Receptor Extracellular Domains. *Structure* **2007**, *15* (8), 942–954.

56. Mi, L.-Z.; Lu, C.; Li, Z.; et al. Simultaneous Visualization of the Extracellular and Cytoplasmic Domains of the Epidermal Growth Factor Receptor. *Nat. Struct. Mol. Biol.* **2011**, *18* (9), 984–989.
57. Arkhipov, A.; Shan, Y.; Das, R.; et al. Architecture and Membrane Interactions of the EGF Receptor. *Cell* **2013**, *152* (3), 557–569.
58. Yarden, Y.; Schlessinger, J. Self-Phosphorylation of Epidermal Growth Factor Receptor: Evidence for a Model of Intermolecular Allosteric Activation. *Biochemistry* **1987**, *26* (5), 1434–1442.
59. Moriki, T.; Maruyama, H.; Maruyama, I. N. Activation of Preformed EGF Receptor Dimers by Ligand-Induced Rotation of the Transmembrane Domain. *J. Mol. Biol.* **2001**, *311* (5), 1011–1026.
60. King, A. C.; Cuatrecasas, P. Resolution of High and Low Affinity Epidermal Growth Factor Receptors. Inhibition of High Affinity Component by Low Temperature, Cycloheximide, and Phorbol Esters. *J. Biol. Chem.* **1982**, *257* (6), 3053–3060.
61. Wood, E. R. A Unique Structure for Epidermal Growth Factor Receptor Bound to GW572016 (Lapatinib): Relationships among Protein Conformation, Inhibitor Off-Rate, and Receptor Activity in Tumor Cells. *Cancer Res.* **2004**, *64* (18), 6652–6659.
62. Tomas, A.; Futter, C. E.; Eden, E. R. EGF Receptor Trafficking: Consequences for Signaling and Cancer. *Trends Cell Biol.* **2014**, *24* (1), 26–34.
63. Citri, A.; Yarden, Y. EGF–ERBB Signalling: Towards the Systems Level. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (7), 505–516.
64. Lemmon, M. A.; Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **2010**, *141* (7), 1117–1134.
65. Kitano, H. Biological Robustness. *Nat. Rev. Genet.* **2004**, *5* (11), 826–837.
66. Kim, G. P.; Grothey, A. Targeting Colorectal Cancer with Human Anti-EGFR Monoclonal Antibodies: Focus on Panitumumab. *Biologics* **2008**, *2* (2), 223–228.
67. Montagut, C.; Dalmases, A.; Bellosillo, B.; et al. Identification of a Mutation in the Extracellular Domain of the Epidermal Growth Factor Receptor Conferring Cetuximab Resistance in Colorectal Cancer. *Nat. Med.* **2012**, *18* (2), 221–223.
68. Wakeling, A. E.; Barker, A. J.; Davies, D. H.; et al. Specific Inhibition of Epidermal Growth Factor Receptor Tyrosine Kinase by 4-Anilinoquinazolines. *Breast Cancer Res. Treat.* **1996**, *38* (1), 67–73.
69. Huang, H.-W.; Mohan, S. K.; Yu, C. The NMR Solution Structure of

- Human Epidermal Growth Factor (hEGF) at Physiological pH and Its Interactions with Suramin. *Biochem. Biophys. Res. Commun.* **2010**, *402* (4), 705–710.
70. Guardiola, S.; Díaz-Lobo, M.; Seco, J.; et al. Peptides Targeting EGF Block the EGF-EGFR Interaction. *ChemBioChem* **2016**, *17* (8), 702–711.
 71. Rodriguez, P. C.; Popa, X.; Martínez, O.; et al. A Phase III Clinical Trial of the Epidermal Growth Factor Vaccine CIMAvax-EGF as Switch Maintenance Therapy in Advanced Non-Small Cell Lung Cancer Patients. *Clin. Cancer Res.* **2016**, *22* (15).
 72. Lynch, T. J.; Bell, D. W.; Sordella, R.; et al. Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non-Small-Cell Lung Cancer to Gefitinib. *N. Engl. J. Med.* **2004**, *350* (21), 2129–2139.
 73. Jänne, P. A.; Yang, J. C.-H.; Kim, D.-W.; et al. AZD9291 in EGFR Inhibitor-Resistant Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **2015**, *372* (18), 1689–1699.
 74. Padfield, E.; Ellis, H. P.; Kurian, K. M. Current Therapeutic Advances Targeting EGFR and EGFRvIII in Glioblastoma. *Front. Oncol.* **2015**, *5*, 5.
 75. Ohashi, K.; Sequist, L. V.; Arcila, M. E.; et al. Lung Cancers with Acquired Resistance to EGFR Inhibitors Occasionally Harbor BRAF Gene Mutations but Lack Mutations in KRAS, NRAS, or MEK1. *Proc. Natl. Acad. Sci.* **2012**, *109* (31), E2127–E2133.
 76. Misale, S.; Yaeger, R.; Hobor, S.; et al. Emergence of KRAS Mutations and Acquired Resistance to Anti-EGFR Therapy in Colorectal Cancer. *Nature* **2012**, *486* (7404), 532.
 77. Zhao, B.; Wang, L.; Qiu, H.; et al. Mechanisms of Resistance to Anti-EGFR Therapy in Colorectal Cancer. **2016**, 19–21.
 78. Ng, K. P.; Hillmer, A. M.; Chuah, C. T. H.; et al. A Common BIM Deletion Polymorphism Mediates Intrinsic Resistance and Inferior Responses to Tyrosine Kinase Inhibitors in Cancer. *Nat. Med.* **2012**, *18* (4), 521–528.
 79. Xu, W.; Bi, Y.; Kong, J.; et al. Combination of an Anti-EGFRvIII Antibody CH12 with Rapamycin Synergistically Inhibits the Growth of EGFRvIII+PTEN-Glioblastoma in Vivo. *Oncotarget* **2016**, *7* (17), 24752–24765.
 80. Bock, C.; Lengauer, T. Managing Drug Resistance in Cancer: Lessons from HIV Therapy. *Nat. Rev. Cancer* **2012**, *12* (7), 494–501.
 81. Weickhardt, A. J.; Price, T. J.; Chong, G.; et al. Dual Targeting of the

- Epidermal Growth Factor Receptor Using the Combination of Cetuximab and Erlotinib: Preclinical Evaluation and Results of the Phase II DUX Study in Chemotherapy-Refractory, Advanced Colorectal Cancer. *J. Clin. Oncol.* **2012**, *30* (13), 1505–1512.
82. Miles, J. A.; Yeo, D. J.; Rowell, P.; et al. Hydrocarbon Constrained Peptides –Understanding Preorganisation and Binding Affinity. *Chem. Sci.* **2016**, *7* (6), 3694–3702.
 83. Singh, J.; Petter, R. C.; Baillie, T. A.; et al. The Resurgence of Covalent Drugs. *Nat. Rev. Drug Discov.* **2011**, *10* (4), 307–317.
 84. Bauer, R. A. Covalent Inhibitors in Drug Discovery: From Accidental Discoveries to Avoided Liabilities and Designed Therapies. *Drug Discov. Today* **2015**, *20* (9), 1061–1073.
 85. Huhn, A. J.; Guerra, R. M.; Harvey, E. P.; et al. Selective Covalent Targeting of Anti-Apoptotic BFL-1 by Cysteine-Reactive Stapled Peptide Inhibitors. *Cell Chem. Biol.* **2016**, *23* (9), 1123–1134.
 86. Donmienne L.; Giovanni A. and Fairlie*, D. P. Protease Inhibitors: Current Status and Future Prospects. **2000**.
 87. Verma, S.; Dixit, R.; Pandey, K. C. Cysteine Proteases: Modes of Activation and Future Prospects as Pharmacological Targets. *Front. Pharmacol.* **2016**, *7*, 107.
 88. Tyndall, J. D. A.; Nall, T. and Fairlie, D. P. Proteases Universally Recognize Beta Strands In Their Active Sites. **2005**.
 89. Juillerat-Jeanneret, L. Prolyl-Specific Peptidases and Their Inhibitors in Biological Processes. *Curr. Chem. Biol.* **2008**, *2* (2), 97–109.
 90. Iván, G.; Szabadka, Z.; Ördög, R.; et al. *Four Spatial Points That Define Enzyme Families*; 2009; Vol. 383.
 91. Irazusta, J.; Larrinaga, G.; González-Maeso, J.; et al. Distribution of Prolyl Endopeptidase Activities in Rat and Human Brain. *Neurochem. Int.* **2002**, *40* (4), 337–345.
 92. Walter, R.; Shlank, H.; Glass, J. D.; et al. Leucylglycinamide Released from Oxytocin by Human Uterine Enzyme. *Science* **1971**, *173* (3999), 827–829.
 93. García-Horsman, J. A.; Männistö, P. T.; Venäläinen, J. I. On the Role of Prolyl Oligopeptidase in Health and Disease. *Neuropeptides* **2007**, *41* (1), 1–24.
 94. Toide, K.; Shinoda, M.; Fujiwara, T.; et al. Effect of a Novel Prolyl Endopeptidase Inhibitor, JTP-4819, on Spatial Memory and Central Cholinergic Neurons in Aged Rats. *Pharmacol. Biochem. Behav.* **1997**, *56* (3), 427–434.

95. López, A.; Mendieta, L.; Prades, R.; et al. Peptide POP Inhibitors for the Treatment of the Cognitive Symptoms of Schizophrenia. *Future Med. Chem.* **2013**, *5* (13), 1509–1523.
96. Schulz, I.; Gerhartz, B.; Neubauer, A.; et al. Modulation of Inositol 1,4,5-Triphosphate Concentration by Prolyl Endopeptidase Inhibition. *Eur. J. Biochem.* **2002**, *269* (23), 5813–5820.
97. Jalkanen, A. J.; Puttonen, K. A.; Venäläinen, J. I.; et al. Beneficial Effect of Prolyl Oligopeptidase Inhibition on Spatial Memory in Young but Not in Old Scopolamine-Treated Rats. *Basic Clin. Pharmacol. Toxicol.* **2006**, *0* (0), 061214140717002.
98. Di Daniel, E.; Glover, C. P.; Grot, E.; et al. Prolyl Oligopeptidase Binds to GAP-43 and Functions without Its Peptidase Activity. *Mol. Cell. Neurosci.* **2009**, *41* (3), 373–382.
99. Savolainen, M. H.; Yan, X.; Myöhänen, T. T.; et al. Prolyl Oligopeptidase Enhances α -Synuclein Dimerization via Direct Protein-Protein Interaction. *J. Biol. Chem.* **2015**, *290* (8), 5117–5126.
100. Hannula, M. J.; Myöhänen, T. T.; Tenorio-Laranga, J.; et al. Prolyl Oligopeptidase Colocalizes with α -Synuclein, β -Amyloid, Tau Protein and Astroglia in the Post-Mortem Brain Samples with Parkinson's and Alzheimer's Diseases. *Neuroscience* **2013**, *242*, 140–150.
101. López, A.; Herranz-Trillo, F.; Kotev, M.; et al. Active-Site-Directed Inhibitors of Prolyl Oligopeptidase Abolish Its Conformational Dynamics. *ChemBioChem* **2016**, *17* (10), 913–917.
102. López, A.; Herranz-Trillo, F.; Kotev, M.; et al. Active-Site-Directed Inhibitors of Prolyl Oligopeptidase Abolish Its Conformational Dynamics. *ChemBioChem* **2016**, *17* (10), 913–917.
103. Myöhänen, T.; Hannula, M.; Van Elzen, R.; et al. A Prolyl Oligopeptidase Inhibitor, KYP-2047, Reduces α -Synuclein Protein Levels and Aggregates in Cellular and Animal Models of Parkinson's Disease. *Br. J. Pharmacol.* **2012**, *166* (3), 1097–1113.
104. Schneider, J.; Giardiniere, M.; Morain, P. Effects of the Prolyl Endopeptidase Inhibitor S 17092 on Cognitive Deficits in Chronic Low Dose MPTP-Treated Monkeys. *Neuropsychopharmacology* **2002**, *26* (2), 176–182.
105. Svarcbahts, R.; Julku, U. H.; Myöhänen, T. T. Inhibition of Prolyl Oligopeptidase Restores Spontaneous Motor Behavior in the α -Synuclein Virus Vector-Based Parkinson's Disease Mouse Model by Decreasing α -Synuclein Oligomeric Species in Mouse Brain. *J. Neurosci.* **2016**, *36* (49).
106. Bal, G.; Van der Veken, P.; Antonov, D.; et al. Prolylisoxazoles: Potent

- Inhibitors of Prolyloligopeptidase with Antitrypanosomal Activity. *Bioorg. Med. Chem. Lett.* **2003**, *13* (17), 2875–2878.
107. Vendeville, S.; Goossens, F.; Debreu-Fontaine, M. A.; et al. Comparison of the Inhibition of Human and Trypanosoma Cruzi Prolyl Endopeptidases. *Bioorganic Med. Chem.* **2002**, *10* (6), 1719–1729.
 108. Tadaki, D. K.; Niyogi, S. K. The Functional Importance of Hydrophobicity of the Tyrosine at Position 13 of Human Epidermal Growth Factor in Receptor Binding. *J. Biol. Chem.* **1993**, *268* (14), 10114–10119.
 109. Engler, D. A.; Champion, S. R.; Hauser, M. R.; et al. Critical Functional Requirement for the Guanidinium Group of the Arginine 41 Side Chain of Human Epidermal Growth Factor as Revealed by Mutagenic Inactivation and Chemical Reactivation. *J. Biol. Chem.* **1992**, *267* (4), 2274–2281.
 110. Dudgeon, T. J.; Cooke, R. M.; Baron, M.; et al. Structure-Function Analysis of Epidermal Growth Factor: Site Directed Mutagenesis and Nuclear Magnetic Resonance. *FEBS Lett* **1990**, *261* (2), 392–396.
 111. Murray, M. B.; Tadaki, D. K.; Champion, S. R.; et al. Structure-Function Analysis of a Conserved Aromatic Cluster in the N-Terminal Domain of Human Epidermal Growth Factor. *Protein Eng.* **1998**, *11* (11), 1041–1050.
 112. Ferguson, K. M.; Berger, M. B.; Mendrola, J. M.; et al. EGF Activates Its Receptor by Removing Interactions That Autoinhibit Ectodomain Dimerization. *Mol. Cell* **2003**, *11* (2), 507–517.
 113. Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **2009**, *49* (2), 377–389.
 114. Eugenio Vazquez, M.; Rothman, D. M.; Imperiali, B. A New Environment-Sensitive Fluorescent Amino Acid for Fmoc-Based Solid Phase Peptide Synthesis. *Org. Biomol. Chem.* **2004**, *2* (14), 1965–1966.
 115. Sainlos, M.; Imperiali, B. Tools for Investigating Peptide-Protein Interactions: Peptide Incorporation of Environment-Sensitive Fluorophores through SPPS-Based “Building Block” Approach. *Nat. Protoc.* **2007**, *2* (12), 3210–3218.
 116. Du, Y. Fluorescence Polarization Assay to Quantify Protein-Protein Interactions in an HTS Format. In *Methods in molecular biology (Clifton, N.J.)*; 2015; Vol. 1278, pp 529–544.
 117. Rosano, G. L.; Ceccarelli, E. A. Recombinant Protein Expression in Escherichia Coli: Advances and Challenges. *Front. Microbiol.* **2014**, *5*, 172.
 118. LaVallie, E. R.; DiBlasio, E. A.; Kovacic, S.; et al. A Thioredoxin Gene

- Fusion Expression System That Circumvents Inclusion Body Formation in the E. Coli Cytoplasm. *Biotechnology. (N. Y.)* **1993**, *11* (2), 187–193.
119. Butt, T. R.; Edavettal, S. C.; Hall, J. P.; et al. SUMO Fusion Technology for Difficult-to-Express Proteins. *Protein Expr. Purif.* **2005**, *43* (1), 1–9.
 120. Díaz-Lobo, M. Applications of Surface Acoustic Wave Sensor in Protein Science. In *Encyclopedia of Analytical Chemistry*; John Wiley & Sons, Ltd: Chichester, UK, 2016; pp 1–14.
 121. Domagala, T.; Konstantopoulos, N.; Smyth, F.; et al. Stoichiometry, Kinetic and Binding Analysis of the Interaction between Epidermal Growth Factor (EGF) and the Extracellular Domain of the EGF Receptor. *Growth Factors* **2000**, *18* (1), 11–29.
 122. Eyrisch, S.; Helms, V. Transient Pockets on Protein Surfaces Involved in Protein–Protein Interaction. *J. Med. Chem.* **2007**, *50* (15), 3457–3464.
 123. Sainlos, M.; Iskenderian, W. S.; Imperiali, B. A General Screening Strategy for Peptide-Based Fluorogenic Ligands: Probes for Dynamic Studies of PDZ Domain-Mediated Interactions. *J. Am. Chem. Soc.* **2009**, *131* (19), 6680–6682.
 124. Sainlos, M.; Tigaret, C.; Poujol, C.; et al. Biomimetic Divalent Ligands for the Acute Disruption of Synaptic AMPAR Stabilization. *Nat. Chem. Biol.* **2011**, *7* (2), 81–91.
 125. Turnbull D. and Daranas W. B. T. On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *J. Am. Chem. Soc.* **2003**, *125* (48), 14859–14866.
 126. Loeffler, H. H.; Winn, M. D. Ligand Binding and Dynamics of the Monomeric Epidermal Growth Factor Receptor Ectodomain. *Proteins Struct. Funct. Bioinforma.* **2013**, *81* (11), 1931–1943.
 127. Watts, K. S.; Dalal, P.; Tebben, A. J.; et al. Macrocycle Conformational Sampling with MacroModel. *J. Chem. Inf. Model.* **2014**, *54* (10), 2680–2696.
 128. Miller, S. C.; Scanlan, T. S. Site-Selective N-Methylation of Peptides on Solid Support. *J. Am. Chem. Soc.* **1997**, *119* (9), 2301–2302.
 129. El-Faham, A.; Albericio, F. COMU: A Third Generation of Uronium-Type Coupling Reagents. *J. Pept. Sci.* **2010**, *16* (1), 6–9.
 130. Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
 131. Cheng, P. N.; Pham, J. D.; Nowick, J. S. The Supramolecular

- Chemistry of β -Sheets. *J. Am. Chem. Soc.* **2013**, *135* (15), 5477–5492.
132. Freire, F.; Gellman, S. H. Macrocyclic Design Strategies for Small, Stable Parallel β -Sheet Scaffolds. *J. Am. Chem. Soc.* **2009**, *131* (23), 7970–7972.
 133. Freire, F.; Almeida, A. M.; Fisk, J. D.; et al. Impact of Strand Length on the Stability of Parallel β -Sheet Secondary Structure. *Angew. Chemie - Int. Ed.* **2011**, *50* (37), 8735–8738.
 134. Fooks, H. M.; Martin, A. C. R.; Woolfson, D. N.; et al. Amino Acid Pairing Preferences in Parallel β -Sheets in Proteins. *J. Mol. Biol.* **2006**, *356* (1), 32–44.
 135. Kenner, G. W.; McDermott, J. R.; Sheppard, R. C. The Safety Catch Principle in Solid Phase Peptide Synthesis. *J. Chem. Soc. D Chem. Commun.* **1971**, No. 12, 636.
 136. de Visser, P. C.; Kriek, N. M. a J.; van Hooft, P. a V; et al. Solid-Phase Synthesis of Polymyxin B1 and Analogues via a Safety-Catch Approach. *J. Pept. Res.* **2003**, *61* (6), 298–306.
 137. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; et al. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266* (5186), 776–779.
 138. Clark, R. J.; Craik, D. J. Native Chemical Ligation Applied to the Synthesis and Bioengineering of Circular Peptides and Proteins. *Biopolymers* **2010**, *94* (4), 414–422.
 139. Blanco-Canosa, J. B.; Dawson, P. E. An Efficient Fmoc-SPPS Approach for the Generation of Thioester Peptide Precursors for Use in Native Chemical Ligation. *Angew. Chem. Int. Ed. Engl.* **2008**, *47* (36), 6851–6855.
 140. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015**, *10* (5), 449–461.
 141. Kerrigan, J. E. Molecular Dynamics Simulations in Drug Design. In *Methods in molecular biology (Clifton, N.J.)*; 2013; Vol. 993, pp 95–113.
 142. Chen, J.; Brooks, C. L.; Khandogin, J.; et al. Recent Advances in Implicit Solvent-Based Methods for Biomolecular Simulations. *Curr. Opin. Struct. Biol.* **2008**, *18* (2), 140–148.
 143. Hou, T.; Wang, J.; Li, Y.; et al. Assessing the Performance of the MM / PBSA and MM / GBSA Methods . I . The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model* **2010**, *51* (1), 69–82.

144. Eglen, R. M.; Reisine, T.; Roby, P.; et al. The Use of AlphaScreen Technology in HTS: Current Status. *Curr. Chem. Genomics* **2008**, *1*, 2–10.
145. Zang, R.; Li, D.; Tang, I.-C.; et al. Cell-Based Assays in High-Throughput Screening for Drug Discovery. *Int. J. Biotechnol. Wellness Ind.* **2012**, *1*, 31–51.
146. Kawamoto, T.; Mendelsohn, J.; Le, a; et al. Relation of Epidermal Growth Factor Receptor Concentration to Growth of Human Epidermoid Carcinoma A431 Cells. *J. Biol. Chem.* **1984**, *259* (12), 7761–7766.
147. Reilly, R. M.; Gariépy, J. Factors Influencing the Sensitivity of Tumor Imaging with a Receptor-Binding Radiopharmaceutical. *J. Nucl. Med.* **1998**, *39* (6), 1036–1043.
148. Herzenberg, L. A.; Parks, D.; Sahaf, B.; et al. The History and Future of the Fluorescence Activated Cell Sorter and Flow Cytometry: A View from Stanford. *Clin. Chem.* **2002**, *48* (10), 1819–1827.
149. Bidwell, G. L. Peptides for Cancer Therapy: A Drug-Development Opportunity and a Drug-Delivery Challenge. *Ther. Deliv.* **2012**, *3* (5), 609–621.
150. Brunel, F. M.; Dawson, P. E. Synthesis of Constrained Helical Peptides by Thioether Ligation: Application to Analogs of gp41. *Chem. Commun.* **2005**, No. 20, 2552.
151. Song, J.; Tan, H.; Perry, A. J.; et al. PROSPER: An Integrated Feature-Based Tool for Predicting Protease Substrate Cleavage Sites. *PLoS One* **2012**, *7* (11).
152. Hatahet, F.; Nguyen, V. D.; Salo, K. E. H.; et al. Disruption of Reducing Pathways Is Not Essential for Efficient Disulfide Bond Formation in the Cytoplasm of E. Coli. *Microb. Cell Fact.* **2010**, *9*, 67.
153. Di, L. Strategic Approaches to Optimizing Peptide ADME Properties. *AAPS J.* **2015**, *17* (1), 134–143.
154. Pollaro, L.; Heinis, C. Strategies to Prolong the Plasma Residence Time of Peptide Drugs. *Medchemcomm* **2010**, *1* (5), 319.
155. Sheweita, S. A. Drug-Metabolizing Enzymes: Mechanisms and Functions. *Curr. Drug Metab.* **2000**, *1* (2), 107–132.
156. Coe, K. J.; Koudriakova, T. Metabolic Stability Assessed by Liver Microsomes and Hepatocytes. *Methods Pharmacol. Toxicol.* **2014**, No. 1, 87–99.
157. Riley, R. J.; McGinnity, D. F.; Austin, R. P. A Unified Model for Predicting Human Hepatic, Metabolic Clearance from in Vitro

- Intrinsic Clearance Data in Hepatocytes and Microsomes. *Drug Metab. Dispos.* **2005**, *33* (9).
158. Wishart, D. S.; Bigam, C. G.; Holm, A.; et al. ¹H, ¹³C and ¹⁵N Random Coil NMR Chemical Shifts of the Common Amino Acids. I. Investigations of Nearest-Neighbor Effects. *J. Biomol. NMR* **1995**, *5* (1), 67–81.
 159. Schechter, I.; Berger, A. On the Size of the Active Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27* (2), 157–162.
 160. Lawandi, J.; Gerber-Lemaire, S.; Juillerat-Jeanneret, L.; et al. Inhibitors of Prolyl Oligopeptidases for the Therapy of Human Diseases: Defining Diseases and Inhibitors. *J. Med. Chem.* **2010**, *53* (9), 3423–3438.
 161. Adolpho, L.; Marin, D.; Puigpinos, A.; et al. In Vitro Evaluation of Caffeoyle and Cinnamoyl Derivatives as Potential Prolyl Oligopeptidase Inhibitors. *Planta Med.* **2013**, *79* (16), 1531–1535.
 162. Prades, R.; Munarriz-Cuezva, E.; Urigüen, L.; et al. The Prolyl Oligopeptidase Inhibitor IPR19 Ameliorates Cognitive Deficits in Mouse Models of Schizophrenia. *Eur. Neuropsychopharmacol.* **2017**, *27* (2), 180–191.
 163. Fülöp, V.; Böcskei, Z.; Polgár, L. Prolyl Oligopeptidase: An Unusual Beta-Propeller Domain Regulates Proteolysis. *Cell* **1998**, *94* (2), 161–170.
 164. Venäläinen, J. I.; Garcia-Horsman, J. A.; Forsberg, M. M.; et al. Binding Kinetics and Duration of in Vivo Action of Novel Prolyl Oligopeptidase Inhibitors. *Biochem. Pharmacol.* **2006**, *71* (5), 683–692.
 165. Tsuda, M.; Muraoka, Y.; Nagai, M.; et al. Poststatin, a New Inhibitor of Prolyl Endopeptidase. VII. N-Cycloalkylamide Analogues. *J. Antibiot. (Tokyo)*. **1996**, *49* (9), 909–920.
 166. Li, J.; Wilk, E.; Wilk, S. Inhibition of Prolyl Oligopeptidase by Fmoc-Aminoacylpyrrolidine-2-Nitriles. *J. Neurochem.* **1996**, *66* (5), 2105–2112.
 167. Dubiella, C.; Cui, H.; Gersch, M.; et al. Selective Inhibition of the Immunoproteasome by Ligand-Induced Crosslinking of the Active Site. *Angew. Chemie - Int. Ed.* **2014**, *53* (44), 11969–11973.
 168. Narayanan, A.; Jones, L. H. Sulfonyl Fluorides as Privileged Warheads in Chemical Biology. *Chem. Sci.* **2015**, *6* (5), 2650–2659.
 169. Brouwer, A. J.; Ceylan, T.; Jonker, A. M.; et al. Synthesis and Biological Evaluation of Novel Irreversible Serine Protease Inhibitors Using Amino Acid Based Sulfonyl Fluorides as an Electrophilic Trap. *Bioorganic Med. Chem.* **2011**, *19* (7), 2397–2406.

170. L'heureux, A.; Beaulieu, F.; Bennett, C.; et al. Aminodifluorosulfonium Salts: Selective Fluorination Reagents with Enhanced Thermal Stability and Ease of Handling. *J. Org. Chem.* **2010**, *75* (10), 3401–3411.
171. Juillerat-Jeanneret, L.; Gerber-Lemaire, S. The Prolyl-Aminodipeptidases and Their Inhibitors as Therapeutic Targets for Fibrogenic Disorders. *Mini Rev. Med. Chem.* **2009**, *9* (2), 215–226.
172. Poplawski, S. E.; Lai, J. H.; Li, Y.; et al. Identification of Selective and Potent Inhibitors of Fibroblast Activation Protein and Prolyl Oligopeptidase. *J. Med. Chem.* **2013**, *56* (9), 3467–3477.
173. Pardridge, W. M. The Blood-Brain Barrier: Bottleneck in Brain Drug Development. *NeuroRx* **2005**, *2* (1), 3–14.
174. Di, L.; Kerns, E. H.; Fan, K.; et al. High Throughput Artificial Membrane Permeability Assay for Blood-Brain Barrier. *Eur. J. Med. Chem.* **2003**, *38* (3), 223–232.
175. Arranz-Gibert, P.; Guixer, B.; Malakoutikhah, M.; et al. Lipid Bilayer Crossing-The Gate of Symmetry. Water-Soluble Phenylproline-Based Blood-Brain Barrier Shuttles. *J. Am. Chem. Soc.* **2015**, *137* (23), 7357–7364.
176. Qvit, N.; Rubin, S. J. S.; Urban, T. J.; et al. Peptidomimetic Therapeutics: Scientific Approaches and Opportunities. *Drug Discov. Today* **2017**, *22* (2), 454–462.
177. Curk, T.; Dobnikar, J.; Frenkel, D. Design Principles for Super Selectivity Using Multivalent Interactions. **2016**, *18* (15), 56–1.
178. Teixeira, A.; Benckhuijsen, W. E.; de Koning, P. E.; et al. The Use of DODT as a Non-Malodorous Scavenger in Fmoc-Based Peptide Synthesis. *Protein Pept. Lett.* **2002**, *9* (5), 379–385.
179. Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; et al. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided. Mol. Des.* **2013**, *27* (3), 221–234.
180. Kollman, P.A.; Massova I.; et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc Chem Res.* **2000**, *33* (12), 889–897.
181. Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98* (7), 1978–1988.
182. Tsui, V.; Case, D. A. Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations. *Biopolymers* **2000**, *56* (4), 275–291.

183. Patriksson, A.; van der Spoel, D. A Temperature Predictor for Parallel Tempering Simulations. *Phys. Chem. Chem. Phys.* **2008**, *10* (15), 2073.
184. Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem.* **1999**, *20* (2), 217–230.
185. Berrow, N. S.; Alderton, D.; Sainsbury, S.; et al. A Versatile Ligation-Independent Cloning Method Suitable for High-Throughput Expression Screening Applications. *Nucleic Acids Res.* **2007**, *35* (6), e45.
186. Slamnoiu, S.; Vlad, C.; Stumbaum, M.; et al. Identification and Affinity-Quantification of SS-Amyloid and α -Synuclein Polypeptides Using On-Line SAW-Biosensor-Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (8), 1472–1481.
187. Dragusanu, M.; Petre, B.-A.; Slamnoiu, S.; et al. On-Line Bioaffinity-Electrospray Mass Spectrometry for Simultaneous Detection, Identification, and Quantification of Protein-Ligand Interactions. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (10), 1643–1648.
188. Hwang, T. L.; Shaka, A. J. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *J. Magn. Reson. Ser. A* **1995**, *112* (2), 275–279.
189. Bax, A.; Davis, D. G. MLEV-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. *J. Magn. Reson.* **1985**, *65* (2), 355–360.
190. Jeener, J.; Meier, B. H.; Bachmann, P.; et al. Investigation of Exchange Processes by Two-dimensional NMR Spectroscopy. *J. Chem. Phys.* **1979**, *71* (11), 4546–4553.
191. Filmus, J.; Pollack, M. N.; Cailleau, R.; et al. MDA-MB-468 a Human Breast Cancer Cell Line with a High Number of Epidermal Growth Factor (EGF) Receptors Has an Amplified EGFR Receptor Gene and Is Growth Inhibited by EGF. *Biochem Biophys Res Comm* **1985**, *128* (2), 898–905.
192. Tarragó, T.; Frutos, S.; Rodriguez-Mias, R. A.; et al. Identification by ^{19}F NMR of Traditional Chinese Medicinal Plants Possessing Prolyl Oligopeptidase Inhibitory Activity. *ChemBioChem* **2006**, *7* (5), 827–833.
193. Toide, K.; Iwamoto, Y.; Fujiwara, T.; et al. JTP-4819: A Novel Prolyl Endopeptidase Inhibitor with Potential as a Cognitive Enhancer. *J. Pharmacol. Exp. Ther.* **1995**, *274* (3).
194. Checler, F.; Vincent, J. P.; Kitabgi, P. Inactivation of Neurotensin by Rat Brain Synaptic Membranes Partly Occurs Through Cleavage at

the Arg8-Arg9 Peptide Bond by a Metalloendopeptidase. *J. Neurochem.* **1985**, *45* (5), 1509–1513.

**SUMMARY IN
CATALAN**


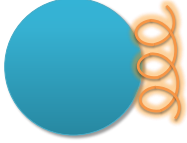

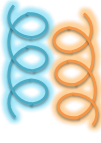

Modulació d'interaccions proteïna-proteïna amb pèptids

Tradicionalment, la indústria farmacèutica, així com la majoria de grups de recerca acadèmica, han focalitzat la recerca de nous fàrmacs en la inhibició d'enzims i receptors implicats en processos patològics. En aquestes dianes terapèutiques, el lloc d'unió del fàrmac és una cavitat típicament hidrofòbica i de mida reduïda, que pot ser modulada d'una forma eficaç amb molècules orgàniques petites.

Tot i l'enorme progrés aconseguit en el camp de la química mèdica fent servir aquesta mena de fàrmacs, les revolucions dels darrers anys en genòmica i proteòmica han descobert desenes de milers de noves proteïnes, que constitueixen els elements clau en les immenses xarxes de regulació cel·lular. Al mateix temps, això ha expandit l'univers de les dianes terapèutiques, obrint les portes d'una nova era en el desenvolupament de fàrmacs.

En les cèl·lules humanes, s'estima que les xarxes de regulació cel·lular comprenen unes 650.000 interaccions proteïna-proteïna (PPIs). Tot i això, coneixem actualment només una petita part d'aquesta xarxa, i els casos de fàrmacs que modulen PPIs es poden comptar per desenes. A banda de la recent descoberta de moltes d'aquestes PPIs, una de les principals raons d'aquest baix percentatge d'èxit radica en la mateixa natura de les PPIs. A diferència de les dianes tradicionals (enzims i receptors), la superfície d'interacció (o interfície) entre dues proteïnes és gran (1000-3.000 Å²) i relativament plana, i sovint no es disposa de substrats naturals que puguin servir d'inspiració per al disseny d'inhibidors. A més, l'espai químic tradicionalment explorat per les molècules petites orgàniques és considerablement diferent dels requeriments fisicoquímics que caracteritzen les PPIs.

Malgrat l'extensa superfície de contacte que caracteritza una PPI, existeixen una sèrie d'aminoàcids que contribueixen en gran mesura a l'energia global de la interacció; aquests punts crítics s'anomenen *hotspots*. Una anàlisi global dels milers de PPIs que es coneixen ha permès identificar un enriquiment dels *hotspots* en residus de triptòfan, tirosina i arginina i, en menor mesura, en àcid aspàrtic i histidina. Des d'un punt de vista estructural, les proteïnes poden interaccionar mitjançant epítops continus (un fragment lineal de cadena peptídica) o discontinus, en els que regions llunyanes de l'estructura primària formen una mateixa regió gràcies al plegament tridimensional de la proteïna. Una classificació més detallada, segons la forma i flexibilitat de les PPIs, està recollida a la *Taula I*.

Classe de PPI	Descripció	Esquema	Exemples
Proteïna-pèptid (epítop continu)	Un epítop continu (fulles β i bucles) s'uneix a una superfície proteica		<ul style="list-style-type: none"> • Integrasa VIH - LEDGF • Integrines • RAD51-BRCA2 • Dominis PDZ
Proteïna-pèptid (epítop discontinu)	Un domini helicoidal (epítop discontinu) s'uneix a una cavitat		<ul style="list-style-type: none"> • MDM2-p53 • BCL-XL-BAK • ZipA-FtsZ • MCL1-BH3
Proteïna-pèptid (residu àncora)	Un únic residu en el pèptid fa de grup s'ancora i domina la interacció		<ul style="list-style-type: none"> • Bromodominis • PDEδ-KRAS • Dominis SH2 • VHL-HIF1α
Pèptid-pèptid	Dos motius helicoidals formen una interfície elongada		<ul style="list-style-type: none"> • MYC-MAX • NEMO-IKK • Anexina II-P11
Proteïna globular - proteïna globular	Dos proteïnes amb epítops discontinus		<ul style="list-style-type: none"> • IL-2-IL-2R • TNF-TNF • EGF-EGFR

Taula I. Classificació estructural de les PPIs i exemples. Adaptat de Scott, D. et al. (ref. 12).

El potencial dels pèptids per modular PPIs

Com hem esmentat, les PPIs es caracteritzen per interfícies extenses i irregulars que dificulten l'acció eficaç de les molècules petites com a moduladors de la interacció. Tanmateix, els pèptids s'han erigit en candidats potencials per compensar les mancances dels fàrmacs tradicionals. D'una banda, els pèptids tenen una major flexibilitat estructural que les molècules orgàniques, el que els permet adaptar-se millor a la irregularitat de les superfícies proteiques. Ja que estan formats d'aminoàcids, els pèptids són capaços de mimetitzar a la perfecció els trets fisicoquímics de les proteïnes, a la vegada que la seva estructura modular permet generar una enorme diversitat estructural. Finalment, els pèptids són molècules biocompatibles (el seu metabolisme en el cos genera aminoàcids), el que es tradueix en una baixa toxicitat.

Malgrat els seus avantatges, els pèptids naturals pateixen d'una estabilitat en sang i teixits molt baixa, ja que són ràpidament degradats per proteases i altres enzims metabòlics. A més, són poc adients per la seva administració oral i tenen dificultats per crear membranes biològiques. Per tal de superar

aquestes mancances, els investigadors han desenvolupat en els darrers anys un ventall de modificacions i millores per tal d'aconseguir un perfil biofarmacèutic més adient.

En aquesta línia, les estratègies que han donat millors fruits són la introducció d'aminoàcids no naturals i de ciclacions en l'esquelet de les molècules peptídiques (*Figura 1*). En el cas dels pèptids cíclics, la manca d'extrem lliures els confereix una millor estabilitat davant de proteases, mentre que una menor flexibilitat redueix el cost entròpic d'adoptar la conformació bioactiva, el que es tradueix en una major activitat biològica.

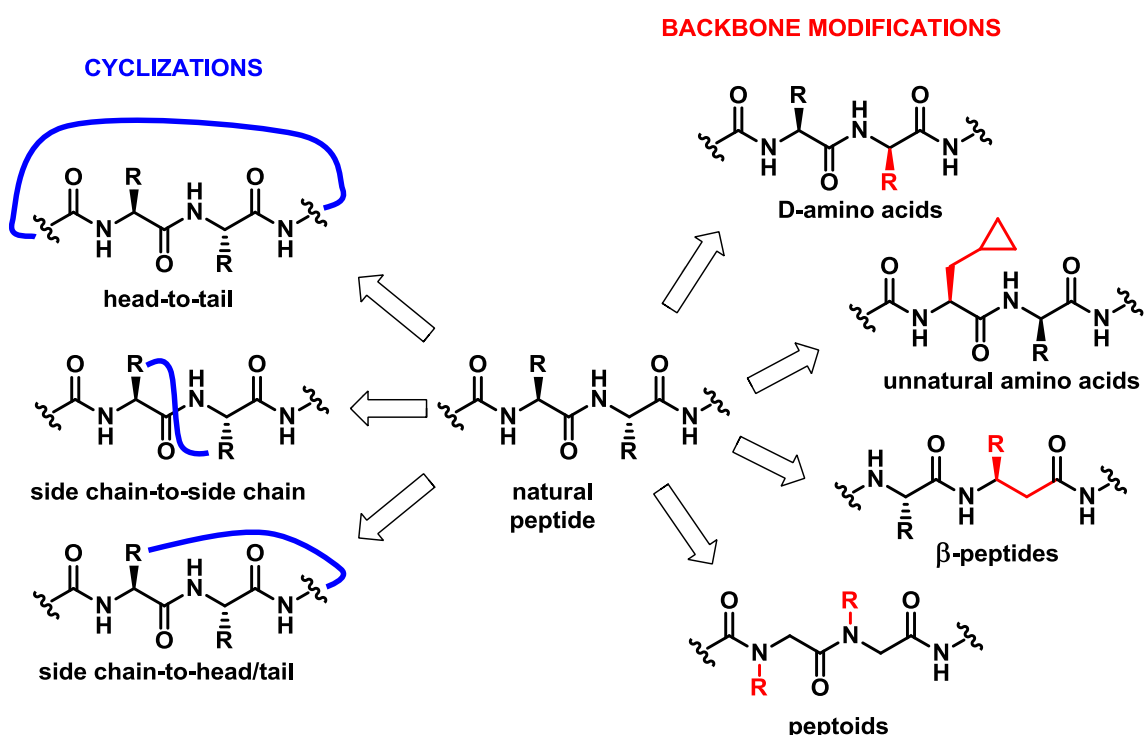


Figura 1. Principals tipus de modificacions estructurals efectuades sobre els pèptids per millorar la seva activitat moduladora de PPIs. Adaptat de Wójcik, P. et al. (ref. 16).

En la majoria de PPIs de les què tenim informació estructural, la regió interactora en una de les proteïnes adopta una estructura secundària ben definida, habitualment una hèlix α . En aquesta situació, un pèptid dissenyat per mimetitzar les característiques principals de l'hèlix α aconseguirà actuar com a inhibidor competitiu de la interacció. Per tant, l'estabilitat de l'estructura secundària del pèptid pot jugar un paper crucial en la seva capacitat de reconeixement molecular. Per tant, una de les estratègies més emprades per aconseguir una millor activitat és la rigidificació de l'estructura peptídica mitjançant motius cíclics.

Per aconseguir mimetitzar una hèlix α , els investigadors han desenvolupat un gran nombre de modificacions químiques que es poden introduir, amb relativa facilitat sintètica, sobre l'estructura dels pèptids. Entre elles, destaca la introducció d'una grapa o *staple* que fixa la posició de les cadenes laterals de dos aminoàcids situats en el mateix costat de l'hèlix α . També s'han descobert diferents formes de forçar el plegament d'un pèptid en forma de gir o de fulla β , tot i que aquests tipus de motius estructurals són menys abundants en les PPIs.

Totes aquestes estratègies de disseny i síntesi han fet que, en els darrers 20 anys, els pèptids en general, i els pèptids cíclics en particular, hagen proporcionat eines químiques per modular PPIs que abans es tenien per impossibles. Malgrat que encara cal millorar molts aspectes relacionats amb la farmacocinètica i les propietats ADME (administració, distribució, metabolisme i eliminació) dels pèptids, actualment comptem amb nombrosos exemples de pèptids cíclics en fases clíniques de desenvolupament (*Taula II*).

La interacció EGF-EGFR i el seu paper en el càncer

El factor de creixement epidèrmic (EGF) va ser descobert per Cohen i col·laboradors en 1962, qui van veure que estimulava la proliferació de cèl·lules epitelials en la pell i còrnia de ratolins, un descobriment pel qual Cohen va rebre el Premi Nobel de Medicina en 1986. Més tard, es va demostrar que EGF porta a terme la seva acció biològica mitjançant la unió a un receptor de membrana, el EGFR.

Les cèl·lules tumorals tenen l'habilitat de proliferar descontroladament i de colonitzar teixits i òrgans distants (un procés que es coneix com a metastasi), alterant la seva funció fisiològica. L'amplificació genètica del EGFR va ser el primer oncogen relacionat amb els càncer epitelials humans, i la hiperactivació d'aquest receptor està documentada a diferents tipus de tumor, com el càncer colorectal i el càncer de pulmó no microcític (sense cèl·lules petites). Tot i el seu paper en el càncer, l'activitat del EGFR és essencial per diversos processos fisiològics, com el desenvolupament neuronal i dels teixits epidèrmics durant la fase embrionària.

Nom	Àrea terapèutica	Fase clínica	Empresa
Aplidin®	Oncologia	III	PharmaMar
Debio025	Malalties infeccioses	III	Novartis
MK4261	Malalties infeccioses	III	Merck
PT141	Hematologia	III	Palatin Technologies
ALRN6924	Oncologia	II	Aileron Therapeutics
APL1	Oftalmologia	II	University of Pennsylvania & Alcon Laboratories
APL2	Hematologia	II	University of Pennsylvania & Potentia
ASP3291	Gastroenterologia	II	Astellas Pharma
AT1001	Desordres metabòlics	II	Alba Therapeutics
AZP531	Cardiovascular	II	Eli Lilly
MEN11420	Gastroenterologia	II	Menarini
PL3994	Cardiovascular	II	Palatin Technologies
POL6326	Immunologia	II	Polyphor
RG7929	Malalties infeccioses	II	Roche
SCY635	Malalties infeccioses	II	Waterstone Pharmaceuticals
Voclera	Dermatologia	II	Roche
Vosoritide	Desordres metabòlics	II	BioMarin
ALRN5281	Desordres metabòlics	I	Aileron Therapeutics
OBP801	Oncologia	I	Astellas Pharma
RA101495	Hematologia	I	Ra Pharmaceuticals

Taula II. Llistat de pèptids cíclics que es troben en fases clíniques de I a III (extret de Medtrack, a 23 de febrer de 2016).

En un context oncogènic, el EGFR esdevé hiperactivat per diferents mecanismes, i és un dels principals responsables de la proliferació de les cèl·lules tumorals, així com del desenvolupament dels vasos sanguinis que aporten nutrients i oxigen al tumor (un procés que es coneix com angiogènesi). En alguns tumors, el EGFR representa el paradigma de l'addició oncogènica, en què les cèl·lules tumorals són completament dependents d'aquesta via de senyalització. La diversitat genètica i bioquímica del càncer explica que el EGFR estigui sobreexpressat i/o mutat més freqüentment en alguns tipus de càncer (*Taula III*).

La família de receptors de membrana ErbB, i els seus lligands

El EGFR (o ErbB1) és el membre principal de la família de receptors membranals ErbB, que consta d'altres tres membres: ErbB2, ErbB3 i ErbB4. Totes aquestes proteïnes comparteixen una estructura semblant, amb una regió extracel·lular on s'uneix el lligand, un domini transmembrana i una regió intracel·lular amb activitat tirosina-quinasa, que transmet la senyal activadora a l'interior de la cèl·lula. L'excepció a la regla és el ErbB2, la regió extracel·lular del qual no s'uneix a cap lligand i es troba en una conformació permanentment activa (*Figura II*).

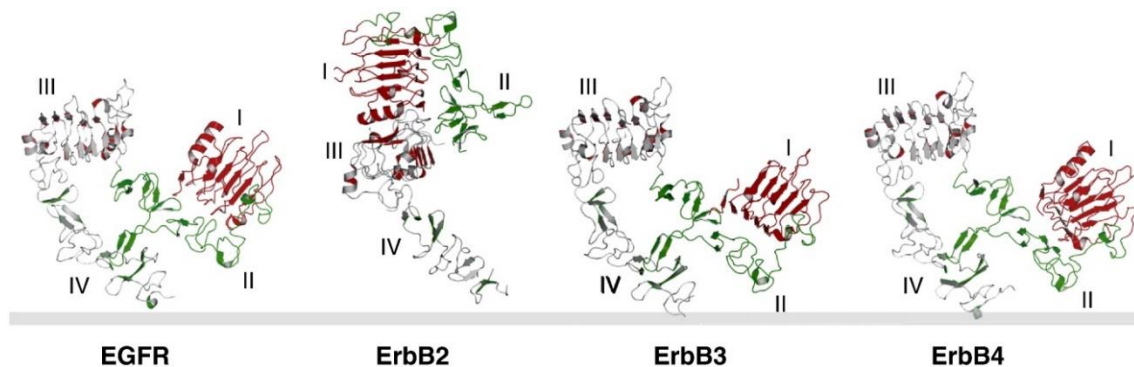


Figura II. Estructura extracel·lular dels quatre receptors ErbB. ErbB2 és l'únic que en absència de lligand adopta una conformació activa. Extret de Lemmon, M. (ref. 43).

Fins a data d'avui s'han descrit 13 lligands extracel·lulars diferents que s'uneixen als receptors ErbB1, els quals s'agrupen sota la categoria de factors de creixement. Aquestes proteïnes comparteixen una o més repeticions d'una seqüència peptídica consens ($CX_7CX_{4-5}CX_{10-13}CX_8GXRC$), on X representa qualsevol aminoàcid. Com es pot observar, aquesta seqüència conté 6 cisteïnes que formen 3 ponts disulfurs intramoleculars, el que genera una estructura globular amb 3 bucles o *loops* que són essencials per la seva alta afinitat pels receptors ErbB. Tots ells tenen diferents afinitats i especificitats per cadascun dels receptors ErbB. En qualsevol cas, la unió d'aquestes proteïnes produeixen la homo- o heterodimerització dels ErbB. De tots ells, EGF és qui té una major afinitat pel EGFR, i el que més sovint s'ha identificat com a principal promotor del desenvolupament tumoral.

Estructura i dinàmica del EGFR

Com passa amb altres receptors amb activitat tirosina-quinasa, la unió de l'EGF o d'altres factors de creixement es produeix a la regió extracel·lular del receptor, la qual està dividida en quatre dominis (I-IV). Els dominis I i

III comparteixen un 37% d'homologia en la seva seqüència, i participen en les interaccions amb el lligand (EGF). En el domini II, existeix una regió amb estructura de β -hairpin (anomenat braç de dimerització) que juga un paper clau en la dimerització del receptor.

En l'estat inactiu de l'EGFR, el domini II està protegit per una sèrie de contactes intramoleculars amb el domini IV, però un cop s'uneix l'EGF es produeix un canvi conformacional massiu en el receptor que genera l'exposició d'aquest braç de dimerització (*Figura III*). En la forma activa i dimèrica del receptor, aquest β -hairpin estableix nombroses interaccions amb el β -hairpin del receptor veí, estabilitzant aquest dímer.

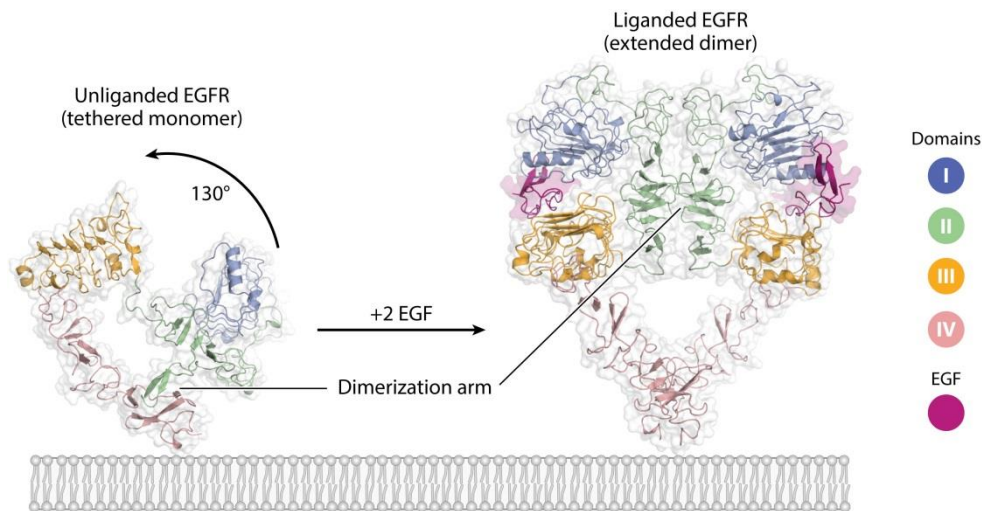


Figura III. La unió de l'EGF induïx una rotació de $\sim 130^\circ$ de rotació en els dominis I i II de l'EGFR, respecte als dominis III i IV, el que possibilita la formació del dímer i l'activació del receptor. Adaptat de Kovacs, E. et al. (ref. 53).

Com a proteïna, EGFR té una estructura molt dinàmica i pot explorar un ventall molt divers de conformacions. A diferència d'altres mecanismes d'activació de receptors semblants (com el VEGF), l'EGF no fa una contribució directa a la interfície de dimerització de l'EGFR, sinó que induïx el canvi conformacional que permet l'associació dimèrica del receptor. Aquest és el model que representa l'encaix induït o *induced fit*. Segons el model alternatiu (però no exclouent) de la selecció conformacional, existirien a la membrana de les cèl·lules dos poblacions conformacionals de EGFR: una majoritàriament inactiva, i una minoria de receptors en un estat pre-dimèric, amb els llocs d'unió al lligand exposats.

De fet, aquest darrer cas és el que explicaria la no-linearitat de les representacions de Scatchard de la unió de EGF a EGFR. Aquest fet revela l'existència d'un petit percentatge de receptors amb alta afinitat per l'EGF ($K_D \sim 50$ pM), i una majoria de receptors ($\sim 95\%$) amb una afinitat més baixa

($K_D \sim 3$ nM). En qualsevol cas, la dimerització produeix l'aproximació espacial dels respectius dominis intracel·lulars, el que activa la fosforilació d'una sèrie de residus de tirosina situats en aquesta part del receptor.

Vies de senyalització i respostes cel·lulars

Una vegada activat l'EGFR, els residus de fosfotirosina esdevenen llocs de reconeixement per una sèrie de proteïnes efectores, com el complex Ras/Raf o les quinases PI3K i MAPK. La visió clàssica de la senyalització cel·lular defensava la propagació vertical de la senyal, mitjançant diferents proteïnes, produint un efecte lineal en la cèl·lula. Tanmateix, els avanços en biologia de sistemes han permès descriure la via de senyalització del EGFR sota la forma de rellotge de sorra, on un conjunt d'*inputs* (els lligands o factors de creixement) produeixen la seva acció a través dels receptors ErbB1 (*Figura IV*). Després, la dimerització d'aquests receptors produeix l'activació d'una gran diversitat de proteïnes efectores, el que genera una xarxa de senyalització altament interconnectada i complexa. Aquesta interconnexió confereix al sistema d'una gran robustesa, fet que fa que el bloqueig d'un node de la xarxa sigui fàcilment suplert per l'activació de vies complementàries.

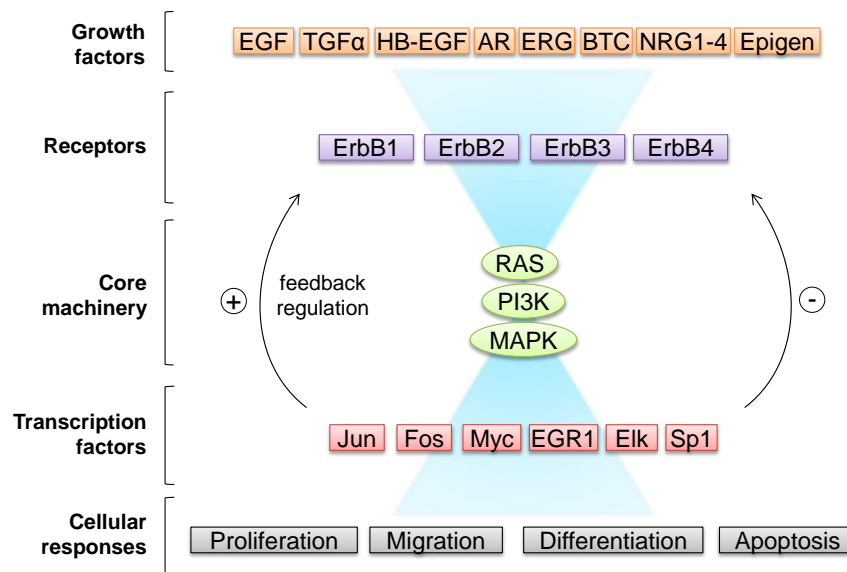


Figura IV. Els elements de la via de senyalització de l'EGFR estan interconnectats en una xarxa en forma de rellotge de sorra, on una àmplia gama d'inputs modula un nombre petit de processos principals. A l'altra banda, les senyals d'aquests processos és llegida per una varietat de proteïnes que constitueixen el output de la xarxa. Adaptat de Citri, A. i Yarden, Y. (ref. 63).

Desenvolupament d'inhibidors de l'EGFR pel tractament dels càncer

Des de la identificació del EGFR amb el càncer, els investigadors han estat activament implicats en la descoberta de fàrmacs capaços de bloquejar l'excés d'activitat d'aquesta proteïna en els processos tumorals. Amb l'avanc de la biotecnologia, aquest interès en nous fàrmacs s'ha traduït en el desenvolupament d'anticossos monoclonals (mAb) contra el EGFR (*Figura V*). L'any 2004 es va aprovar el primer mAb anti-EGFR (cetuximab) pel tractament del càncer colorectal, i des de llavors el seu ús terapèutic s'ha expandit a altres tipus de neoplàsies. El cetuximab s'uneix al domini extracel·lular del receptor ($K_D = 0.39$ nM) amb una afinitat més alta que el EGF ($K_D = 1.9$ nM), bloquejant la seva unió i la seva activació. Posteriorment, s'han desenvolupat altres anticossos completament humanitzats contra l'EGFR (com a panitumumab) amb afinitats més altes que cetuximab, capaços de mantindre l'activitat davant de mutacions al receptor (com a S492R) que confereixen resistència a cetuximab.

Un segon tipus d'agents terapèutics que bloquegen EGFR són els inhibidors tirosina-quinasa (TKI), molècules orgàniques petites que actuen com a inhibidors competitius de l'ATP pel seu lloc d'unió al domini tirosina-quinasa intracel·lular (*Figura V*). El primer compost, gefitinib, va ser aprovat al 2003 pel seu ús en pacients amb càncer de pulmó no microcític. Des de llavors, altres fàrmacs TKI més potents i menys tòxics s'han desenvolupat, alguns amb un mecanisme d'acció irreversible. Tanmateix, l'administració d'aquests fàrmacs està fortament relacionada a l'aparició de resistències secundàries, el que limita la seva efectivitat en el tractament.

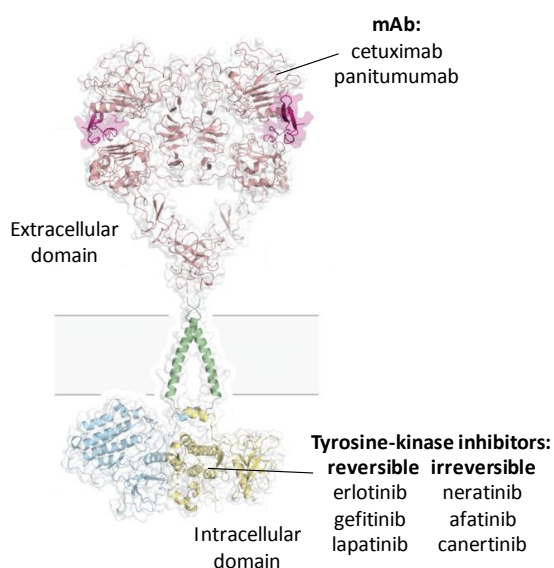


Figura V. Esquema general del mode d'acció dels diferents fàrmacs anti-EGFR actualment comercialitzats.

Una altra estratègia per inhibir l'activació de l'EGFR (i precisament la que es persegueix en aquesta tesi) és bloquejar directament el seu principal lligand, l'EGF. A diferència d'altres factors de creixement, com el VEGF—on s'han publicat casos d'èxit amb fàrmacs inhibidors, n'hi han escassos precedents d'inhibidors de l'EGF a la literatura. Aquesta manca d'èxit és majorment deguda al fet que EGF és una proteïna globular petita amb molt poques regions i cavitats adients per interaccionar amb fàrmacs. De fet, una droga originalment desenvolupada pel tractament de les parasitosis, la suramina, és l'únic lligand químic de l'EGF que s'ha descrit a data d'avui. Tot i això, la suramina té una afinitat molt feble per l'EGF, i interacciona de forma molt poc selectiva, ja que també interfereix amb altres proteïnes com el PDGF o el VEGF.

Estratègies per superar la resistència del càncer als actuals fàrmacs anti-EGFR

Tot i les grans aportacions que els fàrmacs anti-EGFR han fet al tractament del càncer, no tots els pacients amb càncer responen bé al tractament amb aquest tipus de fàrmacs i, els que ho fan inicialment, acaben desenvolupant resistència. Existeixen dos tipus de resistència: primària o intrínseca, i secundària o adquirida. La resistència primària és la manca de resposta inicial a la teràpia, i es dona en alguns genotips de tumors que presenten mutacions en el EGFR o en proteïnes que es troben en la seva via de senyalització.

La resistència secundària o adquirida, d'altra banda, es dona en pacients que inicialment milloren amb la teràpia anti-EGFR però que, després d'un temps de tractament, esdevenen insensibles al tractament. La causa principal d'aquestes mutacions és la inestabilitat genètica de les cèl·lules tumorals, la qual s'incrementa sota la pressió selectiva imposada pels fàrmacs anti-EGFR. Des d'un punt de vista mecanístic, aquest tipus de mutacions es poden agrupar en 4 categories (*Taula III*), de les quals la primera i segona són les més habituals. Tanmateix, el càncer és una malaltia molt heterogènia, i diferents tipus de mutacions poden fàcilment coexistir en un mateix pacient.

	Lung	Colon	Glioblastoma	Head and neck
Primary resistance	EGFR-T790M BIM deletion	Activating mutations in KRAS, PI3K, BRAF, PTEN loss	R84K A265V/D/T Extracellular domain deletions (EGFRvIII)	
Acquired resistance				
EGFR mutation	T790M	S492R	EGFRvIII	EGFRvIII
Alternative pathway activation	ErbB2, ErbB3, VEGF, IGF, FGF, BRAF, PTEN loss	ErbB2, IGF, KRAS	PDGF, PI3K, STAT3, MET, PTEN loss	ErbB2, ErbB3, Aurora
Histologic transformation	Acquisition of stem cell properties Epithelial-to-mesenchymal transition SCLC transformation			Epithelial-to-mesenchymal transition

Taula III. Esquema general dels mecanismes de resistència a l'EGFR en diferents tipus de càncer humà. A la figura de la dreta es mostren les mutacions més habituals a l'EGFR. Adaptat de Chong, C. i Jänne, P. A. (ref. 38).

Per tant, aquesta problemàtica relacionada amb la manca d'eficàcia a llarg termini dels actuals fàrmacs anti-EGFR urgeix el desenvolupament de noves aproximacions terapèutiques per combatre aquesta malaltia. En concret, la recerca sobre nous mecanismes d'acció, així com la combinació de diferents tipus de fàrmacs amb mecanismes complementaris, hauria de permetre un millor control sobre la progressió del càncer i un major temps i qualitat de vida pels malalts de càncer.

Pèptids covalents com a inhibidors específics de proteïnes: el cas de la prolil-oligopeptidasa (POP)

Com hem vist a la primera part de la introducció, els pèptids són estructures modulars i diverses que poden ser dissenyats per modular específicament interaccions proteïques que estan involucrades en processos patològics. En aquest context, el pèptid competeix amb un substrat endogen o una proteïna, actuant com a inhibidor. Degut a la natura transitòria de les PPIs, aquests pèptids presenten constants de dissociació elevades del seu lloc d'unió.

Tot i que aquesta classe d'inhibició és adient per modular PPIs, la unió a cavitats o regions específiques (com el centre actiu d'un enzim) està governada per interaccions molt més fortes. En aquest context, una ràpida dissociació del pèptid inhibidor de la seva diana representa una gran limitació. Com a solució a aquest problema, alguns investigadors han ideat els pèptids covalents, en els que un grup químic electrofílic s'incorpora a l'esquelet peptídic, amb la finalitat que reaccioni covalentment amb algun residu de la proteïna diana.

Malgrat que els inhibidors covalents han estat tradicionalment jutjats com a perillosos i poc adients per l'ús en humans, la cerca de fàrmacs més actius i amb una durada d'acció més llarga ha revifat l'interès en els fàrmacs covalents, especialment aquells que incorporen electròfils poc reactius (*mild electrophiles*). De fet, 3 dels 27 fàrmacs aprovats l'any 2013 per la FDA americana van ser inhibidors covalents. A pesar d'aquestes dades, el nombre d'exemples de pèptids i peptidomimètics amb mecanisme d'acció covalent és molt reduït. Recentment, el grup de Walenksy ha desenvolupat uns pèptids grapa α -helicoidals que incorporen una acrilamida com a grup reactiu. Aquests pèptids són capaços de reaccionar selectivament amb un residu de cisteïna situat a la interfície de la proteïna BFL-1, i bloquejar la PPI amb BH3.

Les proteases com a dianes terapèutiques

Les proteases són enzims catabòlics participen en la síntesi, reciclatge i degradació dels pèptids i proteïnes endògens. La disfunció en algun d'aquests enzims està implicat en un gran nombre de malalties, i actualment existeixen inhibidors de proteases dirigits a tractar malalties tan diverses com la hipertensió, les infeccions pel VIH i per l'hepatitis C, el càncer i la diabetis de tipus II. Les proteases es caracteritzen per tindre una estructura dividida en diferents dominis, on el domini catalític està íntimament connectat a un o més dominis reguladors. Aquests dominis reguladors són el lloc d'unió de diferents molècules i proteïnes missatgeres, que modulen l'activitat catalítica de l'enzim, el que tot plegat fa de les proteases enzims molt dinàmics que exploren una gran diversitat de conformacions.

POP, un enzim amb efectes pleiotròpics

La prolil-oligopeptidasa (POP) és una proteïna citosòlica que hidrolitza pèptids de menys de 30 aminoàcids, específicament a l'extrem C-terminal dels residus de prolina. Existeixen uns pocs enzims, agrupats en la família

S9a de proteases serines, que són capaços de tallar la cadena peptídica a aquest nivell, i juguen un paper molt important en la regulació dels neuropèptids que contenen prolina, com per exemple oxitocina, vasopressina o angiotensina.

La POP s'expressa de manera general a tot l'organisme, tot i que és al cervell on els seus nivells d'expressió són més elevats. A dia d'avui, la funció biològica de la POP no és del tot coneguda, així com les seves implicacions en l'aparició de malalties. A més del seu paper regulador dels nivells de diferents neuropèptids, es pensa que la POP està involucrada en la regulació dels nivells d'inositol trifosfat i de calci intracel·lular, uns mecanismes que estan activats durant els processos de memòria i aprenentatge. Més recentment, s'ha postulat que la POP pot actuar a través de PPIs, ja que s'ha demostrat en diferents assajos *in vitro* (i actualment *in vivo*) que la POP co-localitza amb proteïnes com GAP-43 i α -sinucleïna, aquesta darrera una de les principals senyals bioquímiques de la malaltia de Parkinson.

Experimentalment, s'ha vist que l'administració d'inhibidors de POP produeix efectes anti-amnèsics i potenciadors de la memòria en rates i primats. Igualment, l'administració d'aquests fàrmacs a models animals de Parkinson ha aconseguit disminuir els agregats cerebrals d' α -sinucleïna, així com millorar la resposta motora dels animals. En resum, tot i que els mecanismes no es coneixen amb detall, la hipòtesi més acceptada indica una relació directa entre la disfunció de la POP i els trastorns cognitius i neurodegeneratius.

La química mèdica dels inhibidors de la POP

Degut a l'interès biològic d'aquesta proteasa, la comunitat científica ha mostrat un gran interès en el desenvolupament de fàrmacs capaços de bloquejar o modular l'activitat de la POP. La majoria d'inhibidors descrits són peptidomimètics que deriven del Z-prolilprolinal (ZPP), un dels primers inhibidors descrits. De fet, el ZPP es comporta com un inhibidor anàleg de l'estat de transició, on el grup aldehid reacciona amb la serina 554 del centre actiu formant un enllaç hemiacetal.

L'estructura del ZPP s'ha agafat com a referència pel disseny de molts dels inhibidors descrits de la POP, on els grups farmacòfors P1', P1, P2 i P3 s'uneixen a les cavitats S1, S2 i S3 del centre actiu de la POP, d'acord a la nomenclatura de Berger i Schechten per aquests enzims. En la posició P1', s'han desenvolupat una diversitat de grups funcionals, alguns d'ells capaços de reaccionar de forma covalent amb la cadena lateral de la serina 554. En

els inhibidors reversibles, aquest enllaç covalent és hidrolitzat al cap d'un temps, de manera que l'enzim recupera la seva activitat. Respecte als inhibidors irreversibles, existeixen molt pocs exemples i es limiten a fàrmacs amb acció antiparàsita, mai testats en humans. Una breu descripció de les relacions estructura-activitat dels inhibidors de la POP es recull a la *Figura VI*.

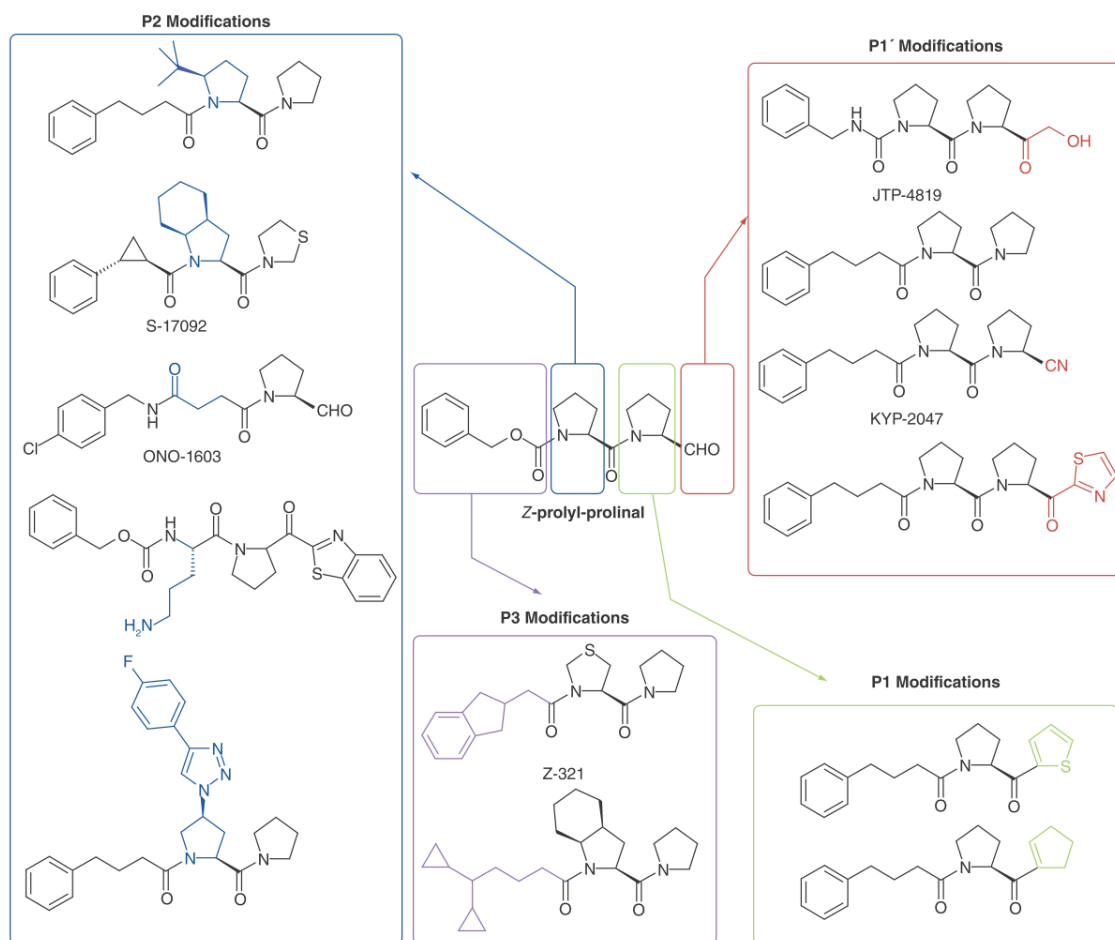


Figura VI. Exemples d'inhibidors de la POP amb estructura de peptidomimètic, obtinguts a partir de modificacions del compost Z-prolilprolinal. Extret de López, A. et al. (ref. 95).

Tot i l'efecte d'aquests inhibidors peptídics en assajos *in vitro*, les escasses proves que s'han fet en animals superiors i humans han estat decebedores. Una de les principals causes d'aquesta manca d'eficàcia *in vivo* és la baixa concentració de fàrmac que arriba al cervell. Habitualment, els esforços de millora van dirigits a una major activitat *in vitro*, mentre que altres paràmetres, com la capacitat de creuar la barrera hematoencefàlica, no són tinguts en compte. En definitiva, calen noves eines químiques per comprendre millor la funció biològica de la POP, així com per validar aquesta diana terapèutica en l'àmbit de les malalties cognitives.

Capítol 1. Ús de tècniques de *docking* pel disseny de nous lligands peptídics d'EGF

Les tècniques computacionals són d'una gran utilitat en la recerca de nous fàrmacs, especialment en aquells casos on no existeixen molècules descrites que interactuen amb la proteïna diana, com és el cas d'EGF, a partir de les quals poder iniciar el procés de disseny. Entre les diferents tècniques computacionals, els experiments de *docking* tenen la finalitat de predir l'orientació relativa d'un lligand sobre una proteïna, i de donar una estimació de l'energia d'interacció del complex. A més, les tècniques de *docking* són més simples i econòmiques (en termes d'hores de càlcul) que les simulacions per dinàmica molecular (MD).

Fent servir l'estructura de RMN de l'EGF (PDB 2KV4) com a model, vam identificar regions potencialment útils per albergar pèptids inhibidors (el que es coneix com a *binding sites*). Sobre una d'aquestes regions, una cavitat hidrofòbica situada entre els residus de tirosina 13 i 29, vam fer el *docking* d'una llibreria de tripèptids que combinava tots els aminoàcids naturals en els extrems, mantenint un residu de prolina en la posició central. Amb el millor resultat de la llibreria (el tripèptid Arg-Pro-Arg) vam realitzar successius cicles de modificacions i *docking* per tal de millorar l'encaix del pèptid i l'afinitat teòrica del complex. Finalment, vam arribar al tripèptid Ser-2Nal-Arg (**d1**) que presentava una energia de -8.25 kcal/mol. L'addició d'un residu de glutàmic al N-terminal del pèptid, formant el tetrapèptid Glu-Ser-2Nal-Arg (**d2**) semblava millorar, tot i que lleugerament, l'energia d'interacció fins a -8.44 kcal/mol (*Figura VII*).

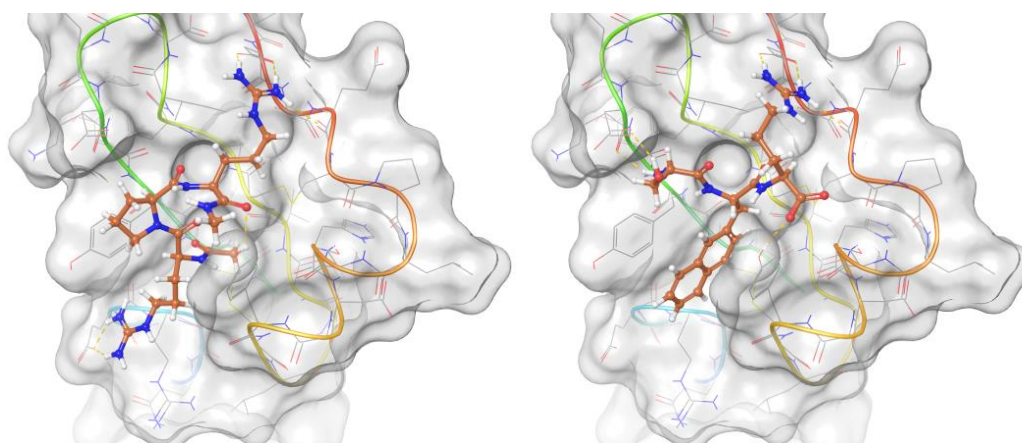


Figura VII. Representació de les interaccions de les poses de mínima energia dels tripèptids Arg-Pro-Arg (esquerra) i Ser-2Nal-Arg (dreta), resultat del *docking* sobre la superfície d'EGF.

Per tal d'avaluar experimentalment la interacció d'aquests pèptids amb EGF, al laboratori no disposàvem de cap assaig biofísic implementat. Per tant, vam centrar inicialment els esforços en posar a punt una tècnica de fluorescència, basada en els canvis d'intensitat d'un grup fluoròfor (4-DMAP) segons la polaritat de l'entorn. En un entorn polar (medi aquós), el 4-DMAP genera una emissió molt feble, però en entorns hidrofòbics (quan el lligand s'uneix a la proteïna) es produeix un increment dramàtic en la fluorescència. A banda d'aquesta propietat, la gran avantatge d'aquest fluoròfor és que té unes mides semblants a l'anell de triptòfan, el que fa que es pugui incorporar a un lligand peptídic en substitució d'un aminoàcid aromàtic, sense causar grans canvis en l'estructura del pèptid.

Un cop vam sintetitzar els anàlegs fluorescents de **d1** i **d2**, vam realitzar l'assaig de titració del pèptid amb EGF. Malgrat que vam arribar a concentracions altes d'EGF, vam observar només petits canvis en l'espectre de fluorescència del lligand. Aquests resultats negatius podien ser causa de manca d'afinitat per part dels lligands, de canvi insuficient en la polaritat de l'entorn del fluoròfor unit, o de una combinació dels dos. El fet que EGF sigui una proteïna petita i exposada a l'aigua del dissolvent, dificulta sens dubte l'aplicació d'aquest mètode, el que ens va portar a explorar tècniques biofísiques alternatives.

La calorimetria és una tècnica basada en la mesura directa de l'energia calorífica que es desprèn o s'absorbeix durant una reacció química. Igualment, la mesura d'aquesta energia durant la formació d'una interacció entre un lligand i una proteïna, el que és la base de la calorimetria de titració isotèrmica (ITC), permet calcular paràmetres interessants sobre el procés d'unió, com ara el component entròpic i entàlpic de l'energia d'interacció.

Pels nostres lligands, l'addició de **d1** i **d2** a una dissolució d'EGF a una concentració de 15 μM es va traduir en canvis en la senyal isotèrmica del calorímetre, el que mostrava l'existència d'una interacció entre lligand i proteïna. Malgrat utilitzar fins a 40 equivalents de pèptid, no vam aconseguir arribar al plateau de saturació, el que dificultava el càlcul de les constants termodinàmiques. De fet, l'aplicació de l'ITC per quantificar interaccions febles és problemàtica. En aquestes condicions, la forma típicament sigmoide de la corba es perd, i cal fixar el coeficient estequiomètric n per tal de poder estimar els valors de K_D i ΔG .

En paral·lel, vam dedicar una part dels esforços a posar a punt un sistema d'expressió recombinant d'EGF, amb l'objectiu últim de realitzar

experiments de RMN amb una forma isotòpicament marcada de la proteïna. L'expressió en *E. coli* no va ser una tasca trivial, degut al baix rendiment de l'expressió i a l'habitual aparició d'espècies d'EGF polimèriques o mal plegades, amb els ponts disulfur incorrectament formats. Finalment, la utilització d'una soca d' *E. Coli* amb un ambient citoplàsmic pro-oxidatiu (SHuffle T7), la qual expressa constitutivament l'enzim disulfur-isomerasa, ens va permetre aïllar l'EGF correctament plegada, tot i que amb rendiments molt baixos. La fusió del gen de l'EGF a la proteïna SUMO es va traduir en un increment del rendiment.

Amb aquesta darrera metodologia, vam aconseguir expressar EGF uniformement marcada en ^{15}N pel seu ús en RMN. L'espectre de RMN de la proteïna coincidia amb l'anteriorment publicat, confirmant l'efectivitat del nostre sistema d'expressió. Mitjançant l'adquisició d'un espectre d'HSQC ^1H - ^{15}N , obtenim una empremta característica de senyals, on cada senyal representa un aminoàcid de la proteïna. A més a més, aquest experiment ens permet detectar la unió de lligands a EGF i, el que és més important, situar la zona d'interacció a la proteïna. Incubant EGF amb una concentració elevada dels pèptids **d1** i **d2**, l'experiment de RMN ens va mostrar canvis en les senyals d'alguns aminoàcids situats en el bucle A de la proteïna (L8, S9, G12, H16, D17). Altres residus, situats al bucle C d'EGF, però que es troben en proximitat dels anteriors, també van resultar modificats per la interacció amb els pèptids (V34, V35, Y44).

Per tal d'identificar les característiques dels pèptids essencials per la interacció, vam sintetitzar una sèrie d'anàlegs amb diferents modificacions (*Taula IV*). En **d3**, l'aminoàcid central va ser reemplaçat per Leu, el que posseeix una cadena lateral hidrofòbica però no aromàtica, que en principi no hauria de ser capaç d'entrar a l'esclatxa entre Tyr13 i Tyr29 de la proteïna. En **d4**, els extrems peptídics van ser acetilat i amidat per tal d'eliminar les càrregues electrostàtiques. En **d5**, els residus 2 i 3 es van intercanviar per tal de determinar el grau d'especificitat de seqüència de la interacció. Finalment, en **d6** la càrrega positiva que proporciona l'Arg va ser eliminada, al canviar aquest residu per Leu.

Com van confirmar els resultats de SAW, de tots els anàlegs **d3-d6** només **d4**, en el que es conserven tots els trets estructurals del pèptid **d1**, va ser capaç de produir canvis similars a l'espectre d'HSQC ^1H - ^{15}N de la proteïna. Per contra, **d5** va produir canvis molt més modestos, mentre que **d3** i **d6**, que manquen dels elements claus per efectuar la unió modelitzada per *docking*, no van afectar l'espectre de la proteïna.

Malgrat els sòlids resultats obtinguts per RMN, aquesta tècnica consumeix quantitats elevades de proteïna isotòpicament marcada. A més, la baixa afinitat dels nostres pèptids dificulta la determinació de la constant d'afinitat (K_D) degut a les altes concentracions de lligand que són necessàries a l'experiment. Per tant, com a mètode biofísic alternatiu, vam testejar l'equip d'ona acústica superficial (SAW) recentment adquirit pels serveis d'Espectrometria de Masses del IRB.

En aquest biosensor de SAW, la proteïna s'enganxa (mitjançant la cadena lateral de les Lys) a un microxip, a través del qual viatgen ones sonores amb una fase i amplitud determinades (*Figura VIII*). Com que les ones acústiques recorren la superfície del xip, qualsevol canvi en la massa unida a la superfície, així com canvis importants en la conformació de la proteïna enganxada, poden ser detectats en forma de modificacions en la fase i amplitud de les ones. A la pràctica, a l'injectar diferents concentracions d'un lligand, aquest fenomen es tradueix en corbes que representen l'associació i dissociació del lligand, el que permet determinar les constants cinètiques (K_{on} , K_{off}) de la interacció, així com la K_D .

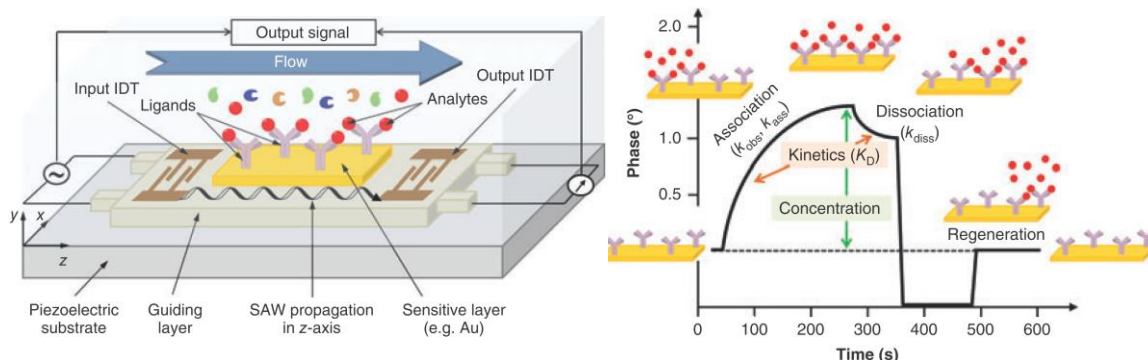
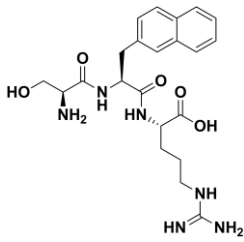
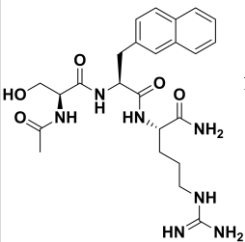
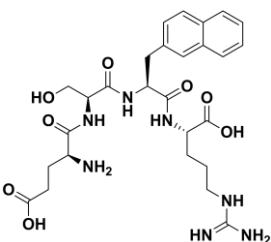
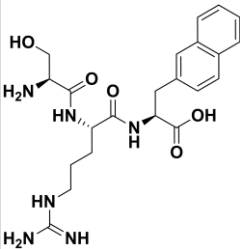
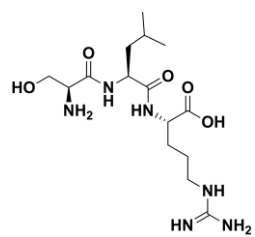
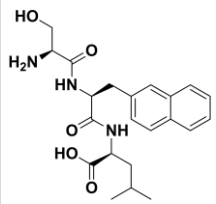


Figura VIII. Esquema d'un dispositiu SAW. Un transductor genera ones acústiques que es propaguen sobre la superfície d'un cristall piezoelèctric. Aquestes ones viatgen a través del xip i reben les modificacions dels vents bioquímics que ocorren a la superfície del xip. Un segon transductor, situat al final del xip, converteix la fase i amplitud de l'ona en senyal elèctrica. Típic cicle d'associació-dissociació obtingut en un experiment de SAW (dreta). Extret de Díaz-Lobo, M. (ref. 120).

Per validar la tècnica i la funcionalitat de la proteïna immobilitzada, vam injectar el domini extracel·lular de l'EGFR a concentracions nanomolars, obtenint una bona interacció que vam poder quantificar amb una K_D de 8.2 nM, en concordança amb els valors publicats obtinguts per ressonància de plasmó superficial (SPR), una tècnica conceptualment molt similar. Amb la tècnica validada, els pèptids **d1-d6** van ser injectats a diferents concentracions, el que ens va permetre observar canvis en les senyals de

SAW per alguns d'ells, i poder calcular les afinitats (K_D) de les interaccions. Com es recull a la *Taula IV*, el pèptid **d1** va mostrar l'afinitat més gran per EGF, confirmant la tendència que havíem vist per RMN.

Id	Estructura	Canvis	K_D	Id	Estructura	Canvis	K_D
d1		Pèptid dissenyat per <i>docking</i>	1.4 ± 0.2 mM	d4		Capping N- i C-terminal (elimina les càrregues iòniques)	3.4 ± 0.5 mM
d2		Extra Glu per unir-se a Lys28 d'EGF	1.5 ± 0.8 mM	d5		Intercanvi dels residus 2-3	2.1 ± 0.2 mM
d3		Mutació de 2Nal a Leu (sense anell aromàtic)	No unió	d6		Arg mutada a Leu	2.6 ± 1.4 mM

Taula IV. Llibreria de pèptids **d1-d6**. Els valors d'afinitat van ser calculats per SAW.

Capítol 2. Disseny de lligands d'EGF basats en mimetitzar epítops interactors de l'EGFR

L'EGFR és una proteïna transmembrana de 170 kDa formada per molts dominis flexibles, entre els quals es diferencia la regió extracel·lular, lloc d'unió de factors de creixement com l'EGF. La regió extracel·lular de l'EGFR està constituïda per quatre dominis; els dominis I, II i III de l'EGFR estan disposats en forma de "C", amb l'EGF situat al mig envoltat pels dominis I i III. Com es tracta d'una proteïna molt dinàmica i difícil de cristal·litzar, els investigadors han trencat l'estructura del receptor en diferents parts, que han estat estudiades per cristal·lografia de raigs X, principalment.

La unió de l'EGF al domini I del receptor, en la seva forma activa (PDB 1IVO), es dona al llarg d'una extensa superfície de 720 Å² i presenta una sèrie d'interaccions específiques (*Figura IX*). Per exemple, els residus 16-18 de l'EGFR i els aminoàcids 31-33 de l'EGF formen una petita fulla β, amb diversos ponts d'hidrogen que estableixen la interacció. A més, les cadenes laterals de Leu14, Tyr45, Leu69, i Leu98 en l'EGFR formen interaccions hidrofòbiques amb Met21, Ile23 i Leu26, en el bucle B de l'EGF.

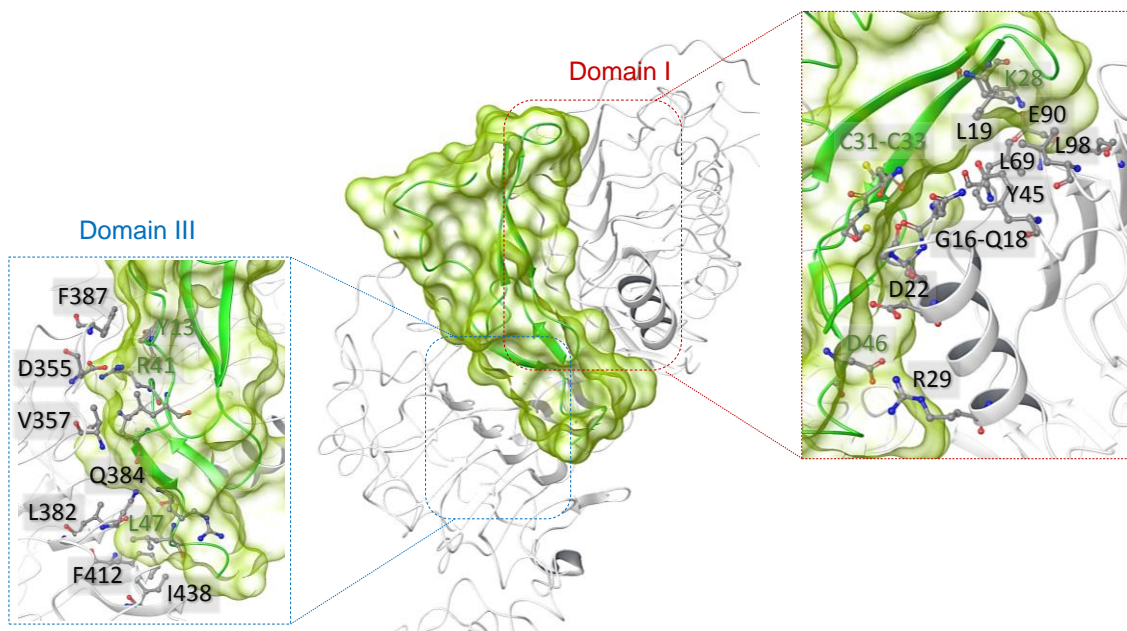


Figura IX. Detall de l'estructura del complex EGF-EGFR en la seva forma activa (PDB 1IVO). EGF es mostra com a superfície gris, EGFR en cinta gris. S'especifiquen els residus principals de la interacció.

La unió de l'EGF al domini III del receptor és més forta, malgrat abastir una superfície semblant (720 \AA^2) que el domini I (*Figura IX*). En aquesta regió, les cadenes laterals de Val350 i Phe357 interaccionen hidrofòbicaament amb Leu15 i Tyr13, respectivament, de l'EGF. A més, la cadena lateral de l'Asp355 forma una interacció iònica crucial amb l'Arg41 de l'EGF.

Reproducció d'un bucle heptamèric de l'EGFR amb pèptids cíclics

En la cerca d'epítops interactors de l'EGFR, vam focalitzar primer la nostra atenció en un epítop continu, situat al domini I del receptor, que en la forma unida a la proteïna adopta una estructura de bucle quasi-cíclica. D'aquesta forma, aquesta regió estableix interaccions per pont d'hidrogen que són claus pel reconeixement molecular entre EGF i EGFR. La utilització d'un heptapèptid lineal per mimetitzar aquest domini es traduiria probablement en un elevat cost entròpic per adoptar la conformació bioactiva, degut a la flexibilitat intrínseca que tenen els pèptids lineals.

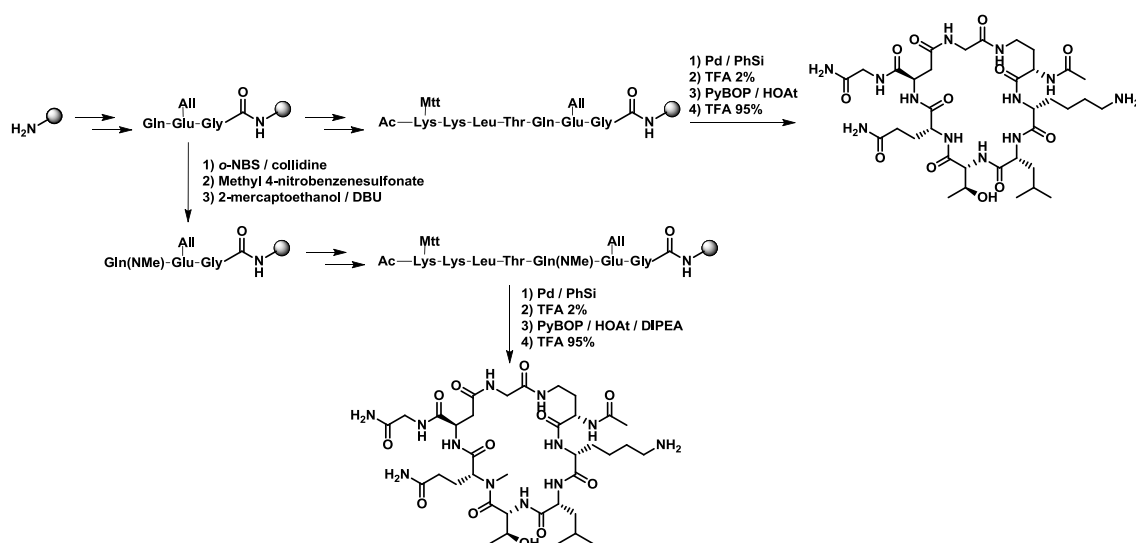
Per tant, vam plantejar una estratègia de ciclació a fi de limitar la flexibilitat conformacional, aprofitant l'estructura quasi-cíclica d'aquest bucle. Per tal de triar la combinació òptima d'aminoàcids per formar el cicle, vam recórrer a tècniques computacionals. Per això, les coordenades d'aquest fragment van ser extretes del PDB 1IVO i els residus d'Asn12 i Leu17 van ser mutats a Lys i Glu (i altres anàlegs amb diferent longitud de cadena lateral), per tal de formar un enllaç amida intramolecular a través d'aquests aminoàcids que feien de nexa de ciclació. En tots els casos, l'esquelet peptídic d'aquests nous anàlegs va ser sotmès a una breu minimització, per tal de resoldre les anomalies estèriques que podien haver sorgit durant la ciclació.

Aprofitant l'eina Macrocylic Conformational Sampling, que és part del programa de modelització Schrodinger, vam dur a terme un mostreig conformacional exhaustiu per tal d'identificar les conformacions de mínima energia per cada pèptid. D'aquesta manera, vam triar les configuracions que mostraven una major similitud amb la conformació bioactiva del receptor, expressada en forma de desviació estàndard mitjana (RMSD) de l'esquelet peptídic (*Taula V*).

Id	Nexe de ciclació	Energia potencial (kcal/mol)¹	RMSD²
Lp1	Asp-Gly-Dab	-1770.8	2.0
Lp2	Dab-Gly-Asp	-1753.1	1.9
Lp3	Dab-Pro-Asp	-1618.0	1.9
Lp4	Dab-Pro-Glu	-1609.0	1.9
Lp5	Dab-DPro-Glu	-1602.6	2.4
Lp6	Dab-Gly-Asp + NMe-Gln	-1602.0	2.4
Lp7	Pro-Dab-Glu	-1592.3	3.1
Lp8	Lys-Glu	-1592.2	1.9
Lp9	DPro-Dab-Glu	-1588.4	3.5
Lp10	Pèptid lineal	-1524.5	2.3
Lp11	Dap-Asp	-1468.2	2.0
Lp12	Dab-DPro-Glu + NMe-Gln	-1457.0	2.3
Lp13	Dab-Pro-Glu + NMe-Gln	-1449.5	2.0
Lp14	Lys-Glu + NMe-Gln	-1433.5	2.0
Lp15	Pro-Dab-Glu + NMe-Gln	-1426.1	3.2
Lp16	DPro-Dab-Glu + NMe-Gln	-1424.2	3.5

Taula V. Llistat de pèptids cíclics samplejats per simulació computacional.¹Mitja de les 5 estructures de mínima energia. ²RMSD de l'esquelet peptídic de la regió interactora (Lys13-Gln16), utilitzant de referència la conformació bioactiva.

La síntesi dels pèptids seleccionats es va realitzar utilitzant grups protectors ortogonals per les cadenes laterals dels residus involucrats en el nexe de ciclació. Aquesta estratègia ens va permetre fer la ciclació en resina, el que evita els problemes de polimerització típicament associats amb la les reaccions de ciclació en dissolució (*Esquema 1*).



Esquema I. Síntesi general dels anàlegs cíclics. En alguns d'ells, es van introduir grups N-metil per tal de rigidificar l'estructura de l'esquelet peptídic, mitjançant una reacció de Mitsunobu (part inferior).

Malgrat els esforços dedicats a l'optimització de la síntesi d'aquests compostos, els assajos d'interacció duts a terme per RMN i SAW van resultar negatius, indicant una manca d'interacció d'aquests pèptids cíclics amb EGF. En conseqüència, vam decidir canviar d'estratègia i intentar mimetitzar regions interactores més grans del receptor.

Disseny de miniproteïnes que reproduïxen un domini de cíclic de l'EGFR

En el domini I del receptor d'EGF, existeix una regió d'estructura cíclica de 28 aminoàcids, caracteritzada per un motiu helicoidal i una regió d'estructura més flexible. En l'estructura de cristall del complex EGF-EGFR, es veu com aquesta part del receptor es troba en proximitat de l'EGF, contribuint al procés d'interacció entre les dos molècules (*Figura X*). Amb l'objectiu d'estudiar l'estructura d'aquest pèptid (que d'ara endavant anomenarem **cp28**) en dissolució, fora del seu entorn al receptor, vam aplicar tècniques computacionals.

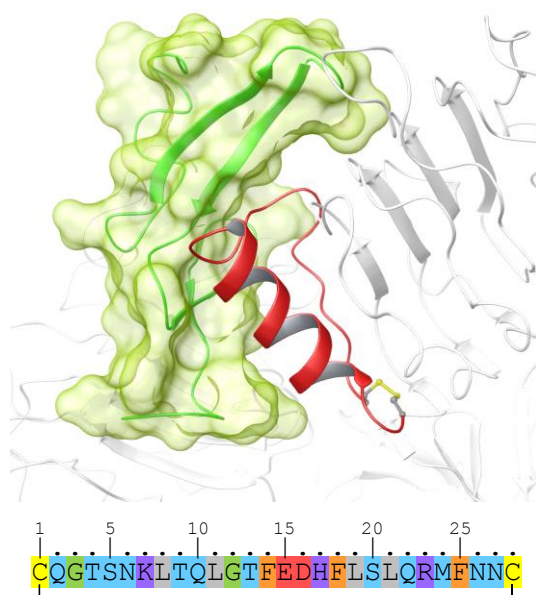


Figura X. Estructura i seqüència de **cp28** (en roig) com a part del domini I de l'EGFR. L'EGF es mostra en superfície verda. Aprecieu l'enllaç disulfur que connecta el primer i últim residu de Cys del pèptid, originant l'estructura cíclica.

En tots els programes de predicció d'estructura secundària que vam consultar, la regió suposadament helicoidal del pèptid mostrava una elevada tendència a adoptar aquest tipus d'estructura secundària. Aquestes prediccions van ser confirmades amb experiments de simulació per intercanvi de rèpliques per dinàmica molecular (REMD), que van mostrar com les conformacions més visitades durant la dinàmica conservaven els trets originals d'aquest domini del receptor.

Encoratjats per aquests resultats preliminars, vam procedir amb la síntesi del pèptid **cp28**, la qual es va portar seguint els mètodes convencionals en fase sòlida, portant a terme la ciclació final en dissolució, sota condicions d'alta dilució. La tendència del pèptid a adoptar un plegament en hèlix α es va confirmar mitjançant experiments de dicromisme circular. A més a més, incubant el pèptid amb la versió isotòpicament marcada de l'EGF, vam observar importants canvis en una àmplia regió de la proteïna (G12, V19, V34, Q43, D46, W50, L52), incloent l'extrem C-terminal flexible de l'EGF, el que inequívocament revelava l'existència d'una interacció (*Figura XI*). Com a estudi addicional, vam calcular l'afinitat de la interacció per SAW, resultant una K_D de 286 μM , un ordre de magnitud més forta que l'obtinguda pel millor dels candidats de *docking* (**d1**).

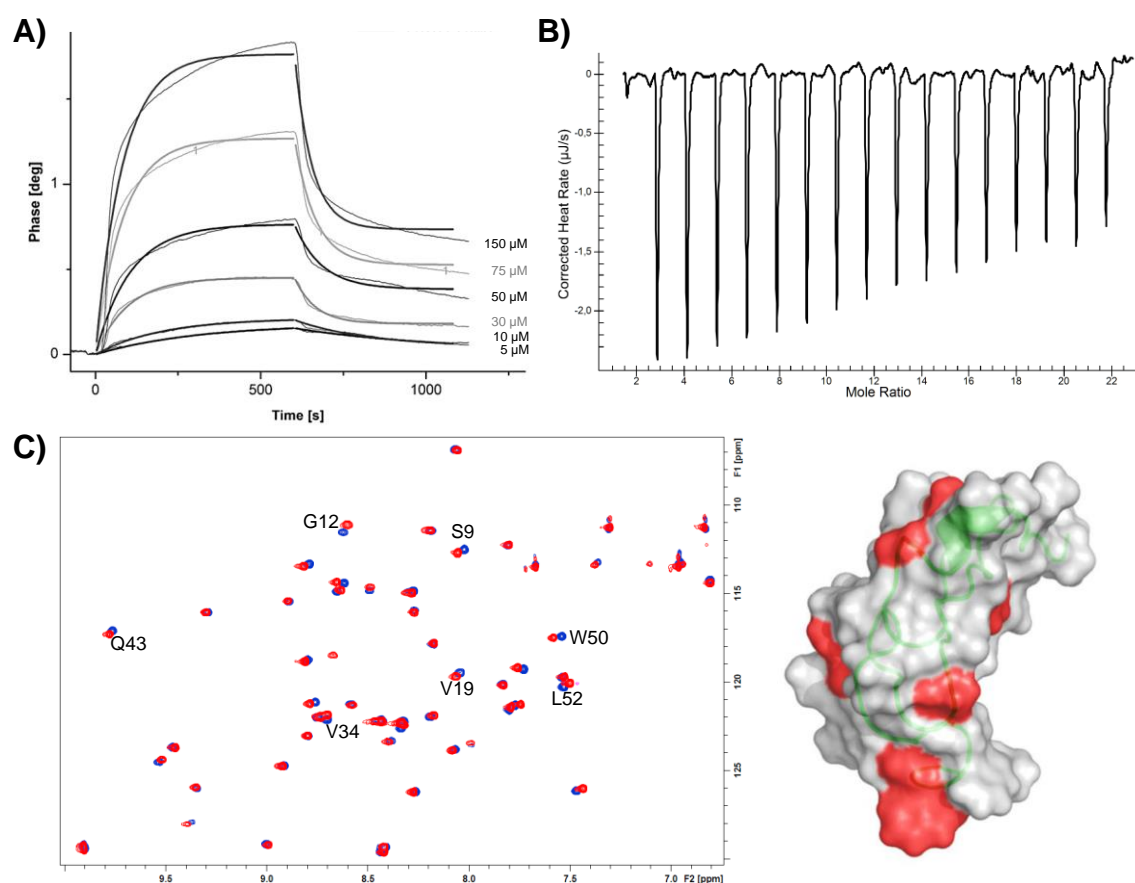
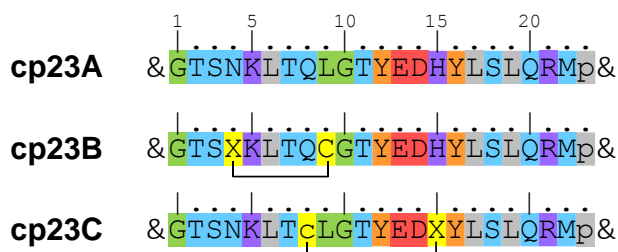


Figura XI. A) Sensograma de SAW de **cp28** injectat a diferents concentracions. B) Experiment de calorimetria de **cp28** amb EGF. C) Espectre de RMN de HSQC ^1H - ^{15}N de l'EGF en absència (blau) i presència (roig) de **cp28** (1.5 mM); els residus amb modificacions significatives es mostren en roig (a la dreta) sobre l'estructura de la proteïna.

Tot i la seva moderada afinitat, el pèptid **cp28** representava el primer cas d'èxit d'un pèptid lligand d'EGF, i ens proporcionava un valuós punt de partida per poder optimitzar el seu perfil. Altres aspectes a millorar de **cp28**

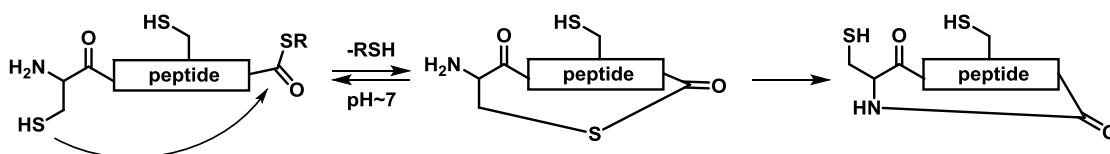
eren la seva baixa solubilitat en medi aquós, així com la seva potencial inestabilitat davant de proteases, com correspon a pèptids formats íntegrament d'aminoàcids naturals. Per tal d'adreçar aquests aspectes i buscar anàlegs més actius, vam començar una estratègia de disseny basada en la millora de l'estructura de **cp28**.

En la primera generació de pèptids anàlegs, vam introduir diferents modificacions estructurals, com ara l'eliminació de les regions no interactores del pèptid, i la seva substitució per un motiu D-Pro-Gly, que força el gir de la cadena peptídica. A més, vam substituir el pont disulfur per un enllaç amida, molt més estable química i biològicament. Finalment, vam eliminar alguns residus hidrofòbics que no contribuïen a la interacció, i vam introduir un segon punt de ciclació, que ajudava a incrementar la rigidesa de l'estructura de la molècula (*Esquema II*).



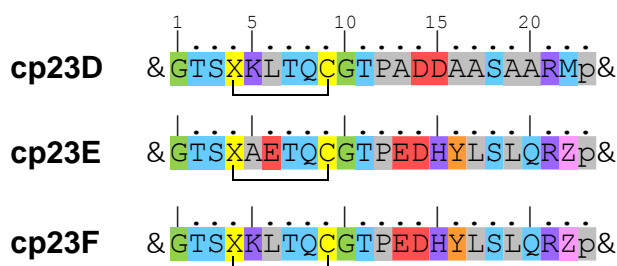
Esquema II. Seqüència dels pèptids anàlegs de primera generació: **cp23A**, **cp23B** i **cp23C** ($X = L$ -homocisteïna).

Per la síntesi d'aquests anàlegs bicíclics, vam aprofitar-nos de les avantatges de la reacció de lligació química nativa (NCL), desenvolupada al laboratori de Kent als anys 90. En aquesta reacció, la cisteïna N-terminal d'un pèptid no protegit reacciona amb un tioèster situat a l'extrem C-terminal del pèptid (si aquest grup funcional està en un altre pèptid, la reacció serà intermolecular en lloc d'intramolecular). Un cop es porta a terme aquesta reacció de transtioesterificació, es produeix un ràpid intercanvi $S \rightarrow N$ que forma l'enllaç amida natiu de l'estructura peptídica, molt més estable termodinàmicament (*Esquema III*).



Esquema III. Reacció de lligació química nativa (NCL); aprecieu que el primer pas és reversible i determina la velocitat de la reacció. Aprecieu també que la presència de tiols lliures situats en altres regions del pèptid no interfereixen amb la reacció.

D'aquest primer joc d'anàlegs, el pèptid **cp23B** va mostrar el perfil més actiu amb una K_D lleugerament inferior a **cp28**, però mantenint el patró d'interacció i l'estructura del pèptid original. Per tal de millorar el perfil d'aquest pèptid, vam generar una àmplia llibreria d'anàlegs i vam aplicar tècniques de simulació computacional per REMD i per càlcul d'energies d'interacció (MM/PBSA), per tal d'identificar els pèptids més actius. En concret, vam veure que la introducció d'un segon residu de Pro, aquest cop a l'extrem C-terminal de l'hèlix α , fixava la conformació en aquesta regió del pèptid i millorava la interacció amb EGF. A més, vam modificar la Met22 del pèptid **cp23B** ja que presentava freqüents problemes d'oxidació a la seva cadena lateral. Aquests canvis van donar lloc al pèptid **cp23F** (*Esquema IV*), que mostrava una afinitat millorada front a EGF ($K_D = 279 \mu\text{M}$) i un millor perfil biofarmacèutic.



Esquema IV. Seqüència dels pèptids bicíclics de segona generació ($X = L$ -homocisteïna; $Z = L$ -norvalina).

Finalment, vam dur a terme un tercer i últim cicle d'optimització computacional d'aquesta estructura bicíclica, explorant una àmplia gama de substitucions sobre l'estructura del pèptid **cp23F**. Aquesta vegada, vam incloure en el nostre repertori aminoàcids D i altres aminoàcids no naturals. En resum, la mutació de Thr2 per D-Arg va resultar en un pèptid molt més semblant a la conformació bioactiva del receptor, mentre que l'energia d'interacció teòrica amb la proteïna (calculada per MM/PBSA) no resultava afectada, o fins i tot lleugerament potenciada. L'espectre de dicroisme d'aquest anàleg (**cp23G**) ens mostrava la solidesa d'aquesta estructura peptídica per tolerar noves modificacions, sense veure's modificada l'estructura global del pèptid (*Figura XII*).

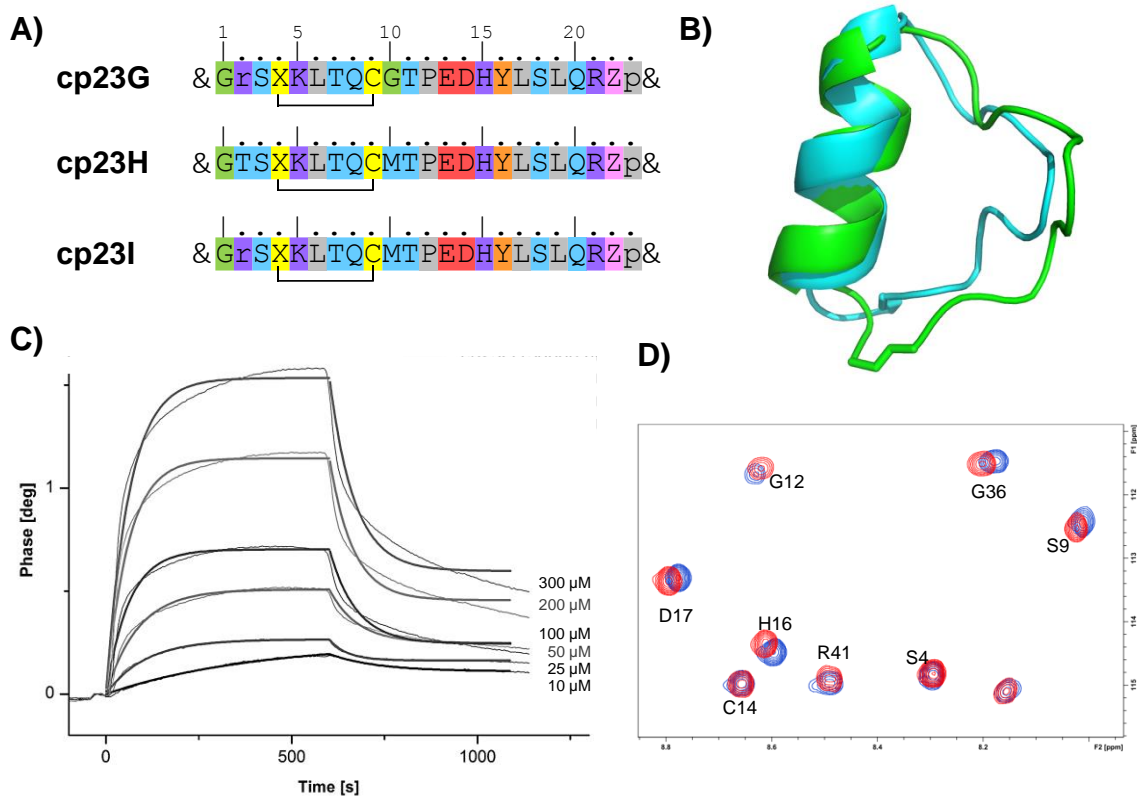


Figura XII. A) Seqüència dels pèptids bicíclics de tercera generació (X = L-homocisteïna; Z = L-norvalina). B) Estructura més poblada durant l'experiment de REMD de **cp23G** (en blau) superposada amb l'estructura bioactiva del pèptid (en verd). Apreciu la similitud entre ambdues estructures. C) Sensograma de SAW de **cp23G** injectat a diferents concentracions. D) Detall de l'espectre de HSQC ¹H-¹⁵N de l'EGF en absència (blau) i presència (roig) de **cp23G** (1 mM).

Com a conclusió d'aquest procés d'optimització del pèptid **cp28**, els anàlegs **cp23B**, **cp23F** i **cp23G** van ser els que van mostrar un millor perfil de tots els pèptids assajats de primera, segona i tercera generació, respectivament, i van ser seleccionats per una avaluació més detallada de la seva activitat *in vitro*, així com de les seves propietats ADME (administració, distribució, metabolisme i eliminació).

Capítol 3. Caracterització *in vitro* de l'activitat i l'estabilitat dels pèptids dissenyats

Per tal de bloquejar l'activació de l'EGFR i produir un efecte biològic, els lligands peptídics que hem dissenyat han de bloquejar la interacció entre EGF i EGFR. Aquesta interacció, tot i donar-se a l'espai extracel·lular, presenta una alta afinitat de l'ordre baix nanomolar i, tot que en teoria és reversible, la unió entre EGF i EGFR és tan forta que la dissociació només te lloc al pH àcid dels endosomes, un cop el receptor s'ha internalitzat a la cèl·lula.

Per estudiar l'acció dels nostres pèptids sobre la interacció EGF-EGFR, vam emprar un assaig *in vitro* basat en la transferència d'energia entre dos petites boles complementàries que estan funcionalitzades amb les proteïnes d'interès. Si aquestes proteïnes formen una interacció, les boles es transfereixen energia i s'emet fluorescència. Aquest assaig d'AlphaScreen te l'avantatge que es pot miniaturitzar en plaques de 96 pouets i utilitza concentracions molt baixes de proteïna (3 nM d'EGF i EGFR, en el nostre cas). En aquest experiment, tot els pèptids assajats (**cp28**, **cp23B**, **cp23F**, **cp23G**) van ser capaços de bloquejar, almenys parcialment, la interacció EGF-EGFR, essent **cp23G** el candidat que va mostrar una major activitat ($IC_{50} = 149 \mu M$).

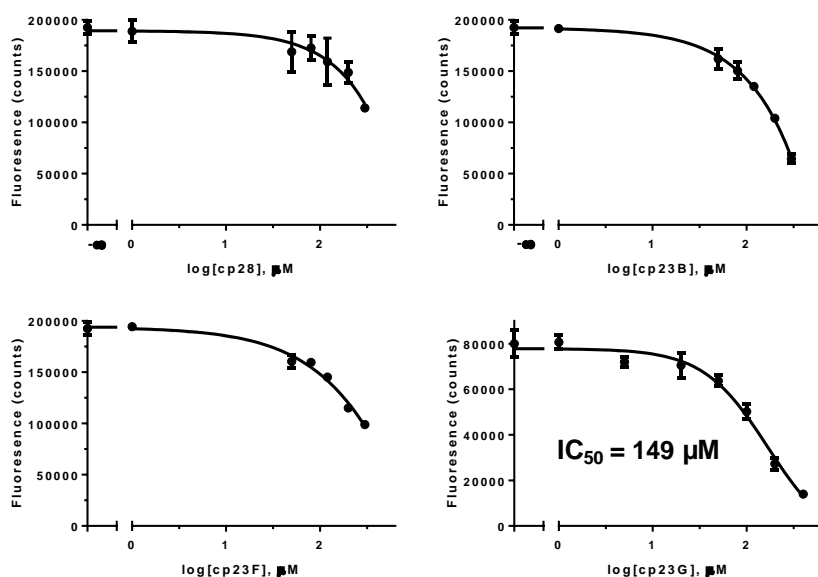


Figura XIII. Assaig d'AlphaScreen efectuat a diferents concentracions de **cp28**, **cp23B**, **cp23F**, **cp23G**. Només per **cp23G** l'ajust de la corba va permetre calcular un valor d' IC_{50} amb precisió.

Per tal d'estudiar aquesta interacció en un context biològic més rellevant, vam posar en marxa un experiment amb cèl·lules tumorals on la finalitat era detectar la quantitat d'EGF unit a les cèl·lules, en absència i presència d'inhibidors. Per quantificar l'EGF unit, vam derivatitzar la proteïna amb AlexaFluor 488, un fluoròfor estable i adient pels assajos in vitro. L'efectivitat de la tècnica va ser verificada incubant cèl·lules tumorals amb diferents concentracions d'EGF marcat, el que ens va permetre calcular una K_D de 12.3 nM, en concordança amb els valors d'aquesta interacció.

Tot i així, la determinació de la fluorescència total de la mostra (106 cèl·lules per tub) comportava una variabilitat molt elevada, degut als errors de manipulació i al fet que no podíem mesurar la fluorescència individual de cada cèl·lula. Com a solució a aquest problema, vam recórrer a l'ús de la citometria en flux per quantificar la fluorescència de cadascuna de les cèl·lules de la mostra. Aquesta tècnica ens va permetre obtenir mesures molt més acurades, i la reversibilitat i especificitat de la unió de l'EGF fluorescent va ser confirmada. Finalment, l'experiment va ser validat fent servir cetuximab, un anticòs anti-EGFR, el qual va donar un valor d' IC_{50} de 5.4 nM. Malgrat les òptimes condicions de l'experiment, l'administració dels nostres pèptids no va comportar canvis significatius en l'EGF unit a cèl·lules. Una possible explicació a aquest fet rau en l'excés d'EGFR d'aquestes línies cel·lulars, en comparació amb les concentracions d'EGFR que anteriorment havíem utilitzat a l'assaig d'AlphaScreen.

De qualsevol manera, vam decidir fer un estudi més general de l'efecte dels nostres pèptids sobre la proliferació de cèl·lules humanes de càncer. En concret, vam utilitzar la línia cel·lular de càncer epitelial A431, que sobreexpresa EGFR i és altament susceptible a l'activació o inhibició d'aquesta via de senyalització. Amb aquestes cèl·lules, vam observar un efecte anti-proliferatiu dels nostres pèptids, més marcat pels pèptids **cp28** i **cp23G**, els quals van obtenir valors d' IC_{50} només un ordre de magnitud per sobre de l'erlotinib (*Figura XIV*). De forma rellevant, aquest efecte inhibitori no es va produir en altres cèl·lules tumorals EGFR(-), el que evidencia l'activitat EGF-específica d'aquests pèptids.

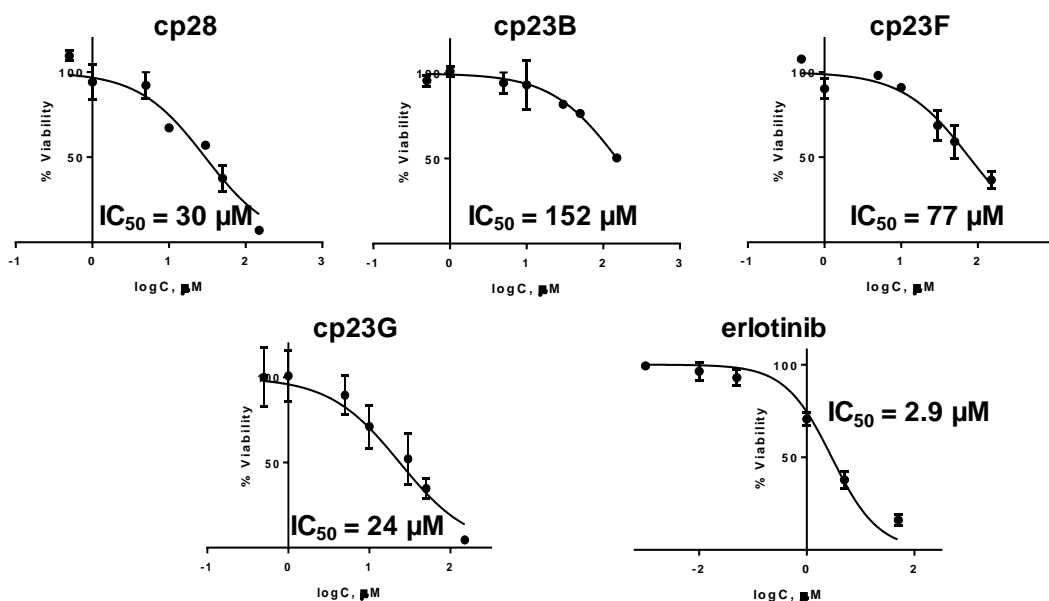


Figura XIV. Viabilitat de les cèl·lules tumorals A-431 incubades amb diferents concentracions de pèptids. Erlotinib va ser utilitzat com a control positiu.

Un dels principals obstacles en el desenvolupament de pèptids terapèutics és la seva manca d'estabilitat en fluids i teixits biològics. Habitualment, els pèptids naturals presenten semivides de minuts en plasma, ja que són ràpidament degradats a aminoàcids i excretats per orina. Per avaluar l'estabilitat dels nostres lligands, vam realitzar un assaig d'incubació dels pèptids en sèrum humà a 37 °C. A intervals de temps definits, vam agafar alíquotes i les vam analitzar. El pèptid **cp28** va presentar una semivida en sèrum de 89 ± 6 min, el que mostra una estabilitat notable per un pèptid natural format exclusivament d'aminoàcids L.

Mitjançant una anàlisi per espectrometria de masses dels principals metabòlits de **cp28**, vam observar que el pèptid era hidrolitzat enzimàticament a la regió de Arg23-Met24, donant lloc a productes de menor pes molecular. Afortunadament, en els anàlegs bicíclics aquesta regió es troba protegida de l'atac enzimàtic per la incorporació d'un aminoàcid no natural (la norvalina) i d'un D-aminoàcid (la D-Pro), el que impedeix el reconeixement i la hidròlisi per les proteases sèriques.

Tal com estava esperat, els anàlegs bicíclics **cp23B-G** van presentar una gran estabilitat en sèrum. A les 8 hores d'incubació, pràcticament el 100% del pèptid restava intacte en sèrum, i fins i tot després de 24 hores aquest percentatge era superior al 60% (*Figura XV*). Aquest experiment posa de manifest la gran bioestabilitat de les nostres estructures bicíclics, en comparació amb el pèptid natural de partida.

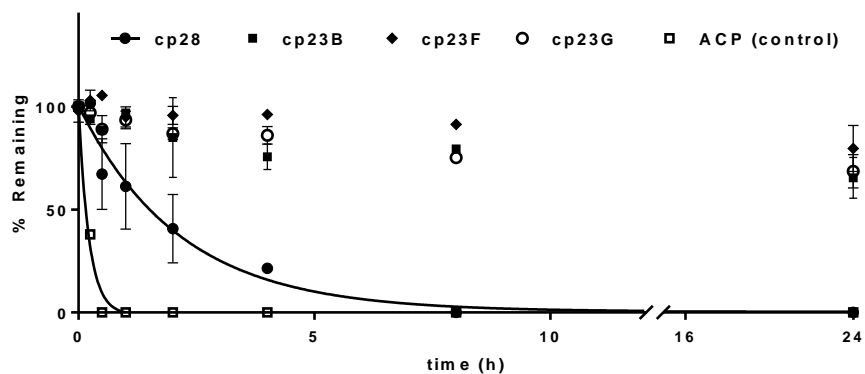


Figura XV. Estabilitat dels pèptids en sèrum humà a 37 °C. El pèptid lineal ACP es va utilitzar com a control.

L'estabilitat d'aquests pèptids també va ser estudiada en un mitjà biològic molt més agressiu, com és el fetge. Aquest òrgan compta amb una gran diversitat de proteases, així com altres tipus d'enzims oxidatius com el citocrom P450. Generalment, el metabolisme hepàtic es produeix sobre molècules apolars per tal d'incrementar la seva hidrofilitat i afavorir la seva excreció, tot i que la informació sobre el metabolisme hepàtic de pèptids és reduïda. En aquest context, tots tres anàlegs **cp23B-G** van presentar un metabolisme hepàtic molt inferior al pèptid **cp28**, el que es podria explicar per la hidrofobicitat més marcada d'aquest darrer.

Capítol 4. Disseny i avaluació de pèptids inhibidors covalents de la POP

Com hem esmentat anteriorment, els inhibidors peptídics competeixen amb un substrat endogen per la seva unió a una proteïna. Una característica típica dels pèptids és que presenten constants de dissociació molt ràpides, fet que resulta desitjable per la modulació d'interaccions transitòries com a les PPI, però que pot resultar un inconvenient major per la inhibició de centres d'unió a receptors i enzims. Aquest dilema s'ha resolt parcialment en l'àmbit de les molècules orgàniques, mitjançant el desenvolupament d'inhibidors covalents, però ha estat molt poc explorat en el camp de la química mèdica de pèptids.

Com a prova de concepte, ens vam plantejar de dissenyar una nova família d'inhibidors de la prolil-oligopeptidasa (POP), un enzim que diferents estudis han implicat amb els trastorns cognitius. Per dissenyar aquests compostos, vam partir de l'estructura canònica del Z-prolilprolinal, així com de la gran quantitat d'estudis estructura-activitat que en els darrers anys s'han fet sobre nous inhibidors de la POP. En les nostres molècules, vam explorar tres grups aromàtics diferents en la posició P3 de l'inhibidor, els quals han mostrat una gran activitat inhibidòria. En P2, vam mantindre l'anell de Pro que és essencial per l'activitat, tot i que vam introduir diferents substituents polars a la posició 4 de l'anell, ja que aquesta modificació podia repercutir en una major permeabilitat. En posició P1, vam incorporar un grup electròfil de fluorur de prolilsulfonil, destinat a inhibir irreversiblement l'activitat de la POP (*Figura XVI*).

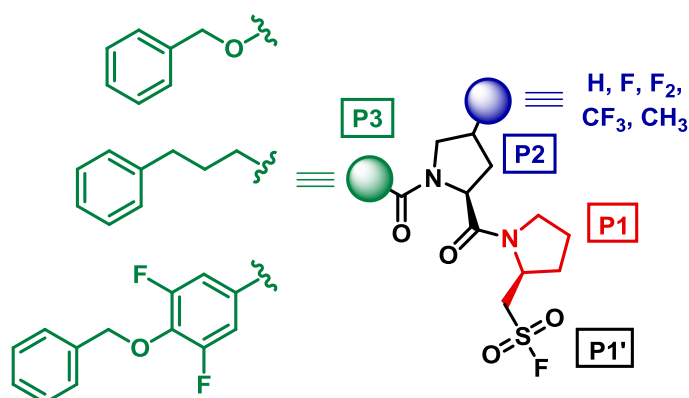


Figura XVI. Estructura general dels inhibidors covalents de la POP. Diferents grups químics van ser combinats en P3 i P2, el que va generar un total de 15 inhibidors potencials.

La determinació de l'activitat enzimàtica de la POP es va realitzar mitjançant un assaig fluorimètric, que mesura la incorporació d'un substrat fluorescent. Per poder comparar tota la família d'inhibidors, vam utilitzar les mateixes concentracions de substrat i el mateix temps d'incubació a totes les mesures. Tal com esperàvem pel disseny racional, els nostres compostos van produir un efecte molt potent sobre l'activitat de la POP, amb tots els candidats aconseguint valors d'IC₅₀ nanomolars (*Taula VI*). Fins i tot alguns compostos, com **11** i **18**, van arribar a potències subnanomolars.

Des d'un punt de vista cinètic, la inhibició per part d'un fàrmac covalent té lloc en almenys 2 etapes. En primer lloc, el procés de reconeixement molecular entre l'enzim i l'inhibidor, el qual és reversible i està governat per interaccions no covalents. Seguidament, l'atac nucleòfil per la proteïna a l'inhibidor, el que dona pas a la formació del complex covalent E-I. Si la inhibició és irreversible, aquest complex serà definitiu fins la degradació de la proteïna. Donat que la determinació cinètica és un procés complex i tediós, aquest experiment el vam efectuar només per l'inhibidor **18**, el més actiu de la sèrie. En aquest cas, el rati k_2/K_i , que es considera el paràmetre que millor representa la potència d'un inhibidor, va resultar de $2 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$, el que indica una activació ràpida i definitiva de l'enzim (*Figura XVII*).

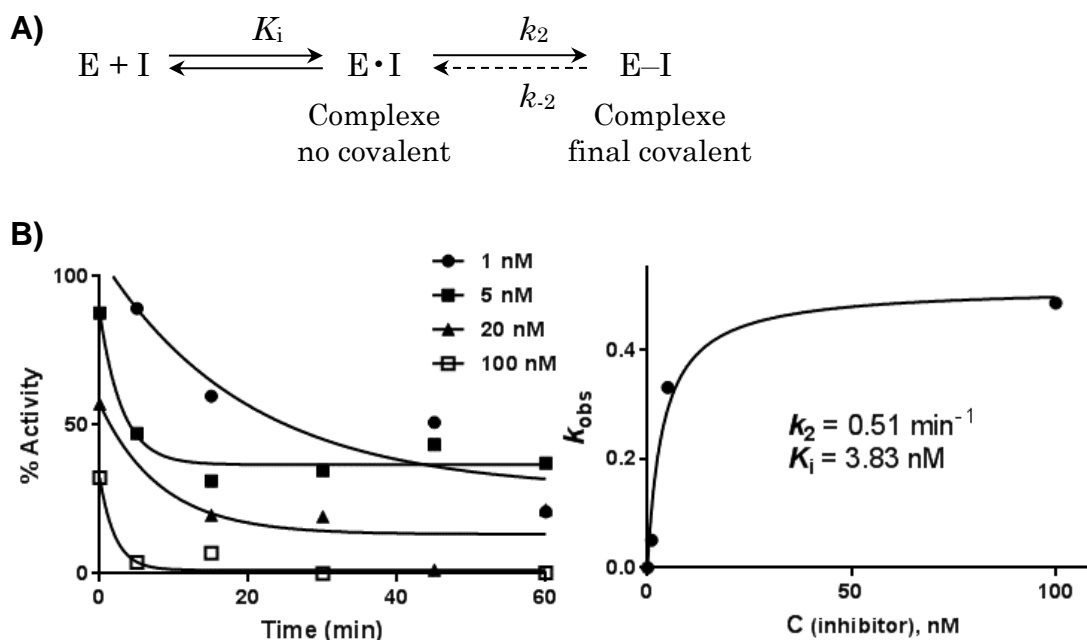
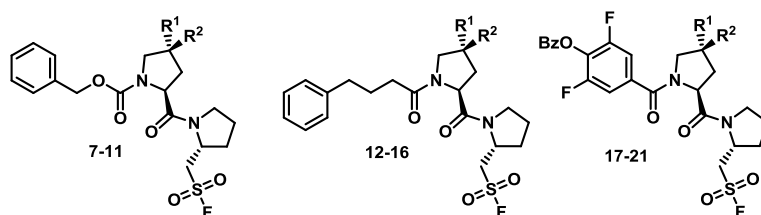


Figura XVII. A) Equilibri i constants cinètiques que representen la inhibició d'un enzim (E) per un fàrmac covalent (I). B) Caracterització cinètica de l'inhibidor **18**.

De les ~400 proteases humanes que es coneixen, unes poques són capaces d'hidrolitzar pèptids i proteïnes que contenen Pro. Aquestes proteases, com la POP, la FAP i la DPP-IV, s'agrupen a la família S9 de proteases, i presenten una gran similitud tant des del punt de vista estructural com funcional. En aquest projecte, vam estudiar la selectivitat dels nostres millors inhibidors (**8**, **13** i **18**) davant de la FAP i la DPP-IV, per veure quin grau de discriminació molecular podien assolir. Remarcablement, fins i tot a la concentració més alta assetjada 50 μ M (~50.000 cops la IC₅₀ de POP), la inhibició de DPP-IV i FAP es va mantindre per sota d'un 25%. A concentracions més baixes, la capacitat d'aquests compostos per distingir entre aquestes tres prolilpeptidases va ser excepcional, complint un dels principals criteris en el desenvolupament d'inhibidors selectius de la POP.

Finalment, es van avaluar les propietats d'aquests compostos per arribar al seu lloc d'acció, el cervell. El cervell està protegit de la circulació sistèmica per la barrera hematoencefàlica, una membrana altament selectiva que regula l'entrada de substàncies al sistema nerviós central. En la practica ~100% dels fàrmacs biològics (com anticossos i proteïnes) i el 98% dels fàrmacs convencionals no aconsegueixen travessar aquesta barrera. Per estudiar aquest sistema, vam utilitzar l'assaig PAMPA, que recrea les condicions de la barrera hematoencefàlica d'una manera simple i reproducible.

En aquest assaig, el transport a través d'aquests lípids cerebrals va ser notable per tots els anàlegs assajats, de l'ordre del propranolol, un fàrmac β -bloquejant que s'utilitza en aquests assajos com a control positiu. L'addició de fluor a l'anell de prolina P2 va produir un increment de la permeabilitat, mentre que els grups excessivament apolars van causar una retenció del compost a la membrana i una lleugera disminució del transport (*Taula VI*).



Id	R ¹	R ²	IC ₅₀ POP (nM)	Transport (%)	Retenció (%)	P _e (10 ⁻⁶ cm s ⁻¹)
7	H	H	1.3	19.1	37.1	11.4
8	F	H	4.5	18.8	19.7	11.2
9	F	F	21.2	24.8	2.1	16.3
10	H	CF ₃	7.9	14.0	40.7	7.8
11	H	CH ₃	0.8	23.9	14.2	15.4
12	H	H	3.2	21.5	29.3	13.3
13	F	H	1.9	20.2	23.6	12.2
14	F	F	2.3	25.4	20.4	16.8
15	H	CF ₃	2.1	21.5	51.2	13.3
16	H	CH ₃	1.9	20.3	57.8	12.4
17	H	H	1.8	25.4	26.7	17.2
18	F	H	0.9	26.5	34.0	17.9
19	F	F	1.7	15.3	62.3	8.7
21	H	CH ₃	1.2	19.6	51.4	11.8

Taula VI. Resum de la potència inhibidora front a POP i dels paràmetres de permeabilitat (obtinguts a l'assaig PAMPA) dels inhibidors dissenyats.

En conclusió, aquests inhibidors covalents van ser capaços de bloquejar l'activitat enzimàtica de la POP a concentracions nanomolars baixes, al temps que van mostrar una notable selectivitat sobre altres enzims relacionats. A més, les molècules dissenyades van aconseguir una notable permeabilitat a través d'un model de barrera hematoencefàlica, tot plegat complint els principals requisits per esdevenir potencials candidats a fàrmac.

Conclusions

D'acord als resultats presentats, les conclusions d'aquesta tesi són les següents:

1. Utilitzant tècniques computacionals de *docking*, hem aconseguit modelitzar una sèrie de pèptids petits capaços d'unir-se a EGF. No obstant, les tècniques actuals de *docking* s'han manifestat insuficients per predir acuradament la interacció amb una proteïna tan flexible com EGF. Addicionalment, s'han explorat diverses tècniques biofísiques per estudiar les interaccions dels lligands amb EGF, de les quals la ressonància magnètica nuclear i el biosensor SAW han aconseguit els millors resultats
2. Inspirats en l'estructura del receptor, s'ha dissenyat una família de pèptids que mimetitza diferents regions interactores del EGFR. Aquests pèptids, d'una mida més gran i amb una estructura tridimensional definida, han aconseguit millors resultats que el descrits per *docking*, quant a la seva habilitat per interaccionar amb EGF. El pèptid **cp23G** presenta una estructura bicíclica i incorpora aminoàcids no naturals, i ha resultat l'anàleg més actiu, junt amb el pèptid **cp28**, en els assajos biofísics.
3. A més d'unir-se a EGF, els pèptids dissenyats en els apartats anteriors han de trencar la interacció EGF-EGFR, per tal de produir un efecte biològic. Al respecte, hem estudiat aquesta capacitat en assajos biològics amb cèl·lules i en assajos bioquímics. En cèl·lules tumorals, els millors candidats han estat capaços de disminuir la proliferació fins a un 80%. A més, el pèptid **cp23G** va mostrar una excepcional resistència a la degradació per part de proteases sèriques, així com en microsomes hepàtics.
4. Per tal d'inhibir una altra proteïna amb interès terapèutic (la proliloligopeptidasa, o POP) s'han dissenyat una sèrie de compostos peptidomimètics amb mecanisme d'acció covalent. Aquests composts són capaços d'inactivar la POP a concentracions nanomolars, a la vegada que són selectius per aquest enzim. A més, aconsegueixen valors elevats de permeabilitat en un assaig de difusió passiva a través de lípids cerebrals, el que posa de manifest el seu potencial com a fàrmacs.