UNIVERSITAT DE BARCELONA

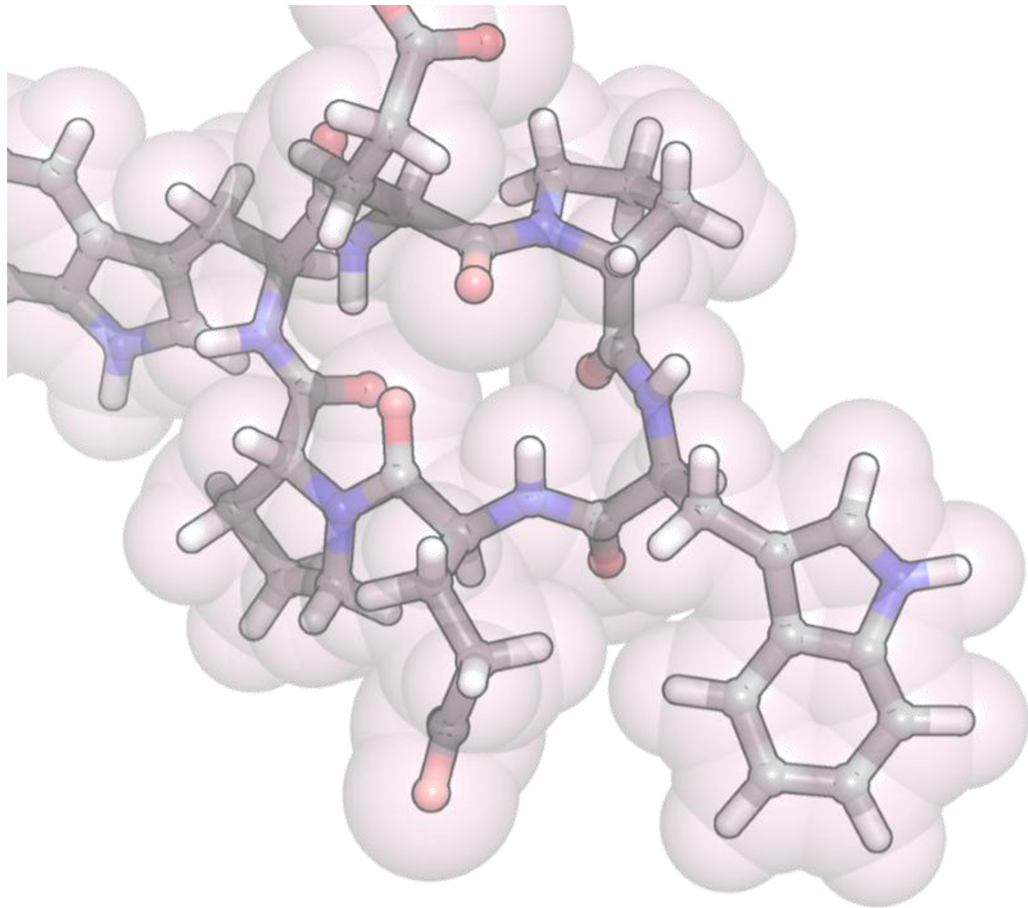
Cyclic peptides and small proteins in molecular recognition

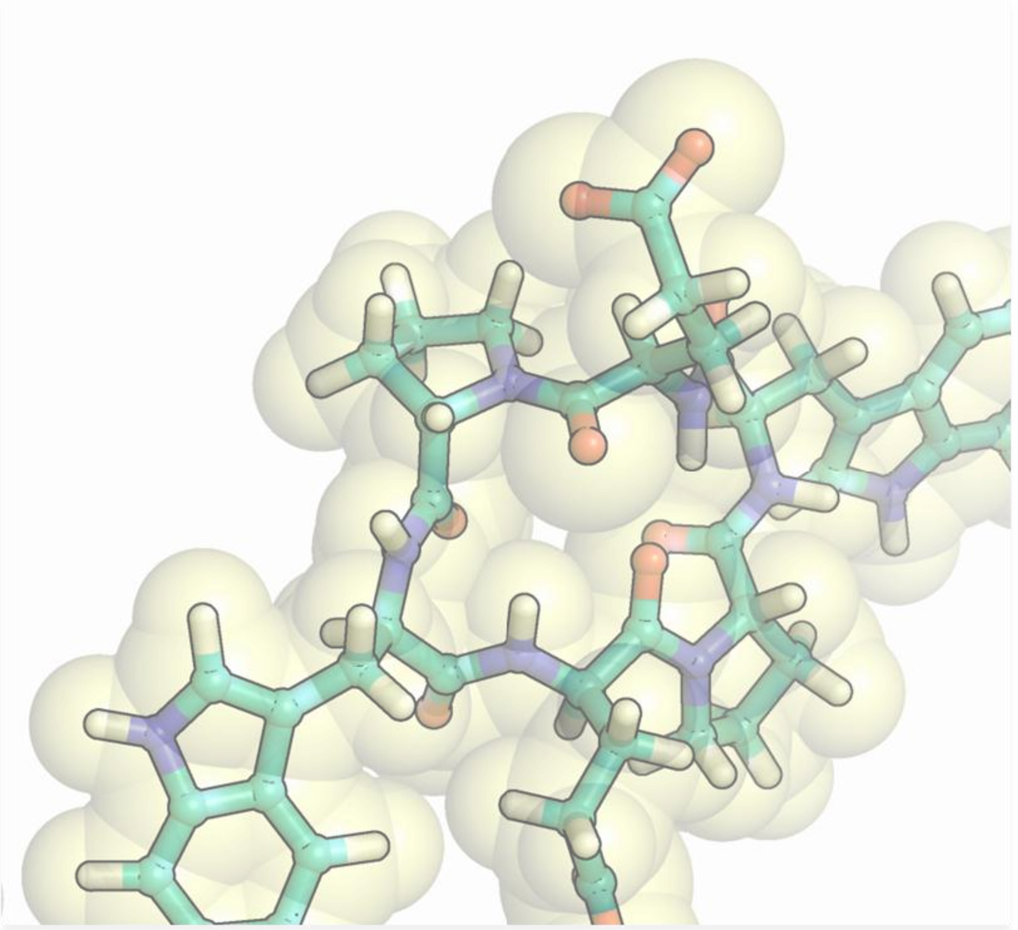
Sonia Ciudad Fernández

Aquesta tesi doctoral està subjecta a la llicència **_Reconeixement- NoComercial – SenseObraDerivada 3.0. Espanya de Creative Commons_**.

Esta tesis doctoral está sujeta a la licencia **_Reconocimiento - NoComercial – SinObraDerivada 3.0. España de Creative Commons._**

This doctoral thesis is licensed under the **_Creative Commons Attribution-NonCommercial-NoDerivs 3.0. Spain License._**

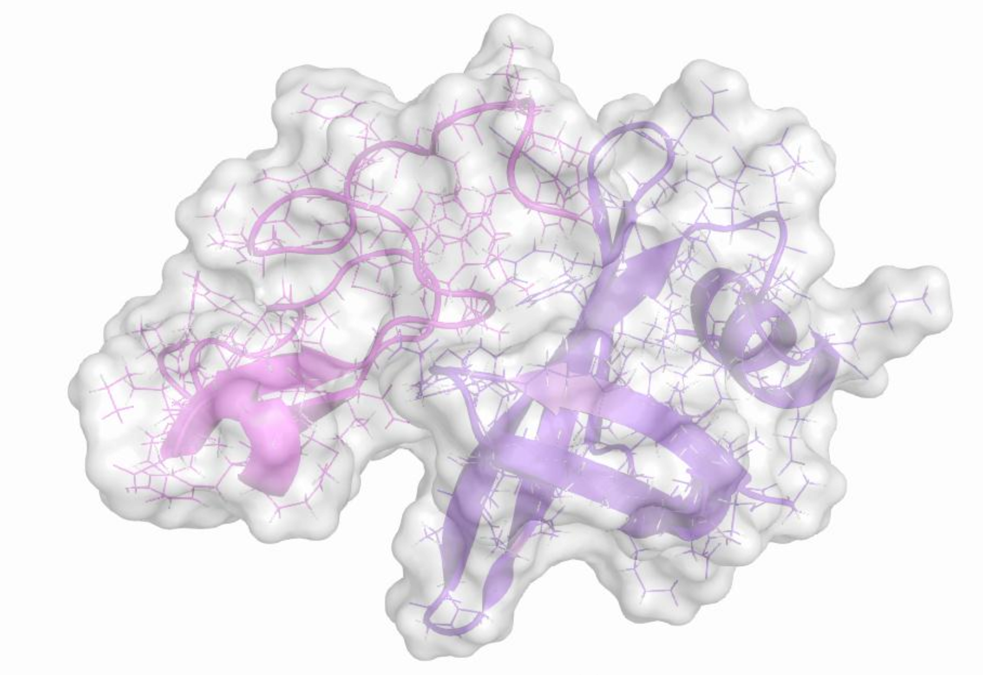




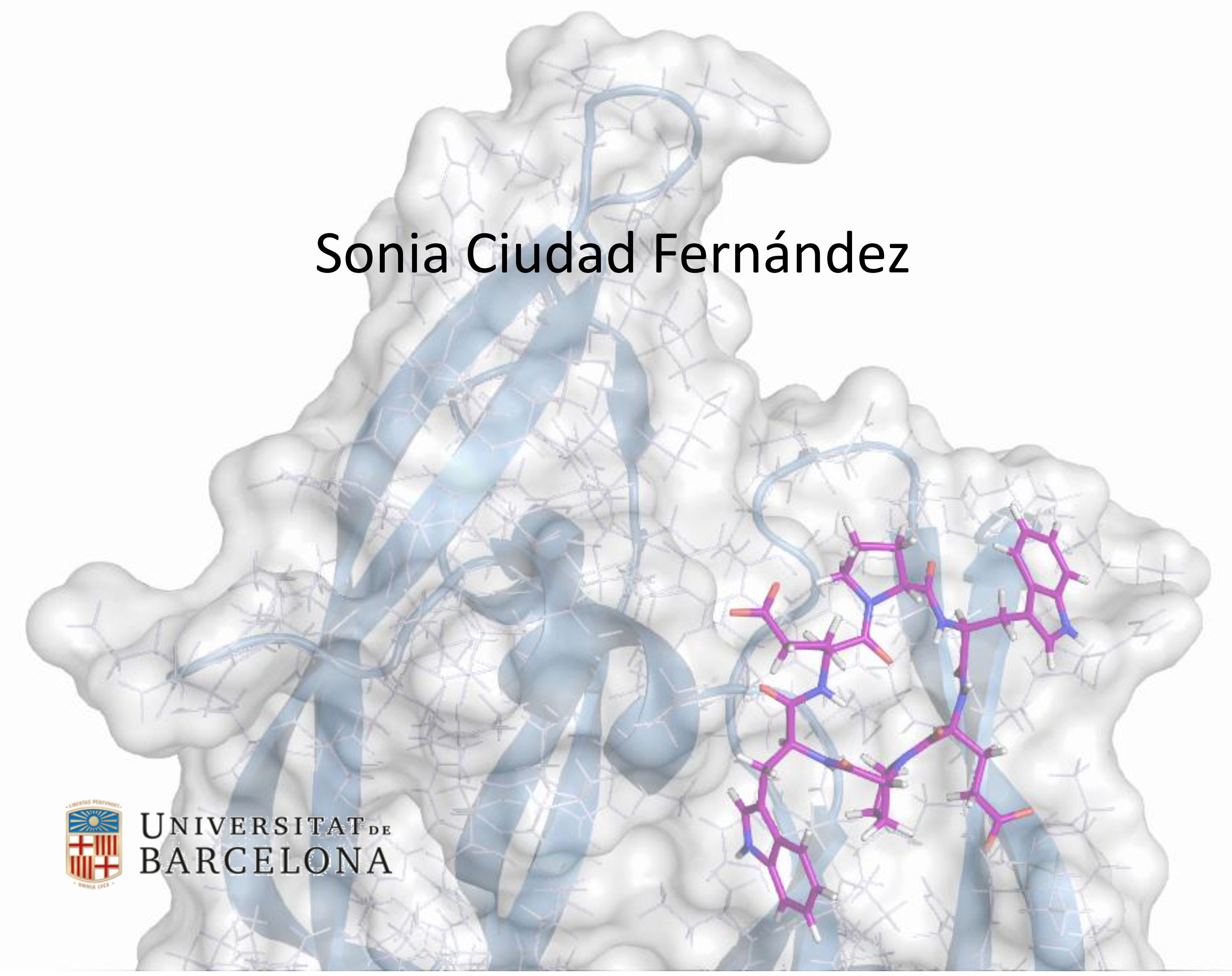
Sonia Ciudad Fernández

Tesi doctoral

2016



Cyclic peptides and small proteins in molecular recognition



Sonia Ciudad Fernández

UNIVERSITAT DE BARCELONA

Programa de doctorat de química orgànica

Cyclic peptides and small proteins in molecular recognition

Sonia Ciudad Fernández

Tesi doctoral dirigida per:

Prof. Ernest Giralt Lledó
Universitat de Barcelona
Facultat de Química
Departament de Química Orgànica

Dr. Jesús García Arroyo
IRB Barcelona
Programa de química i
farmacologia molecular

Barcelona, 2016.

A mis padres, a mi hermana y a mis abuelos

Acknowledgements

Realizar una tesis es como montarse en una montaña rusa con sus altos y bajos, por eso es muy importante rodearse de personas que te ayuden a superar los instantes difíciles y con quienes poder compartir los buenos momentos. Es por ello que quiero aprovechar para dar las gracias a toda la gente que ha estado a mi lado de un modo u otro para ayudarme siempre a lo largo de estos cuatro años y sin los cuales este trabajo no habría sido posible.

Primero quiero agradecer a mis directores de tesis: a **Ernest Giralt** por confiar en mí desde el principio, por darme la oportunidad de realizar esta tesis en tu laboratorio y por adentrarme en el mundo de la RMN de proteínas. A **Jesús García** por estar siempre dispuesto a ayudarme, por tu enorme paciencia a la hora de enseñarme los entresijos de la RMN y por los buenos consejos que me has dado a lo largo de estos años.

Estoy muy agradecida a **la Caixa**, por otorgarme una beca para poder llevar a cabo la investigación de esta tesis y por proporcionarme los medios necesarios para poder asistir a cursos y congresos, de los cuales he podido aprender muchas cosas.

Quiero agradecer muy especialmente a **Michael** por ser mi mentor durante mis primeros meses en el IRB y además un gran amigo. También dar las gracias a los miembros de mi comisión de seguimiento: **Miquel Pons**, **María J. Macías** y **Albert Virgili**; por sus buenos consejos y por guiarme siempre que lo he necesitado. Y agradecer también a **Jordi Redondo** por confiar siempre en mí y por todas las cartas de recomendación que me has escrito durante este tiempo.

A toda la gente del lab con los que he tenido el placer de compartir estos cuatro años. Al BBB Team: **Txell**, **Bernat G.**, **Cristina F.**, **Pep**, **Cristina D.**, **Pol**, **Benjamí y Xavi**; a toda la gente del antiguo Pop Team y ahora IPROTEOS: **Tere**, **Abraham**, **Jesús S.**, **Laura M.**, **Albert**, **Roger**, **Marta**, **Núria** y **Sandra**; a la gente del Amyloid Team: **Martí**, **Sílvia**, **Bernat S.**, **Roberta, Antonela y Eduard;** y en especial a **Natàlia**, por tus grandes consejos y tu incansable ayuda, y por supuesto por confiar en mi para emprender una nueva aventura. Y como olvidarme de mis compañeros del proyecto PPIs: **Salva**, **Júlia**, **Ester** y **Mar**. También agradecer a nuestros postdocs: **Macarena**, **Daniele**, **Jesús S.**, **Monica**, **Montserrat**, **Aurelio y Núria**; y especialmente a **Laura N**. por estar siempre ahí cuando la he necesitado y por ese viajecito a Hyères que nunca vamos a olvidar. También a los nuevos compañeros del lab: **Toni**, **Mark** y **Daniela** con los que, a pesar de no haber tenido la oportunidad de pasar mucho tiempo, he podido compartir algunos buenos momentos y charlas amenas.

Agradecer enormemente a nuestra secretaria **Eva Poca** por su amabilidad y su ayuda incondicional con todo el papeleo y temas administrativos. Por supuesto, no me olvido de nuestras dos técnicas de laboratorio: **Cristina G.** y **Esther Z.** que han hecho lo posible por ayudarnos con todas las comandas y demás temas relacionados. No podría olvidarme tampoco de toda la gente del lab 100 y Pharmamar con los que también he compartido grandes instantes, en especial con **Helena** y **Judit** y por supuesto también a **Iván** por ser mi compañero de aventuras salseras y por todos los cafés que hemos tomado juntos durante estos años y que tanto nos han ayudado a desconectar.

Durante este tiempo, he tenido la enorme suerte de conocer personas con quienes he llegado a entablar una gran amistad y con los que he compartido momentos increíbles dentro y fuera del laboratorio. Quiero agradecer enormemente a **Pol** y **Jesús S**. por vuestra ayuda incondicional y por estar siempre cerca y apoyarme cuando más lo he necesitado, y por supuesto a mis niñas: **Mar**, **Ester** y **Antonela**, porque a pesar de haber coincidido poco tiempo en el laboratorio nuestra amistad ha continuado fuera y espero poder seguir compartiendo muy buenos ratos con vosotras. Agradecer también a todos mis compañeros de desayunos y cafés: **Salva**, **Martí**, **Cristina F**., **Cristina D**., **Antonela**, **Mar**, **Ester**, **Jesús S**., **Pol** e **Iván**; por ayudarme a empezar cada día repleto de energía, por todas las risas que nos hemos echado juntos y por esas barbacoas y escapadas inolvidables.

Quiero también acordarme de todos mis compañeros del **IRB PhD Symposium ’15**, con quienes compartí la emocionante experiencia de organizar un evento internacional. Aprovecho también para dar las gracias a **Patricia**, **Leyre** y **Alba** por vuestra ayuda a la hora de enfrentarnos a este reto y por darme soporte con todos los temas administrativos. También quiero agradecer en especial a **Konstanin** y **Enrique**, con los que empecé esta aventura allá en el 2012 y con quienes siempre he contado a lo largo de estos años.

Por supuesto no podría olvidarme de toda la gente de las diferentes "facilities" del IRB que hacen siempre nuestra vida mucho más fácil. Al grupo de Espectroscopia de Masas, especialmente a **Marta** y **Mireia**; a Expresión de Proteínas: **Nick**, **Raquel** y **Queralt**; también a la gente de los servicios del CCiT, especialmente a **Vicky**. Así como de toda la gente de los Servicios de RMN, especialmente de **Marga**, a quien quiero agradecer su tremenda amabilidad y su infinita paciencia al enseñarme y ayudarme con todo lo relacionado con la RMN.

I want to thank all the people with whom I shared my stay at MIT very much for making me feel like home during my four month's experience in the USA. First of all I want to thank **Prof. Barbara Imperiali** for giving me the opportunity to join her lab and for making me feel really comfortable. Thanks to **Elisabeth** also, for all her help and kindness. To all my lab colleagues for helping always: **Debasis**, **Sonya**, **Amael**, **Cristy**, **Joris**, **Julie** and **Jean Marie**. And of course, I want to specially thank **Silvano** very much for all the moments that we spent together during these amazing four months and for all your patience and confidence in teaching me.

Quiero agradecer muy especialmente también a todos mis amigos (**Mariona**, **Mar**, **Lluís**, **Marc**, **David**, **Anna**, **Maria C**. y **Maria R.**) por vuestra paciencia y por seguir estando siempre a mi lado aunque no tuviéramos mucho tiempo para vernos.

Finalmente, pero no por ello menos importante, quiero dar las gracias enormemente a toda mi familia, especialmente a mis padres (**Aurelio** y **Magdalena**), a mi hermana (**Cristina**) y mi cuñado (**Edu**); por apoyarme siempre incondicionalmente, y porque sin ellos nunca habría llegado hasta aquí. Y por supuesto agradecer infinitamente a mis abuelos (**Aurelio, José, Claudia y Josefa**) por animarme siempre a seguir adelante y a no rendirme nunca.

“Do not ever save anything for a special occasion.
Every day you are alive is a special occasion.”
-Ann Wells-

“It always seems impossible until it is done”
-Nelson Mandela-

CONTENTS

ABBREVIATIONS

Aa	amino acid
ACH	α-cyano-4-hydroxycinnamic acid
ACN	acotonitrile
AcOH	acetic acid
API	active pharmaceutical ingredient
BBB	blood brain barrier
BME	2-mercaptoethanol
BODIPY	boron-dipyrromethene
Boc	tert-butyloxycarbonyl
CD	circular dichroism
CNS	central nervous system
COSY	correlation spectroscopy
CSP	chemical shift perturbation
DCM	dichloromethane
DIEA	N,N-diisopropylethylamine
DMF	dimethylformamide
4-DMN	4-N,N-Dimethylamino-1,8- naphthalimide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DSS	sodium-3-(trimethylsilyl)propanesulfonate
DTT	D,L-dithiothreitol
EC	endothelial cell
E.coli	Escherichia coli
EDTA	ethylenedianminetetraacetic acid
EGF	endothelial growth factor
EGFR	endothelial growth factor receptor
eq.	equivalent
ESI	electrospray ionization
FA	formic acid
FACS	fluorescence activating cell sorting
FBLD	fragment-based ligand discovery
FDA	food and drug administration
Fmoc	9-fluoroenylmethoxycarbonyl
FP	fluorescence polarization

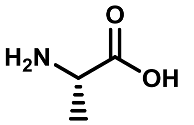
FPLC	fast protein liquid chromatography
FRET	Förster resonance energy transfer
GSH	glutathione
HATU	O-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate
HBS	hydrogen bond surrogate
HEPES	4-(2-hydroxyethyl)-piperazine-1-ethanesulfonic acid
hEGF	human epidermal growth factor
HOAt	1-hydroxy-7-azabenzotriazole
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence spectroscopy
HTS	high throughput screening
Ig	immunoglobulin
IMAC	immobilized metal ion affinity chromatography
IPTG	Isopropil-β-D-1-tiogalactopiranósido
ITC	isothermal titration calorimetry
LB	lysogeny broth
mAb	monoclonal antibody
MALDI	matrix assisted laser desorption ionization
MeCN	acetonitrile
MeOH	methanol
MIT	Massachusetts Institute of Technology
MOE	molecular operating environment
mRNA	messenger ribonucleic acid
MS	mass spectroscopy
MSA	mouse serum albumin
MST	microscale thermophoresis
MW	molecular weight
MWCO	molecular weight cut-off
m/z	mass over charge
NBD	nitrobenzoxadiaxole
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
OD	optical density

PAGE	polyacrylamide gel electrophoresis
PAMPA	parallel artificial membrane permeability assay
PBS	phosphate buffer saline
PDA	photodiode array
PDB	protein data bank
PDGFR	platelet-derived growth factor receptor
pI	isoelectric point
PlGF	placental growth factor
PPI	protein-protein interaction
ppm	parts per million
PyAOP	7-Azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
REMD	replica exchange molecular dynamics
RCM	ring closing metathesis
RMSD	Root-mean-square deviation
RNA	ribonucleic acid
RP	reverse phase
RT	room temperature
SD	standard deviation
SDS	sodium dodecylsulfate
SEC	size exclusion chromatography
SPPS	solid-phase peptide synthesis
SrtA	sortase A
SUMO	small-ubiquitin like modifier
STD	saturation transfer difference
SW	sweep width
TB	terrific broth
tBu	tert-butyl
TCA	trichloroacetic acid
TCEP	Tris-(2-carboxyethyl)phosphine
TEA	triethylamine
TFA	trifluoroacetic acid
TFE	trifluoroethanol
TIS	triisopropylsilane
TKR	tyrosine kinase receptor

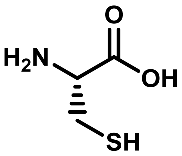
TOCSY	total correlation spectroscopy
TOF	time-of-flight
Trt	trityl
UPLC	ultra high performance liquid chromatography
UV	ultraviolet
VEGF	vascular endothelial growth factor
VEGFR	vascular endothelial growth factor receptor
WT	wild type
YSD	yeast surface display

Proteinogenic amino acids

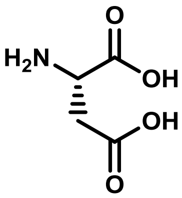
A
Ala
L-Alanine



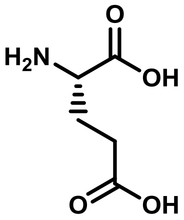
C
Cys
L-Cysteine



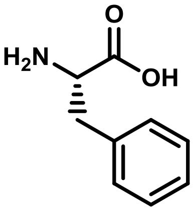
D
Asp
L-Aspartate



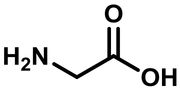
E
Glu
L-Glutamate



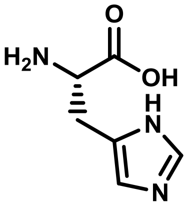
F
Phe
L-Phenylanine



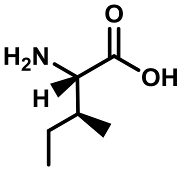
G
Gly
Glicine



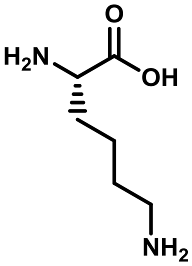
H
His
Histidine



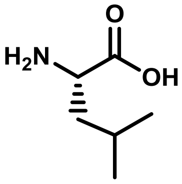
I
Ile
Isoleucine



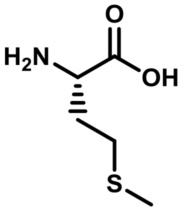
K
Lys
Lysine



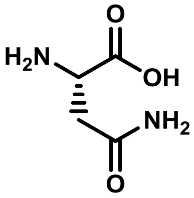
L
Leu
Leucine



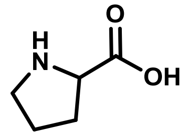
M
Met
Methionine



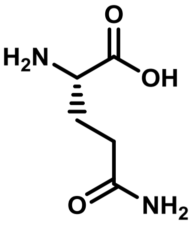
N
Asn
Asparagine



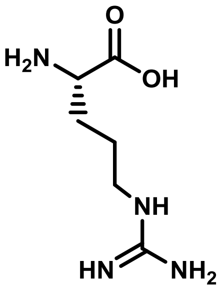
P
Pro
Proline



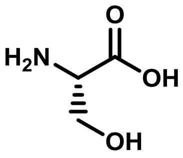
Q
Gln
Glutamine



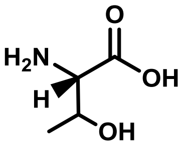
R
Arg
Arginine



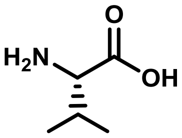
S
Ser
Serine



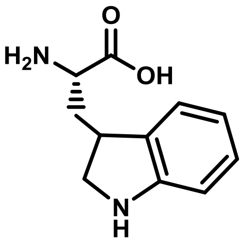
T
Thr
Threonine



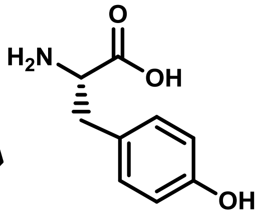
V
Val
Valine



W
Trp
Tryptophan

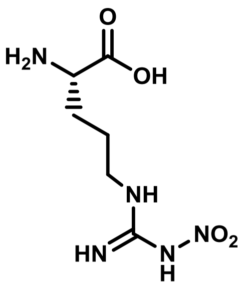


Y
Tyr
Tyrosine

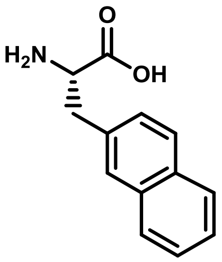


Non-proteinogenic amino acids

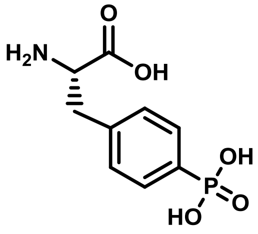
Arg(NO$_2$)
N-gamma-nitro-arginine



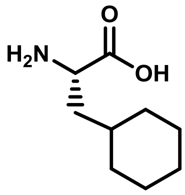
2-Nal
2-naphtylalanine



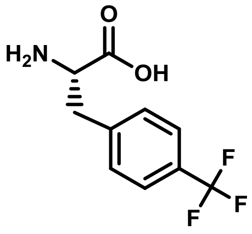
Tyr(PO$_3$H$_2$)
phosphotyrosine



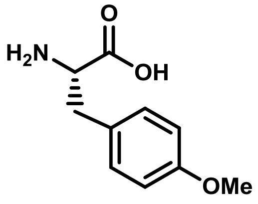
Cha
cyclohexyl-alanine



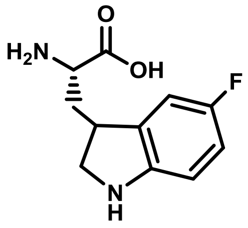
Phe(4CF$_3$)
4-trifluoromethyl-phenylalanine



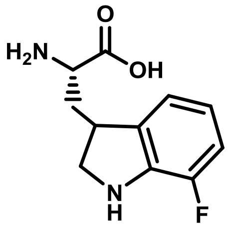
Tyr(Me)
O-methyl-tyrosine



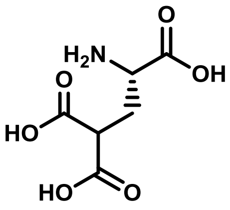
Trp(5F)
5-fluoro-tryptophan



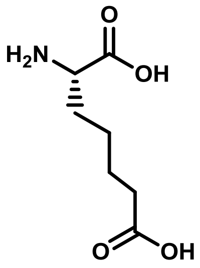
Trp(3F)
3-fluoro-tryptophan



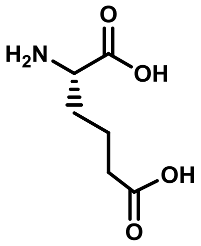
Gla
gamma-carboxy-glutamic acid



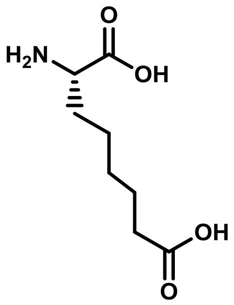
Api
pimelic acid



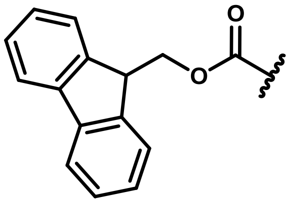
Aad
adipic acid



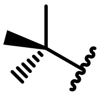
Asu
suberic acid



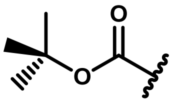
Protecting groups



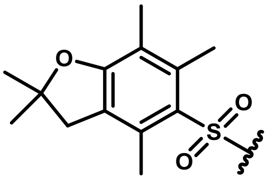
9-fluorenylmethoxycarbonyl
(Fmoc)



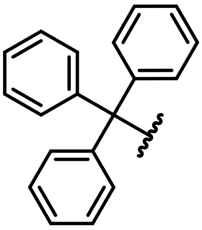
tert-butyl
(tBu)



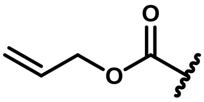
tert-butoxycarbonyl
(Boc)



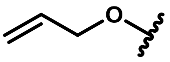
2,2,4,5,7-pentamethyl-
dihydrobenzofurane-5-sulfonyl
(Pbf)



trityl
(Trt)

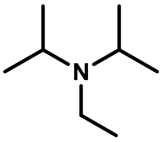


Allyloxycarbonyl
(Alloc)

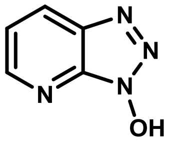


Allyl
(OAl)

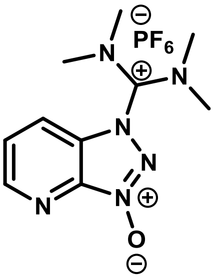
Coupling agents and additives



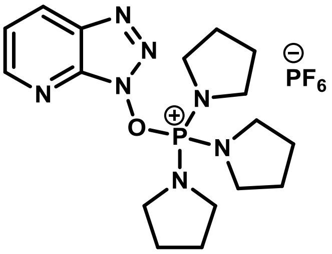
N,N-diisopropylethylamine
(DIEA)



1-hydroxy-7-azabenzotriazole
(HOAt)

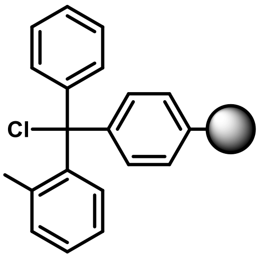


O-(7-Azabenzotriazol-1-yl)-
N,N,N’,N’-tetramethyluronium
hexafluorophosphate
(HATU)



7-Azabenzotriazol-1-yloxy)
tripyrrolidinophosphonium
hexafluorophosphate
(PyAOP)

Resin



2-chlorotrityl resin

INTRODUCTION

Molecular recognition of protein surfaces

In chemistry and biology, molecular recognition refers to the specific interaction between two or more molecules through complementary non-covalent bonding, for example, via hydrogen bonding; metal coordination; van der Waals forces; and π−π, hydrophobic, or electrostatic interactions. Recognition processes between biological molecules guide the selective interactions that are essential for life, including those between receptors and ligands, antigens and antibodies, nucleic acids and proteins, proteins and proteins, enzymes and substrates, and nucleic acids with each other.

Proteins play an essential role in cellular activity, including structural (cytoskeleton), mechanical (muscle), biochemical (enzymes), and cell signaling (hormones) functions. Molecular recognition of protein-protein interactions (PPIs) governs the affinity and specificity of complex formation and determines their biological functions. Examples of pathways mediated by protein interactions include gene expression, proliferation, intracellular communication and apoptosis. Numerous diseases occur because pathways involving particular PPIs malfunction. To date, only a few PPIs have been the subject of a drug discovery initiative. However, owing to the many opportunities presented by modulation of PPIs, this field has gained increasing scientific interest over the past decade and has seen many innovations from both academic and industrial research groups.

Protein interactions are fundamentally characterized as stable or transient, the latter being expected to control the majority of cellular processes. Stable interactions are those in which the individual interacting partners, that can be identical or different, are not observed *in vivo*. Hemoglobin and core RNA polymerase are examples of multi-subunit interactions that form stable complexes. Transient interactions are reversible and typically require a set of conditions that promote the interaction, such as phosphorylation, conformational changes or localization to discrete areas of the cell. In transient interactions the binding partners and the complex are in dynamic equilibrium, although with a wide range of binding affinities and on- and off-rate kinetic constants. While in contact with their binding partners, transiently interacting proteins are involved in a wide range of cellular processes, including protein modification, transport, folding, signaling, and cell cycling.

Proteins bind to each other through a combination of hydrophobic bonding, van der Waals forces, and salt bridges at specific binding sites on each protein. These sites can be small binding clefts or large surfaces and can be just a few residues long or span hundreds of amino acids, and the strength of the binding is influenced by the size of the binding region. In this sense, PPIs can be structurally classified as mediated either by the interaction between two protein domains (domain-domain) or by the interaction between a linear sequence of residues of one partner and a domain of the other one (peptide-domain) (**Figure 1**). In peptide-domain interactions, which accounts for almost 40% of known PPIs, the interacting motif often adopts a specific secondary structure, typically an α-helix, although it can also be completely disordered.[1]

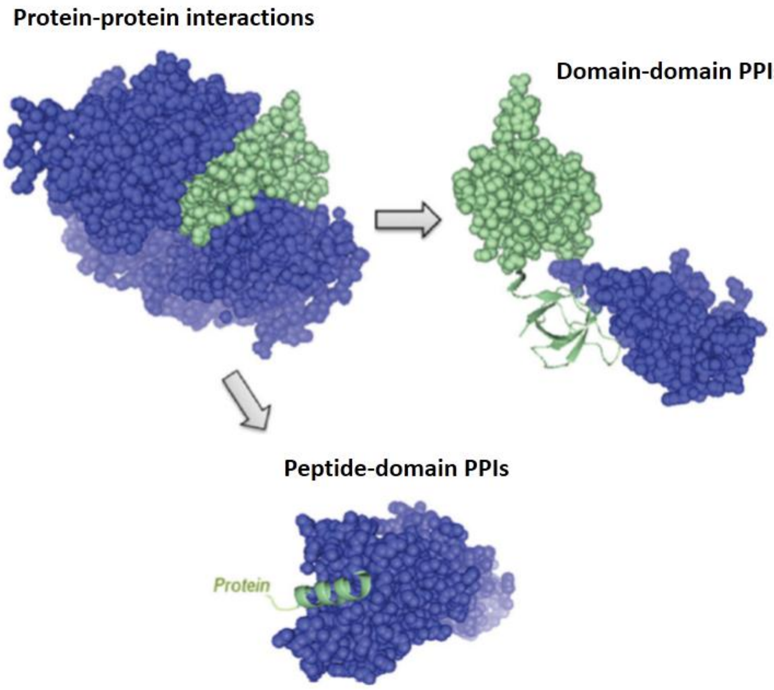


Figure 1. *Protein–protein interactions scheme representing domain-mediated or peptide-mediated interactions. Extracted from Nevola et al.* [2]

For some time, PPI targets came to be thought of as high risk, difficult targets and perhaps even undruggable, by many in the pharmaceutical industry; due to the fact that most PPIs interfaces are relatively flat and featureless, lacking pre-formed and well defined hydrophobic cavities which can fully accommodate a small molecule ligand. Such single, deep binding pockets that entirely surround the bound ligands are usually found on traditional druggable targets, and on average they occupy a volume of ~270 Å^3.[3] Hence, druggable targets were originally referred to any therapeutic protein target that could be modulated with small, orally available molecules.[4] However, this historically definition has been continually evolving.

The residues on the protein-protein interface do not contribute equally to PPIs. There is a collection of a few small binding pockets scattered across the interaction surface, each with a volume about 100 Å^3.[5] It has been shown that only some of these interactions, called "hot spots", are essential for affinity as they contribute to the majority of the total interaction energy. A hot spot is defined as a residue which substitution by an alanine leads to a significant decrease in the free energy of binding (ΔΔGbinding > 1.5 kcal/mol).[6] A survey of alanine scanning data indicated that the amino acid composition of hot spots was enriched in large amino acids such as tryptophan, arginine, and tyrosine.[7] Typically, hot spot density on the protein-protein interface composes 10% of the binding site residues.[8] The free energy of binding between two proteins is not a simple summation of the contribution from individual hot spots, but they tend to occur in clusters. Within the cluster, the tightly packed hot spots are in contact with each other and form a network of conserved interactions called hot regions or hot spot segments.[9] The contributions of hot spots within one hot region are cooperative to stabilize PPIs, while the energetic contributions between two hot regions can be additive[10] or cooperative.[11] The prevalence of hot spot residues in both domain-domain and peptide-domain interfaces[12] made it theoretically feasible to disrupt protein-protein interactions with inhibitors large enough to simultaneously interact with multiple hot spots patches and gain a significant part of the distributed free energy. It is estimated that more than 50% of domain-domain PPIs are mediated by one hot segment that contributes the majority of the binding interaction energy.[1]

Representative examples of successfully targeted PPI interactions include, among others, those of the p53 activation domain with MDM2,[13] Bcl-xL proteins with BH3 peptides[14] and cytokine interleukin-2 IL2 with its receptor IL2Rα.[15]

Peptides as chemical tools to recognize protein surfaces

A number of issues are raised for identifying ligands that recognize protein surfaces: the lack of starting molecules, the topological and chemical features of the surface to bind, the composition of current chemical libraries, the assessment of on-target effects. Over the past few years, these challenges have been tackled using a variety of strategies.

The modulation of PPIs has been pursued through many approaches, in particular those based on the use of small organic molecules, either derived from natural compounds or from pure organic synthesis. Favorable tracks for the development of small molecule PPI inhibitors emerged in the eighties from the energetic analysis of protein binding interfaces. Exploration of the contribution of different contacts established upon binding could be performed by alanine scanning of both partners. The experimental determination of hot-spots remains however tedious and time consuming even if higher throughput approaches have been developed since the early alanine scans.[16] Based on the structure of two partners forming a complex, several groups developed computational methods for the prediction of hot-spots.[17-19] On an average, these methods present a success rate of about 80% in their predictions and combining them can even improve the accuracy of the hot-spot prediction.[20] Other strategies such as the fragment-based ligand discovery (FBLD) approach provides a highly effective way of exploring the chemical space.[21-23] Fragment assembly allows access to very large numbers of potential fragment combinations. Although initial fragment hits are typically weak binders, two or more fragments can be covalently linked to yield medium-sized binders with higher affinity. However, the main difficulty encountered in the discovery of small molecule modulators derives from the large contact surfaces involved in PPIs (~1500-3000 Å^2) when compared with those that participate in protein-small molecule interactions (~300-1000 Å^2).[24] In addition, as we have already introduced, these protein-protein interfaces are generally flat and often lack the grooves and pockets present on the surfaces of proteins that bind to small molecules.[25] Researchers have thus turned their attention to the possibility of using ‘‘more natural’’ alternative PPI modulators, such as recombinant proteins, antibodies, and peptides, which can explore larger surfaces.

Large and diverse PPI surfaces, which feature complex topologies of multiple low energy interaction sites, are well-complemented and effectively modulated by protein-based biologic drugs. Due to their greater size and well-defined three dimensional conformations, biologics can bind their protein targets with high affinity and remarkable selectivity. Development of monoclonal antibodies competing with the PPIs formation has gain increasing importance in the last decade. This approach, necessitating the 'humanization' of immunoglobulins to prevent an immune reaction, was remarkably successful, so that each year, tens of new antibodies started clinical trials[26] or reached FDA drug approvals[27-30] for therapeutic applications or diagnostic purposes. The exploitation of simpler immunoglobulin systems such as the nanobodies found in camelids should give rise to more stable and simpler molecules.[31, 32] Computational methods also contributed significantly to this field. Modeling and design tools such as the Rosetta suite have reached sufficient accuracy to be able to guide the rational design of binders.[33] The combination of rational design strategies with state-of-the-art display technologies such as ribosome or yeast display coupled to deep sequencing technologies open great perspectives for the diversification of the classes of binders.[34-36] Computational methods can also be used to reduce the immunogenicity of designed proteins.[37] Nonetheless, the major limitation of antibodies, nanobodies and other proteins selected for tight binding of a target remains to be the difficulty to bring these proteins into cells. Their applications remain mainly limited to extracellular targets, such as membrane-bound receptors and secreted proteins.

Given the current success of recombinant-protein-based therapeutics, we are already witnessing the erosion of the long-standing and relatively narrow definition of what constitutes a druggable target. Nowadays, the definition of druggability has widened to include targets whose activity can be modulated by larger molecules, such as proteins and peptides.

Peptides, which are distinguished from proteins based on their smaller size (50 amino acids or less), mediate various essential biological functions, such as signal transduction, heart rate regulation, food intake, and growth. Natural peptides such as insulin, oxytocin, erythromycin and somatostatin; and synthetic peptides such as octeotride

(Novartis Pharmaceutical), goselerin (AstraZeneca) and leuprolide (Bayer) are successful drugs, which underlines the great market potential of peptide drugs.

Therapeutic peptides traditionally have been derived from three sources: (i) natural or bioactive peptides produced by plants, animal or human (derived from naturally occurring peptide hormones or from fragments of larger proteins); (ii) peptides isolated from genetic or recombinant libraries and (iii) peptides discovered from chemical libraries. Generally, the size of the peptide determines the most suitable technology for its production: chemical synthesis, recombinant DNA technology, cell-free expression systems, transgenic animals and plants or enzymatic synthesis. With the use of unnatural amino acids and pseudo-peptide bonds, chemical synthesis offers access to a much wider chemical diversity than peptide derivatives produced by recombinant technologies, with a diversified potential for intellectual property. Production of synthetic therapeutic peptides has become possible for the pharmaceutical industry with recent developments of solid-phase peptide synthesis (SPPS), initially developed by Merrifield.[38] Nowadays, SPPS is crucial in the early steps of preclinical research and in the production of peptide-based active pharmaceutical ingredients (APIs).[39]

Furthermore, over the past two decades, library-based display technologies (such as phage display, yeast and bacteria display, ribosome display and mRNA display) have been staggeringly optimized since their appearance in order to mimic the process of natural molecular evolution. These technologies are essential for the isolation of specific high-affinity binding molecules (proteins, polypeptides, nucleic acids and others) for diagnostic and therapeutic applications in cancer, infectious diseases, autoimmune, neurodegenerative, inflammatory pathologies *etc.* Displayed technologies have been extensively applied to find peptide ligands and small proteins for PPIs modulation. With display technologies a new era of biomolecule discovery has commenced. These technologies permit the *de novo* discovery of highly active, high affinity peptides from libraries, improvement of the performance of natural peptide sequences and modification of the binding properties of natural peptide structures. Further optimizations through synthetic chemistry and rational design that may be complementary to such recombinant methods are improving the applicability of display technologies. Moreover, recently, the development of *in vivo* display of phage libraries that can impart delivery to peripheral tissues and circulating immune cells, where they act as ligand mimics, has opened new and very promising avenues in therapeutics and other applications. Many of these technologies are still in their infancy and further improvements to the methodologies, target presentation and panning of libraries against cells, tissues and difficult receptor targets will expand the list of targets that have so far been intractable to therapeutic intervention.

Peptides present several advantages when compared with small molecules and biologics: (i) flexibility, which is translated into adaptability to large surfaces; (ii) easy modularity, which increases structural diversity and consequently allows higher selectivity and potency; (iii) size, which limits accumulation in tissue; and (iv) complete biocompatibility, which means low toxicity in humans.[40] The last two features are highly desirable given the growing interest in PPIs as therapeutic targets. Moreover, non-natural building blocks and various chemical scaffolds can be incorporated into a peptide sequence to create a palette of modified peptides with a wide range of functionalities and chemical diversity.

However, progress towards the development of therapeutic peptide PPI modulators is hindered by the following drawbacks of these molecules: low stability against degradation by proteolytic enzymes of the digestive system and blood plasma; rapid removal from the circulation; poor ability to cross physiological barriers; and potential immunogenicity.[41] In spite of these limitations, the large number of successful peptide PPI modulators reported so far and the great effort to tackle the bottlenecks that impair their use as pharmaceutics are impressive. In addition to features that allow cell and tissue permeability, many chemical modifications and smart linker conjugations[42] have been introduced into PPI modulators in order to reduce proteolytic degradation and improve bioavailability.

Nowadays, increasing stability of the active conformation and decreasing susceptibility toward proteolysis are the most important goals of the introduced peptide modifications. There are two major groups of structural changes: modifications of the backbone and cyclizations (**Figure 2**).

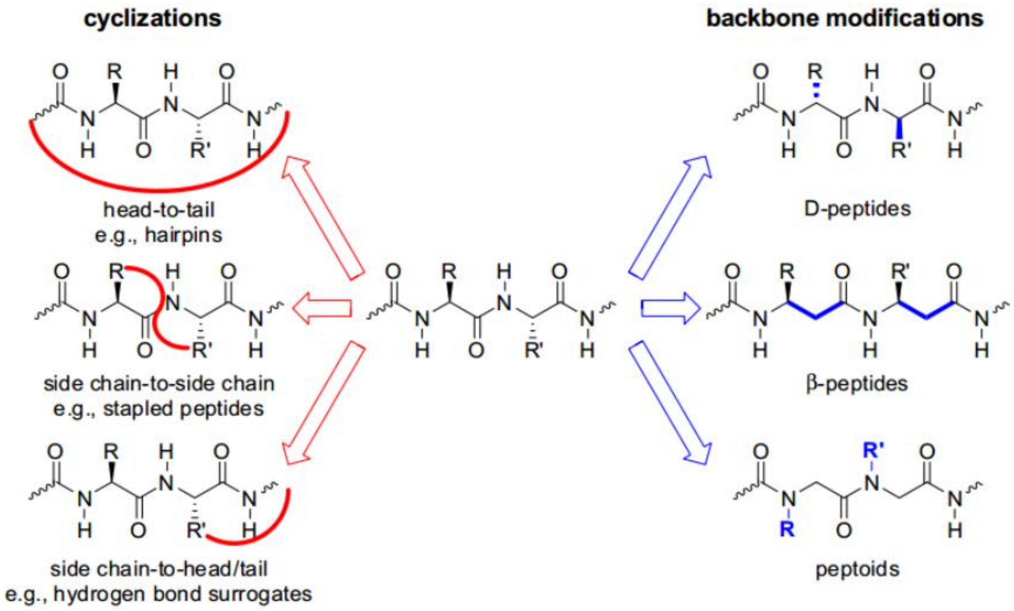


Figure 2. Major modification types of peptide-based protein-protein interaction inhibitors. Extracted from Wójcik et al.[43]

Peptide backbone modifications

The first group of peptide modifications, which is based on changes in the backbone, usually alternates the properties of the compound more profoundly, and the obtained sequences and three-dimensional structures differ significantly from the original protein fragment. The major possibilities are as follows: variation in the stereochemistry (the use of D-amino acids), extension of the backbone (incorporation of β-amino acids), and switch of the side chain position from Cα to amide nitrogen atoms (peptoids).

Use of D-amino acids

D-Peptides usage in medicinal chemistry is one of the longest known strategies for developing peptide-based bioactive compounds. The major advantage of this approach is avoidance of the fast proteolytic degradation that is typical for the L-peptides in blood plasma. Unfortunately, D-peptides exhibit problematic features, namely low conformational stability and mirror three dimensional arrangements in relation to L-peptides that make structure-based design quite challenging. In particular, helices formed by D-peptides have the opposite handedness of the natural ones, and the binding mode for such fragments usually differs from that observed in the native ligand-protein complexes. In some cases, the retro-inverso strategy could be applied,[44] whereas the mirror-image phage display technique is of more general utility.[45]

Incorporation of ß-amino acids

β-peptides, that is, oligomers containing β-amino acids, have several key advantages in the construction of effective PPI inhibitors.[46, 47] The possibility to predict the relationship between the sequence and three-dimensional structure makes the β-peptides very useful for discovering new bioactive compounds.[48, 49] There are two general approaches for designing β-peptides: (i) exclusive usage of β-residues or (ii) combination usage of constrained β-residues and natural α-residues. Due to high conformational stability, peptides containing β-residues can be significantly shorter than a-peptides but achieve a similar degree of folding. An additional advantage of this class of compounds is that they have an unusual resistance to degradation by proteases.[50]

Peptoids

Peptoids consist of repeating N-substituted glycine units.[51] Peptoid monomers differ from natural α-amino acids in that the side chain is attached to the backbone nitrogen instead of the α-carbon. They are able to fold into helices that mimic the peptide structure and function.[46] Due to the lack of the H-bonding capacity of the backbone, the induction of folding into well-defined helical structures must be performed through the careful choice of the N-substituent groups. Peptoids exhibit several advantages, including a particularly high resistance to proteolytic degradation.[52] The major problem in the design of peptoids is the appropriate matching to the surface of the protein, and it often requires many attempts to position the key interacting residues in the right place.

Peptide cyclization

Extensive analysis of the available structural data of PPI complexes shows that the latter are usually mediated via various folded subdomains, and α-helices are the structural motif that mostly facilitates this interaction. For peptidic ligands, conformation is believed to be the primary determinant of bioactivity and bioavailability.[53] However, when a short peptide fragment is removed from their parent protein, they generally lose their well-defined, biologically active conformation and their proteolytic stability. Hence, it is intuitive to start designing bioactive peptides by carefully mimicking structural elements such as turns, β-sheets or α-helices at the PPI recognition interfaces.

Cyclizations that result in peptidic macrocycles is one of the most widely chosen methods for developing PPI inhibitors.[54] Various strategies such as hydrogen bond surrogate, stapling, and hairpins were developed to stabilize turns, helices and extended conformations.[54-56]

Hydrogen bond surrogate (HBS)

This modification induces an α-helical conformation in short peptide sequences.[57] The HBS α-helices feature a carbon-carbon bond derived from a ring-closing metathesis reaction in place of an N-terminal intramolecular hydrogen bond between the peptide i and i+4 residues, forming a macrocycle within the first turn of the α-helix. The HBS approach is an attractive strategy for generation of ligands for protein receptors because placement of the cross-link on the inside of the helix does not block solvent-exposed molecular recognition surfaces of the molecule.[56] Typically, the hydrocarbon linker obtained by a ring closing metathesis (RCM) reaction is used to build the HBS ligand. [58] A hydrazone bridge can also be effectively applied to mimic the hydrogen bond, as shown by Cabezas and Satterthwait.[59] The conformational stability of the HBS helix significantly improves its resistance to proteolysis and increases the cell penetration properties.

Stapled peptides

Stapled peptides are among the most common peptide inhibitors of PPIs that restore and enhance the natural α-helical structure of a peptide.[60-62] The principle of the design of such modifications is based on joining side chains that are in spatial proximity to the α-helical conformation by covalent bonds. Usually, i, i+4 or i, i+7 hydrocarbon staples that contain 8 or 11 carbon atoms, respectively, are formed. Alternatively, disulfide bridges,[63] lactamization,[64] cysteine crosslinking,[65] azide-alkyne cycloadditions,[66] biaryl linkages formation using borylated phenylalanine derivatives can be used.[67] The most important advantage of this approach is the possibility to introduce constraints at any chosen part of the helix, allowing for proper modulation of the conformational stability and the discovery of staples that do not interfere with ligand-protein interactions.[68]

Importantly, properly designed stapled peptides exhibit other important features, including an increased resistance to proteolysis and high cellular uptake.[68-70]

In a recent work, our research group reported the use of light-regulated staple peptides as photoswitchable inhibitors of the β-arrestin-β-adaptin 2 interaction.[71]

β-Hairpin mimetic

Another interesting and often functionally important epitope used for molecular recognition by PPI interfaces is the β-hairpin. The approach to the design of the β-hairpin involves the use of a protein epitope.[56] The starting point for designing this synthetic molecule is the identification of key epitopes that are involved in PPIs. The discovered epitopes are transferred on semi-rigid macrocyclic templates of hairpin loop sequences. The β-hairpin template is a universal scaffold that can be used to mimic epitopes based on various types of secondary structures.[72] To obtain the stable β-hairpin conformation, a cyclic peptide that contains a turn inducing unit must be designed. D-Pro-L-Pro is one of the most commonly used fragments for this purpose,[73] but other rigid units could also be applied, for example, squaramide-based modules have been recently proposed.[74] Typically, cyclization of β-haipins is done by coupling N- and C-termini, although other methods using reactivity of side chains were also applied, including, disulfide formation,[75] azide-alkyne cycloaddition,[76] Trp-Trp crosslinking.[77]

Constrained cyclic peptides mimicking protein structure motifs

The design of macrocyclic peptidomimetics held great promise for structural and functional mimicry of protein surfaces.[60, 78-81] Small synthetic cyclic peptides, structurally fine-tuned with additional molecular constraints, may mimic, in favorable cases, elements of protein secondary structure and thereby downsize protein surfaces to smaller molecules. Combining cyclization methodologies with other strategies to efficiently lock peptides in desired conformations is a long-standing goal in chemical sciences and not only related to the design of PPI inhibitors.[2] Such constrained cyclic peptides can have protein-like biological activities and potencies, enabling their uses as biological probes and leads therapeutics, diagnostics and vaccines.

Mimetics of turn structures

With a length of two to six amino acids, turns are irregular secondary structure elements that differ from helices and β-sheets through the non-repetitive dihedral angles of their backbones. Turn structures were defined as regions that allow a polypeptide chain to fold back on itself, thereby enabling the formation of globular proteins.[82] Over the last few decades, several more general definitions were described, with a widely used one classifying turns in accordance to the hydrogen bond pattern formed between the backbone carbonyl group of the residue at position i and the backbone amide proton at position i+n. This leads to the three more common families of γ-, β-, and α-turns with three to six amino acids in length and n=2, 3, and 4, respectively.

Cyclization of peptides has been the most common method used to stabilize turns and can be accomplished in a variety of ways (head to tail, head to side chain, side chain to tail, side chain to side chain) (**Figure 3**). Many chemical approaches have been used to achieve cyclization, ranging from disulfide or lactam formation to RCM and click chemistry. The chemistry of peptide cyclization has been reviewed in detail.[83-85]

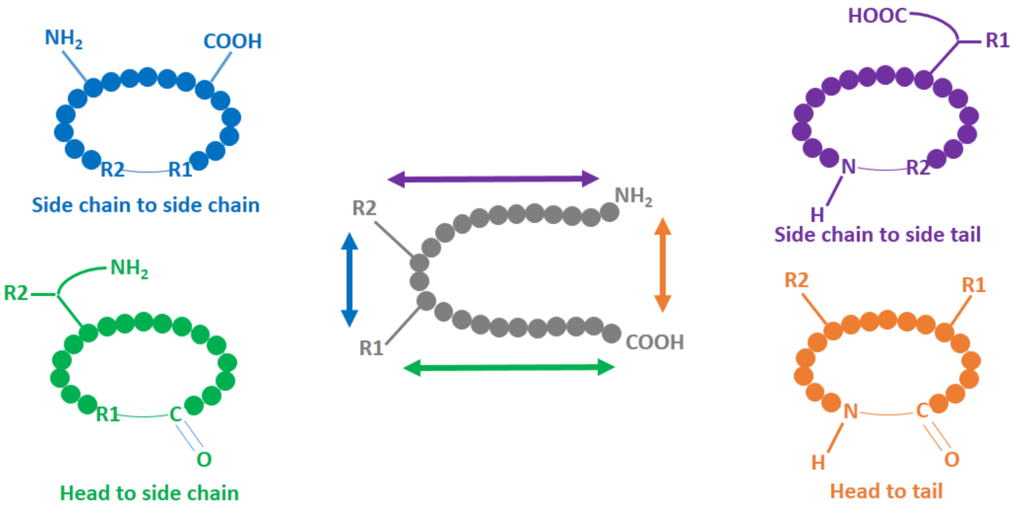


***Figure 3**. Schematic representation of the different ways to cyclize a linear peptide*

Other turn-inducing structural features are often needed to help constrain a cyclic peptide into a turn. One method is to introduce a *cis*-amide bond into a peptide chain, forming a motif analogous to a β-turn. This is often achieved by incorporating a proline ring, glycine,[86] D-amino acid[87] or N-methyl[88] into peptide structures (**Figure 4**).

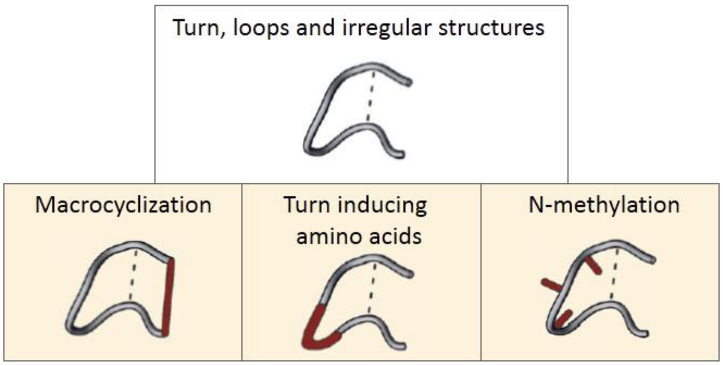


Figure 4. *General strategies for turn stabilization and mimicry. Adapted from Pelay et al.*[89]

N-methylation can also be substituted with an N-(4-azido butyl) group to induce *cis*-amide bond formation.[90] The triazole group formed by click chemistry has also been utilized as a turn-inducing *cis*-proline mimetic.[91] Pseudo-prolines[92] and heterocycles[93] have been used to introduce turn structures into cyclic peptides. 4S-Azidoprolines have also been used to generate all *cis* cyclic peptides.[94]

Mimetics of β-strands and β-sheets

The β-strand/sheet accounts for 30% of protein secondary structure. Peptide β-strands[95] are extended "linear" peptides, usually paired in proteins as antiparallel, parallel or barrel β-sheets,[96] the strands joined together by an array of hydrogen bonds that stabilize structure. Distinctive phi and psi dihedral torsion angles distinguish strands and sheets from α-helix and β-turn structures. Amino acids that favor strands in proteins are branched residues Val, Ile, Thr as well as Tyr, Cys, Trp, Phe. A key feature of the β-strand is that alternating side chains (e.g. i, i+1) point in opposite directions and thus first and third residues (e.g. i, i+2) in a peptide sequence have side chains on the same face of a strand/sheet, bringing them into proximity for potential ring closure to cyclic peptides. Strands are recognized as important recognition motifs in protein-protein interactions.[97-99] The combination of two antiparallel β-strands connected through a turn provides a β-hairpin which is stabilized by an extended pattern of inter-strand hydrogen bonds.[100]

Several methods have been developed to synthesize modified β-strands, β-hairpins, and β-sheets.[101-103] Three general approaches can be distinguished: the use of turn mimetics that nucleate β-sheet formation,[104-107] covalent or noncovalent macrocyclization[108-110] (backbone or side chain to side chain), and the use of β-strand-enforcing amino acids (**Figure 5**).[111-114]

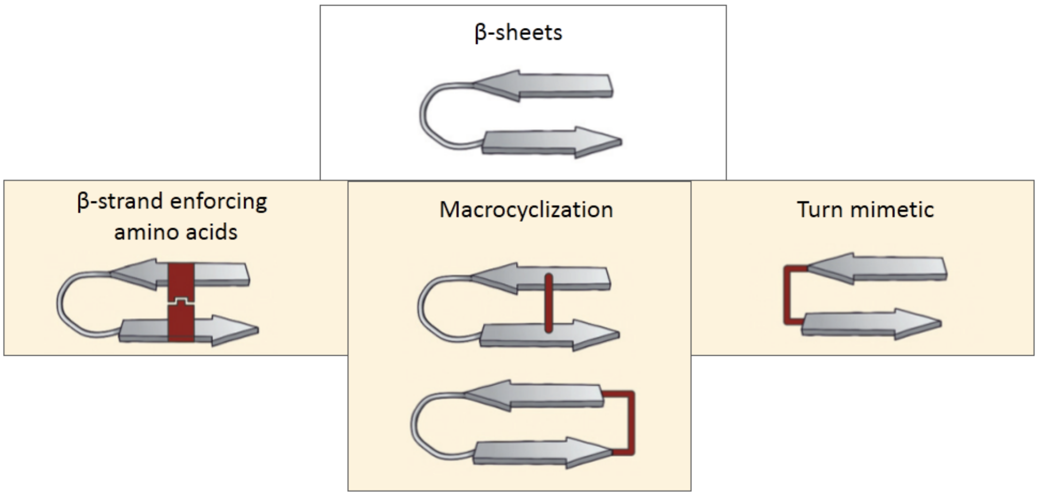


Figure 5. *General strategies to afford β-sheet mimetics. Adapted from Pelay et al.*[89]

There is enormous untapped potential for designing β-strand/sheet mimetics to interfere in PPIs or mimic one of the binding partners. Because one face of a β-strand is often exposed to solvent, every second side chain can potentially be linked to form a β-strand mimicking macrocycle, which might be linked to another β-strand macrocycle, or α-helical macrocycles, or acyclic peptide chains, or non-peptide appendages. Use of macrocycles as templates to produce strand mimetics can be a valuable new approach to engineering tool compounds for investigating the importance of β-strand surfaces of difficult and less druggable protein targets.

Mimetics of helices

Helices are defined by consecutive turns linked together by hydrogen bonds that bridge across turns. Helices are repetitive secondary structure elements which make up more than 30-40% of structured protein domains.[115] In natural proteins, only helices with an integral number between three and five are observed (3_{10}-, 3.6_{13}-, and 4.4_{16}-helix), even if others are theoretically stable.[116] Only the 3_{10}-helix, which consists of repetitive β-turns, retained its nomenclature. The 3.6_{13}- and 4.4_{16}-helices, built up by consecutive α- and π-turns, are better known as α- and π-helices, respectively. Whereas the π-helix is rarely observed in protein secondary structures,[117, 118] 3_{10}-helices contribute to 10% of all helical regions in globular proteins.[119] The remaining 90% are α-helices.

Based on available structural data,[120] helices contribute to the protein-protein interface in 62% of all PPIs,[121] thus highlighting the importance of α-helices in this context. Different stabilization approaches have been reported since the early 1980s.[122] Most prominent strategies towards the preparation of helix mimetics involve side chain to side chain cross-linking of peptides, the introduction of stabilizing N-terminal caps and the use of foldamers, which are peptide and nucleic acid inspired oligomers which exhibit major backbone alterations (**Figure 6**).

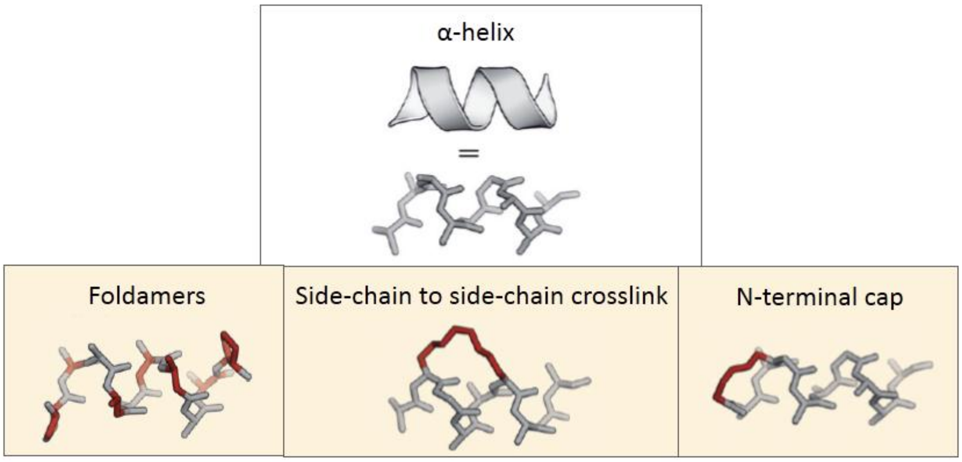


Figure 6. *Schematic representation of an α-helix together with general strategies of helix stabilization and mimicry. Adapted from Pelay et al.*[89]

Side chain to side chain crosslinking can be achieved by: thiol-based cross-link,[71, 123-127] lactam cross-link,[64, 128-130] triazole cross-link[131-134] and α-methylated hydrocarbon cross-link.[70, 135-138] The most successfully used N-caps for the generation of PPI inhibitors are hydrogen-bond surrogates. β-peptides are the most exhaustively studied foldamers with well-characterized folding propensities.

Cyclic peptides as therapeutic agents and biochemical tools

Several cyclic peptides found in nature are used in clinic. Some examples are gramicidin and tyrocidine with bactericidal activity, cyclosporin A with immunosuppressive activity, and vancomycin with antibacterial activity, and so on. Usually, cyclic peptides show better biological activity compared to their linear counterparts due to the conformational rigidity, which decreases the entropy term of the Gibbs free energy, therefore allowing the enhanced binding toward target molecules, or receptor selectivity. Another benefit from cyclic structure is the resistance to hydrolysis by exopeptidases due to the lack of both amino and carboxyl termini. Cyclic peptides can be resistant even to endopeptidases, as the structure is less flexible than linear peptides. In addition, some cyclic peptides, though not all, can cross the cell membrane. Cyclosporin A is a good example of the membrane permeable cyclic peptides.

Cyclic peptides are thus envisaged to overcome some of the limitations of their linear counterparts by addressing biological problems such as the enzymatic hydrolysis and poor oral bioavailability. Improvement of their pharmacological properties in conjunction with the current advances on rational drug design, peptide synthesis, and structure determination have assisted the development of novel and more potent cyclic peptides. The growing interest in peptide-based drugs in medicinal research, as well as the promising *in vitro* and *in vivo* efficacy of cyclic peptides, provides many opportunities for the development of cyclic peptides towards the treatment of several diseases.

Cyclic peptides also have potential applications as biochemical tools such as receptor agonist/antagonist. Cell surface receptors can be good targets for cyclic peptide compounds as the peptide would work on the cell surface and it is not necessary for the peptide to cross the cell membrane.

Outlook and opportunities

During the last three decades there has been an exponential growth in an average number of peptide therapeutics entering clinical studies. Peptide drugs are being investigated in a wide variety of therapeutic areas, with the highest number of them entering clinical trials for metabolic disorders and cancer treatments (**Figure 7**).

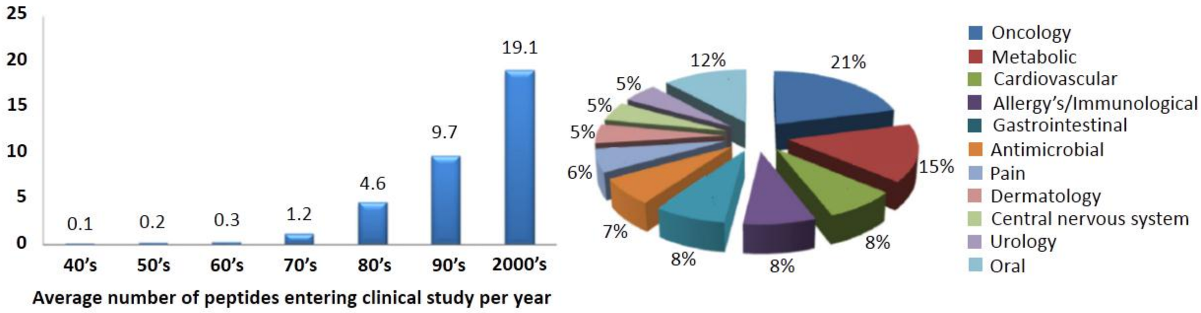


Figure 7. *Average number of peptides entering clinical study per year. The distribution of peptide pipeline drugs by therapeutic areas. Other indications that are not specified on the chart include: Endocrinology, Respiratory, Bones, Hematology, Ophthalmology, OB-GYN (obstetrics and gynecology). Data was taken from a 2010 and 2012 report provided by the Peptide Therapeutic Foundation.*

With the aim of targeting chemically intractable proteins that represent a vast majority (>80%) of the human proteome, great efforts are being devoted in industry and academia to the discovery and development of new peptide ligands to recognize challenging protein surfaces.

Typically, the first design step is based on the crystal structure of the molecular target with the fragment of its natural counterpart. Unfortunately, structure information is only available for an estimated 50% of all protein targets, with a significant under-representation of targets with high therapeutic importance, such as membrane proteins. This combined with the lack of diversity in the commercial compound libraries used for screening, highlight the need for new and diverse scaffold libraries that could ensure more successful screening processes to find PPI modulators in the future.

A major challenge in drug development is the treatment of brain disorders such as brain cancer, Alzheimer's and Parkinson's disease. Most potential drugs for the treatment of central nervous system (CNS) disorders fail due to the restrictive nature of the blood-brain barrier (BBB), a highly specialized biological barrier that insulates the CNS from the other parts of the body. It has been estimated that more than 98% of low molecular weight candidate drugs and almost 100% of large therapeutic candidate drugs cannot cross the BBB.[139] One approach to deliver active molecules into the brain is the use of molecular vectors, also known as BBB-shuttles. These are molecules that have the capacity to carry a cargo across the BBB through either active or passive transport without affecting the BBB integrity. Endogeneous proteins,[140] antibodies[141] and peptides[142] have been used as BBB-shuttles. There is an increasing interest in the development of peptide BBB-shuttles because they overcome some of the weakness of classical protein shuttles, including complex derivatization and characterization, high immunogenicity, and costly production. Although most peptides have a limited capacity to cross the BBB, in the recent years, several research groups, including that of Prof. Giralt, have been actively working on the development of new protease resistant BBB-shuttle peptides.[44, 143]

Besides therapeutic applications, the ability of peptides to recognize specific receptors and proteins can be exploited in various biotechnological and pharmaceutical applications, including peptide-based biosensors or biomarkers.[144] Bio-conjugation strategies can provide an efficient way to convert interaction information between peptides and analytes into a measurable signal, which can be used for fabrication of novel peptide-based biosensors. Many sensitive fluorophores can respond rapidly to environmental changes and stimuli manifest as a change in spectral characteristics, hence environmentally-sensitive fluorophores have been widely used as signal markers to conjugate to peptides to construct peptide-based molecular sensors. Several research groups, including that of Prof. Imperiali, have been working on the use of peptide-based biosensors to study protein interactions.[145]

OBJECTIVES

In the introduction of this thesis we have seen the enormous potential of peptides, and especially cyclic peptides, to be used as therapeutic and diagnostic tools in various pharmaceutical and biotechnological applications such as peptide-based therapeutics, biosensors and functional modulators of proteins. In this context, the present thesis addresses the following objectives.

1) The first objective was to assess the ability of cyclic peptides to recognize protein surfaces of challenging targets such as VEGF.

   In order to accomplish this first objective, we established the following aims:

   i. To screen a library of cyclic hexapeptides against VEGF to find possible VEGF-binding candidates.
   ii. To study the conformational behavior of cyclic hexapeptides to better understand their binding capacity.
   iii. To design and conduct an evaluation process of peptide 7-derivatives to improve the VEGF-binding affinity and specificity of this cyclic peptide.
   iv. To study the effect of stereochemistry on the conformation of cyclic hexapeptides with the aim of gaining insights into the design of secondary structured cyclic peptides.

2) The second objective was to apply the acquired knowledge on peptide characterization by NMR to study the conformational behaviour of apamin analogues.

   To achieve this objective, we established the following aims:

   i. To perform a conformational study of a bicyclic analogue of apamin (MiniAp-1).
   ii. To conduct a conformational study of minimized monocyclic analogues of apamin (MiniAp-3 and MiniAp-4) and to compare their conformational behavior with the one of a linear analogue of apamin (MiniAp-2).

At the end of my thesis I did a short stay at the Massachusetts Institute of Technology (MIT), in Prof. Imperiali's lab, with the aim of expanding my knowledge on the use of library-based display technologies and the application of peptides and small protein scaffolds as biosensors. For this short stay, the following objective was established:

3) The application of direct evolution techniques for engineering protein ligands for challenging therapeutic targets, in order to convert these protein-based ligands into biosensors.

The following aims were proposed to achieve this second objective:

i. To set up a protocol for the expression and purification of soluble Sso7d-based cysteine mutants that can be labeled with cysteine reactive 4-DMN derivatives.
ii. To label the Sso7d-based cysteine mutants with 4-DMN derivatives in order to perform fluorescence spectroscopic assays to assess the binding of these mutants to MSA, as a proof of concept.
iii. To use the yeast surface display methodology to engineer Sso7d-based binders of hEGF.

RESULTS AND DISCUSSION

Cyclic hexpeptides for VEGF protein surface recognition

Angiogenesis, the process that involves the formation and maintenance of blood vessel structures, is essential for the physiological functions of tissues and it is also known to play important roles in the progression of diseases such as cancer and inflammation.[146, 147] In recent decades, a variety of signaling molecules, such as VEGF-VEGFRs, ephrin-Eph receptors, angiopoietin-Tie, and the Delta-Notch system, have been identified as key factors in angiogenesis. Among all these, vascular endothelial growth factors (VEGFs) and receptors (VEGFRs) regulate both vasculogenesis, the process of new blood vessel formation during embryonic development of the cardiovascular system; and angiogenesis, the formation of blood vessels from pre-existing vessels at a later stage.[148]

The VEGF family of genes contains at least 7 members, whereas the VEGFR family of genes has 3 to 4 members depending on the vertebrate species.[149, 150] VEGF is a protein with vascular permeability activity that was originally purified from a fluid secreted by a tumor.[151] Some years later, a protein with angiogenic activity was independently purified and named VEGF.[152] However, subsequent studies revealed that these two proteins were identical and encoded by a single gene.[148] The VEGF family includes VEGF-A, VEGF-B, VEGF-C, VEGF-D, PlGF (placental growth factor), VEGF-E (Orf-VEGF), and Trimeresurus flavoviridis svVEGF. The last two members, do not exist in mammalian genomes, including humans.

VEGF-A, from now on VEGF, is expressed in multiple spliced forms, including polypeptides of 121 ($VEGF_{121}$), 165 ($VEGF_{165}$), 189 ($VEGF_{189}$), and 206 ($VEGF_{206}$) amino acids in humans. Except for $VEGF_{121}$, the other variants have a basic stretch near the carboxyl terminus. The basic stretch of $VEGF_{165}$, the most important isoform of VEGF both quantitatively and qualitatively, has a weak affinity for acidic molecules such as heparin and to neuropilin-1, a membrane protein involved in neuronal cell regulation and a co-receptor for VEGF. In $VEGF_{189}$ this basic stretch has a strong binding affinity to heparin, and thus, most of these molecules are localized on the cell surface or in the extracellular matrix. The VEGF gene is unique in terms of its haploid insufficiency: even if only a single copy of the VEGF gene is deficient the mutant embryo dies at early embryogenesis due to immature formation and dysfunction of the circulatory system.[148] This means that the local concentration of VEGF in tissue is tightly regulated in embryogenesis, and half the level of this protein is insufficient to complete the formation of the closed circulatory system in the body. VEGF binds to and activates both

VEGFR-1 (Flt-1) and VEGFR-2 (KDR/Flk-1), promoting angiogenesis, vascular permeability, cell migration, and gene expression.[150]

VEGF-B and PIGF bind to and activate only VEGFR-1. Both PlGF–/– and VEGF-B–/– mice are alive at birth with no significant defects related to angiogenesis, suggesting that these genes are dispensable at embryogenesis. Nevertheless, under pathological conditions, synergism between PlGF and VEGF has been shown to contribute to angiogenesis.[153] VEGF-B–/– mice have the phenotype of an atrial conduction defect.[154] In addition, VEGF-B was recently reported to protect against the degeneration of sensory neurons.[155] These results indicate that, although PlGF and VEGF-B are not essential at embryogenesis, they have a variety of functions under pathological or stressed conditions.

VEGF-C and VEGF-D are produced as premature forms and are cleaved by proteases such as furin in both the amino- and carboxyl-terminal portions.[156] After processing, these molecules develop a higher affinity for VEGFR-3, which is expressed on lymphatic endothelial cells and stimulates the receptor for lymphangiogenesis. In addition, these proteins have a weak affinity for VEGFR-2, activating angiogenesis to some extent. VEGF-C is expressed during embryogenesis, whereas VEGF-D is expressed after birth during adult stages. This difference in gene expression is thought to be a major cause for lethality in VEGF-C–/– mice but not in VEGF-D–/– mice. VEGF-C–/– mice show severe accumulation of fluid in tissues due to poor development of lymph vessels.[157]

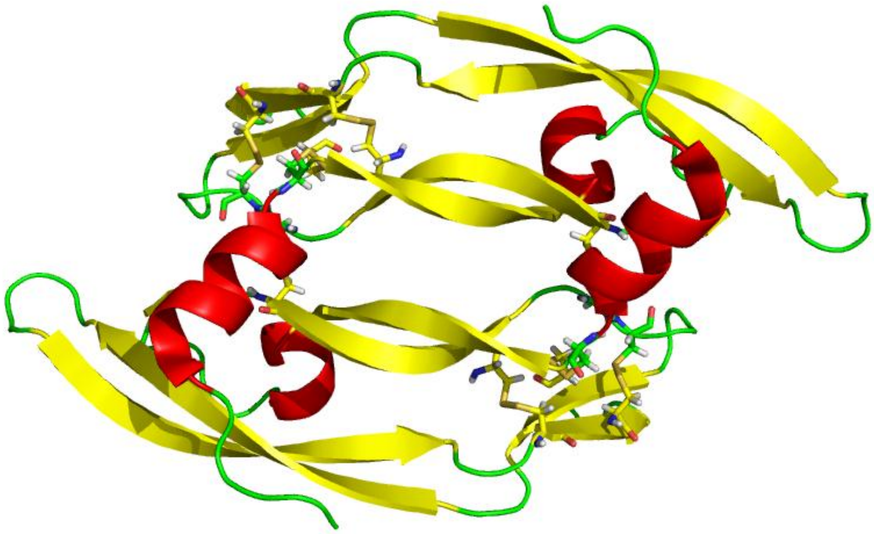


Figure 8. *Cartoon representation of the VEGF homodimer. B-strands are colored in yellow, α-helices in red and turns in green. Disulfide bonds are shown as sticks.*

Essentially, all the VEGFs have eight conserved cysteine residues at fixed positions, which are very similar to the PDGF family. Among the eight cysteine residues, six form three disulfide intramolecular bonds and generate three loop structures,[158] while the remaining two cysteine residues form two disulfide intermolecular bonds, contributing to the stable homodimer structure of VEGF (**Figure 8**). Two symmetric receptor binding surfaces at the poles of the VEGF homodimer have been identified by mutational and structural studies. The crystal structure of the VEGF-VEGFR-1 complex revealed that no major conformational changes occur on VEGF upon receptor binding and that the intermolecular interaction is mediated mainly by hydrophobic contacts.[159] Both X-ray and NMR spectroscopy studies suggest that the receptor binding region of VEGF has intrinsic conformational flexibility.[160, 161] Such flexibility may have important functional implications that allow the protein to interact with multiple receptors.

VEGFRs are typical tyrosine kinase receptors (TKRs) carrying an extracellular domain for ligand binding, a transmembrane domain, and a cytoplasmic domain, including a tyrosine kinase domain.[149] The overall structure of VEGFRs is similar to that of the PDGFR family members; nonetheless, some clear differences can be found between these two receptor families: the PDGFR extracellular region contains five immunoglobulin (Ig)-like domains, whereas VEGFRs bear seven Ig-like domains. Both families share a tyrosine kinase domain with a long kinase insert (KI) of 60 to 70 amino acids; even so, the amino acid sequences in the KI are very different from each other. VEGFR-1 has a high affinity for VEGF (K_D = 1-10 pM), which is one order higher than that of VEGFR-2, whereas its tyrosine kinase activity is approximately 10-fold weaker than that of VEGFR-2.[162]

The major pro-angiogenic signal is generated from the ligand-activated VEGFR-2. Within the KI or carboxyl-terminal region, TKRs have tyrosine autophosphorylation sites, which are important for the downstream signal. Unlike most of the TKRs that activate the Ras pathway or PI3K pathway, it was found that the PLCγ-PKCMAPK pathway is highly activated in VEGF-bound VEGFR-2 and used as a crucial signal for endothelial proliferation.

The signaling cascade of VEGFR-1 is not fully understood and the biological functions of VEGFR-1 have been the topic of several studies. Some results indicate that VEGFR-1 signaling is important for the progression of tumors *in vivo* mostly via bone marrow-derived VEGFR-1 positive cells. In addition, Wu et al.[163] reported that some human tumors, such as breast carcinomas, express VEGFR-1 and utilize its signaling directly for tumor growth. The VEGFR-1 gene expresses two mRNAs of different lengths: the longer of approximately 8 kb, and the shorter of 2.5 to 3.0 kb. The short mRNA is highly expressed in normal placenta, encoding a soluble form of Flt-1 known as sFlt-1.[164] Within the placenta, trophoblasts located between the fetal and maternal blood vessel systems preferentially express sFlt-1. Thus, an interesting possibility is that sFlt-1 functions as a biochemical barrier between fetal and maternal circulation in the placenta by suppressing excess angiogenesis and abnormal vascular permeability. sFlt-1 was also found to be expressed in corneal epithelial cells.[165] This strongly suggests that sFlt-1 suppresses angiogenesis near the lens and maintains the transparency of the eye.

VEGFR-3 has a typical tyrosine kinase domain like other VEGFRs, and upon stimulation with VEGF-C, the PKC pathway and Ras pathway were reported to be activated for lymphangiogenesis. However, it remains to be clarified which autophosphorylation site(s) on the tyrosine residues in VEGFR-3 is responsible for these pathways and is critical for lymphangiogenesis.

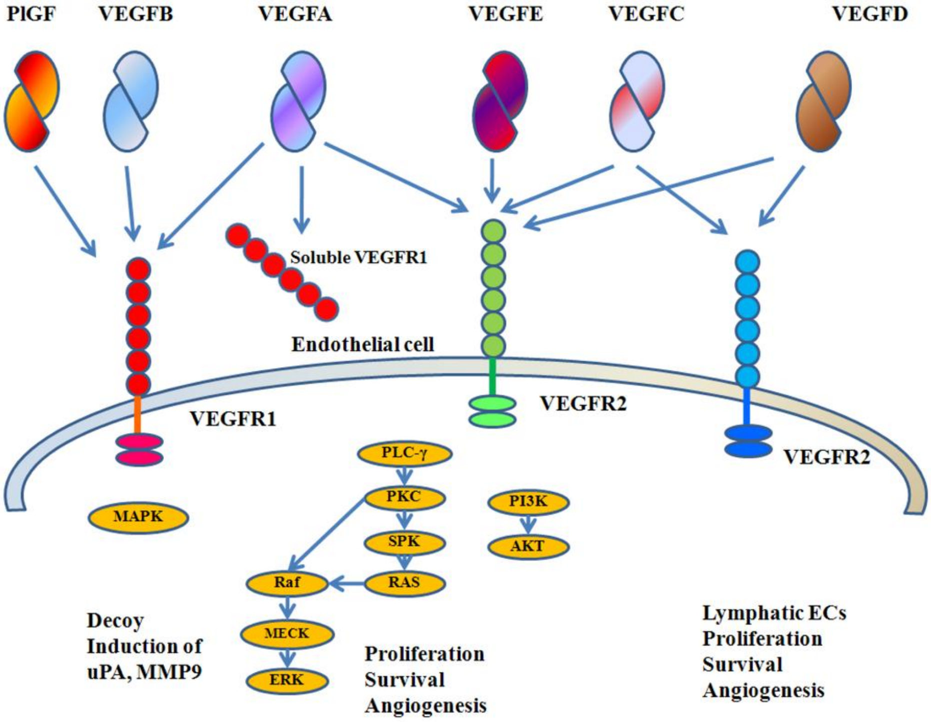


Figure 9. *Schematic view of VEGF family and receptors. VEGF-A binds both VEGFR-1 and VEGFR-2. PlGF and VEGF-B bind only VEGFR-1. VEGF-C and VEGF-D bind VEGFR-2 and VEGFR-3. VEGFR-2 is the major mediator of EC mitogenesis and survival. VEGFR-1 does not mediate an effective mitogenic signal in EC and it may perform an inhibitory role by sequestering VEGF-A and preventing its interaction with VEGFR-2.*[166]

In summary, VEGF and its receptors VEGFR-1 and VEGFR-2 play major roles in physiological as well as pathological angiogenesis, including tumor angiogenesis. VEGF-C/D and their receptor VEGFR-3 can regulate angiogenesis at early embryogenesis but mostly function as critical regulators of lymphangiogenesis.[157] VEGF, has a variety of functions, including pro-angiogenic activity, vascular permeability activity, and the stimulation of cell migration in macrophage lineage and endothelial cells (**Figure 9**).

Nowadays, inhibitors of the VEGF-VEGFR system, such as an anti-VEGF neutralizing antibody and multikinase inhibitors have been developed and widely used for the treatment of solid tumors.[167, 168] Bevacizumab (anti VEGF humanized monoclonal antibody) has been approved for the treatment of colorectal, breast, lung (non–small cell type), and renal cancers as well as for glioblastoma patients.[148, 169] Multikinase inhibitors such as sorafenib and sunitinib are now approved for renal and hepatic cancer patients. In addition to these medicines, others that target the VEGF-VEGFR system, including VEGF-Trap (a fusion protein of VEGFR-1 and VEGFR-2 ligand-binding domains), anti VEGFR-1 or anti VEGFR- 2 neutralizing antibody, soluble VEGFR- 3, VEGFR-1 or VEGFR-2 peptide vaccine therapy,[170] and anti-PlGF antibody,[171, 172] have been developed and are undergoing preclinical and clinical trials.

The clinical efficacy of these medicines has been well evaluated; even so, none of them provide a complete cure for cancer patients. The molecular basis of the refractoriness in some tumors and the acquisition of resistance to these medicines are some of the factors that need to be extensively studied to develop more efficient anti-angiogenic therapies.

Apart from cancer, VEGF is also implicated in other diseases such as the wet form of age-related macular degeneration. Localization of VEGF at sites of injury (choroidal neovascular membranes) strongly suggests its direct role in the progression of age-related macular degeneration related choroidal angiogenesis. Use of anti VEGF therapies for age-related macular degeneration has also been explored in clinical trials.[173]

In this regard, peptide antagonists acting both on VEGF or their receptors appear as an alternative to currently developed antibodies as well as pharmacological tools for studying the regulation of cellular phenomena associated with angiogenesis.[174, 175] The major difficulty when targeting VEGF with peptide ligands is that VEGF-VEGFR binding interface is flat, mainly hydrophobic, and rather large (>800 Å^2 per receptor unit),[160] which thereby suggests that targeting this interaction may be especially challenging. However, a compound interacting with the VEGF binding site in the VEGFR D2 domain and inhibiting complex formation was recently described.[176] This observation thus indicates that this PPI may be enable to disruption by synthetic compounds.

Information on the receptor binding epitope of VEGF may be useful to design VEGF-VEGFR inhibitors. In this sense it is known that several VEGF segments distant in sequence generate a discontinuous binding epitope that contributes to receptor binding. The binding determinants of VEGF for VEGFR are localized on the N-terminal α-helix 1 (residues 16–27), the loop connecting β3 to β4 (residues 61–66), and strand β7 (residues 103–106) of one VEGF monomer as well as strand β2 (residues 46–48) and strands β5 and β6, and the turn connecting them (residues 79–91).

In this chapter we evaluated the use of cyclic peptides as possible candidates to bind the VEGF protein in the receptor binding interface. We started evaluating an in-house made library called the EXORIS library. Some of these peptides were found to bind VEGF with mM affinity. The best candidate from the EXORIS library was selected to perform an exploratory process analyzing how different factor such as symmetry, amino acid replacement, residue configuration and ring size affect VEGF-binding. We also conducted a conformational analysis of a representative set of cyclic hexapeptides having different residue configurations. Different symmetric and non-symmetric β-turn conformations were observed by changing the chirality of the amino acids present in these cyclic hexapeptides.

De novo rationally designed library: EXORIS

Cyclic hexapeptides are probably the most extensively experimentally studied cyclic peptides. This is mainly owing to the ease of synthesis of cyclic hexapeptides and the ability of the cyclic hexapeptide backbone to take up well defined and stable structures. Based on the examination of the crystal structure of more than 100 synthetic cyclic hexapeptides it was early noticed that cyclic hexapeptides without sequence symmetry and those of the type (A_1-A_2-A_3)$_2$, where A_1, A_2 and A_3 are various amino acid residues, adopt conformations featured by two β-turns, many of which are stabilized by their associated hydrogen bonds.[177]

As it is described in the introduction of this thesis, mimetics of turn structures are frequently used for protein surface recognition and modulation of PPIs. In this regard, cyclic hexapeptides have recently demonstrated their potential as PPI modulators for several relevant targets.[178-180]

Encouraged by these considerations, Prof. Ernest Giralt, Dr. Meritxell Teixidó and Dr. Núria Bayó designed in our lab a small library of cyclic hexapeptides (EXORIS library) with the aim of exploring its potential in the recognition of protein surfaces. This library was synthesized by Dr. Nuria Bayó in collaboration with Dr. Monica Varese, and I performed the screening of these peptides as possible VEGF-binders and also the conformational analysis of a selected set of peptides.

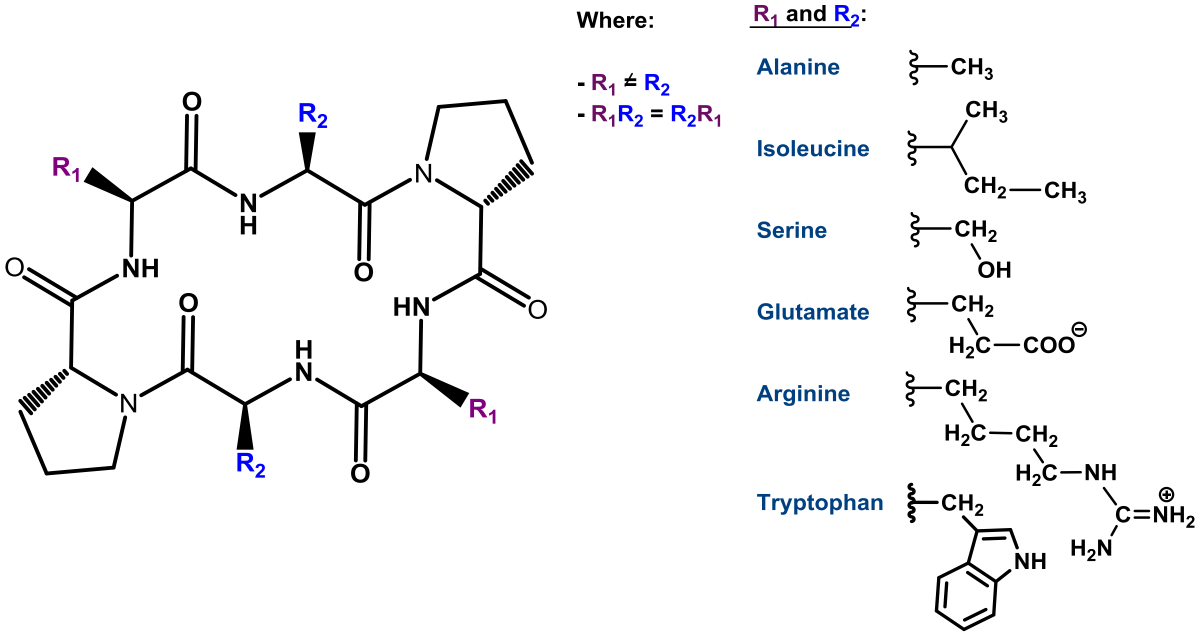


__Figure 10.__ Schematic representation of the EXORIS library peptides.

This library is composed by symmetric cyclic hexapeptides of the type c(A_1-A_2-A_3)$_2$, where A_2 are D-Proline residues, and A_1 and A_3 are residues from a selected reduced set of all proteinogenic amino acids. This reduced set of amino acids was rationally selected considering physicochemical, synthetic and statistical parameters such as propensity of each amino acid to a given secondary structure, amino acid hydrophobic properties, and their presence in protein active sites (**Figure 10**). This decision was based on a previous work developed by Dr. Ignasi Belda in our lab.[181] A summary of these physicochemical and statistical parameters is disclosed in the following table.

Table 1. *Summary of considered parameters for reduced amino acid-based selection.*

Amino acid	Secondary structure propensity[182]	Hydrophobicity[183]	Presence in active sites[184]
Tryptophan	Medium tendency for β-sheet	11	21
Arginine	Similar trend for β-sheet or α-helix	20	13.3
Alanine	High α-helical propensity	7	-
Glutamic acid	High α-helical propensity	17	3.6
Serine	Similar trend for β-sheet or α-helix	10	1.1
Isoleucine	High β-sheet propensity	1	9.6

As it is shown in **Table 1**, this reduced set of amino acids includes hydrophilic, hydrophobic, charged and neutral residues, as well as three of the top amino acids present in protein active sites (tryptophan, arginine and isoleucine).

D-proline was selected as it is the only proteinogenic N-alquil amino acid and it has a tendency to facilitate β-turns. In addition, prolines often participate on crucial hydrophobic-hydrophobic interactions. Tryptophan was selected because it is known to be over-represented at protein interfaces and it is found on many hot spots. Arginine (pI = 10.76) displays, as lysine, a positive charge at physiological pH. However, significant differences in the hydration energy and hydrogen bonding potential of these cationic amino acids make arginine more favorable to establish inter-residue contacts. Alanine is a highly abundant residue in natural systems with high α-helical propensity, and since it has a small lateral side chain it can act as spacer for distinct interaction sites on peptides. Glutamic acid contributes on negative charges and has tendency to form α-helical structures. Serine is a polar non-charged amino acid that behaves better than threonine on SPPS. Finally, isoleucine is an aliphatic β-branched amino acid with a significantly bulky side chain. It has a high tendency to form β-sheet structures and has a higher presence in actives sites than valine and leucine.

All the peptides from the EXORIS library are described in **Table 2**. One letter code abbreviations will be used through this chapter; L- and D-amino acids are represented by capital and low-case letters, respectively.

Table 2. *Description of the EXORIS library peptides.*

Peptide code	Sequence	
	One letter code	**Three letter code**[185]
1	c(RpARpA)	&Arg-D-Pro-Ala-Arg-D-Pro-Ala&
2	c(RpIRpI)	&Arg-D-Pro-Ile-Arg-D-Pro-Ile&
3	c(RpWRpW)	&Arg-D-Pro-Trp-Arg-D-Pro-Trp&
4	c(EpREpR)	&Glu-D-Pro-Arg-Glu-D-Pro-Arg&
5	c(EpAEpA)	&Glu-D-Pro-Ala-Glu-D-Pro-Ala&
6	c(EpIEpI)	&Glu-D-Pro-Ile-Glu-D-Pro-Ile&
7	c(EpWEpW)	&Glu-D-Pro-Trp-Glu-D-Pro-Trp&
8	c(EpSEpS)	&Glu-D-Pro-Ser-Glu-D-Pro-Ser&
9	c(SpISpI)	&Ser-D-Pro-Ile-Ser-D-Pro-Ile&
10	c(SpRSpR)	&Ser-D-Pro-Arg-Ser-D-Pro-Arg&
11	c(SpWSpW)	&Ser-D-Pro-Trp-Ser-D-Pro-Trp&
12	c(ApSApS)	&Ala-D-Pro-Ser-Ala-D-Pro-Ser&
13	c(ApWApW)	&Trp-D-Pro-Ala-Trp-D-Pro-Ala&
14	c(IpAIpA)	&Ile-D-Pro-Ala-Ile-D-Pro-Ala&
15	c(IpWIpW)	&Ile-D-Pro-Trp-Ile-D-Pro-Trp&

The potential of this library was already demostrated in a previous work conducted by Dr. Xavier Vila (from Ernest Giralt's lab) in collaboration with the Hospital Virgen del Rocío in Seville. They identified a cyclic hexapeptide from the EXORIS library able to inhibit the virulence of several *Acinetobacter Baumannii* strains both *in vitro* and *in vivo* trough a novel mode of action based on blocking the binding of bacteria OmpA to host-cell fibronectin.[186]

As part of one of the main ongoing projects in the group of Prof. Giralt, focused on the search of potential modulators of the VEGF-VEGFR interaction, we decided to assess the ability of the EXORIS peptides to recognize the receptor-binding site of VEGF.

The screening of this peptide library against VEGF was done by nuclear magnetic resonance (NMR) spectroscopy. NMR methods able to detect binding of small molecules to biomolecular targets can be classified as ligand- or protein-based methodologies depending of the NMR signals observed in the experiment. The advantage of protein-based methods is that when the assignment of the protein spectrum is available the residues participating in the interaction with the ligand can be easily identified.

Protein-based methods require proteins enriched in NMR-active isotopes, such as ^{15}N or ^{13}C, which are usually obtained by recombinant protein expression. The expression and purification of His-tagged ^{15}N-uniformly labeled VEGF$_{11\text{-}109}$ was already described by Fairbrother et al.[161] Nevertheless, in a previous work in our lab, it was observed that addition of v107 (GGNECDIARMWEWECFERL), a peptide known to bind VEGF in the receptor-binding site, caused considerable perturbations in more than 50% of the ^{1}H-^{15}N HSQC cross-peaks of VEGF. The drastic v107-dependent changes in the ^{1}H-^{15}N HSQC spectra of VEGF may be misleading, hindering the identification of the interaction region on the protein. In addition, this method requires high amounts of ^{15}N-labeled protein and may be time consuming for screening purposes.

Selective isotopic labeling of specific atoms of the target protein may alleviate some of these concerns. Particularly, when NMR-active isotopes are introduced into amino acid side chains, one can obtain utterly simple spectra, retaining most of the information and, at the same time, greatly reducing protein related costs. Moreover, since side chains are less prone to suffer from long-range chemical shift perturbation effects, they may be more reliable than backbone atoms when extracting structural information.

We took advantage of a methodology previously developed in our group by Dr. Ricard Rodríguez.[187] He used ^{13}C-methyl-methionine labeling and obtained high-quality protein NMR spectra observing only five specific signals corresponding to the VEGF methionines (positions 18, 55, 78, 81, and 94), which significantly simplifies the use of NMR experiments for screening purposes. In our case, this labeling strategy benefits from the fortunate location of methionine residues in the VEGF structure. Two out of the five VEGF methionine residues, Met18 and Met81, are known to play relevant roles in complex formation and stabilization. Met81 and Met18 become deeply buried upon interaction with the VEGFR1 receptor or other previously described binding peptides.[188] Moreover, mutation of Met18 or Met81 has deleterious consequences for both complexes.

Beyond their undeniable privileged location in VEGF, from a practical perspective methionine residues are also attractive for labeling purposes. First, introduction of stable isotopes into its thioether methyl group is chemically uncomplicated. Furthermore, from the NMR point of view, ^{13}C-methyl labeling is very appealing due to its high sensitivity and advantageous relaxation properties. In addition, incorporation of labeled methionine into proteins can be easily achieved by expressing recombinant proteins in methionine-auxotrophic *E. coli* strains.

The activity of selectively ^{13}C-methyl- methionine labeled VEGF was previously assessed by Dr. Ricard Rodríguez. He analyzed the perturbations induced in the ^{1}H-^{13}C HSQC spectra of VEGF upon addition of the v107 peptide (**Figure 11**).[189]

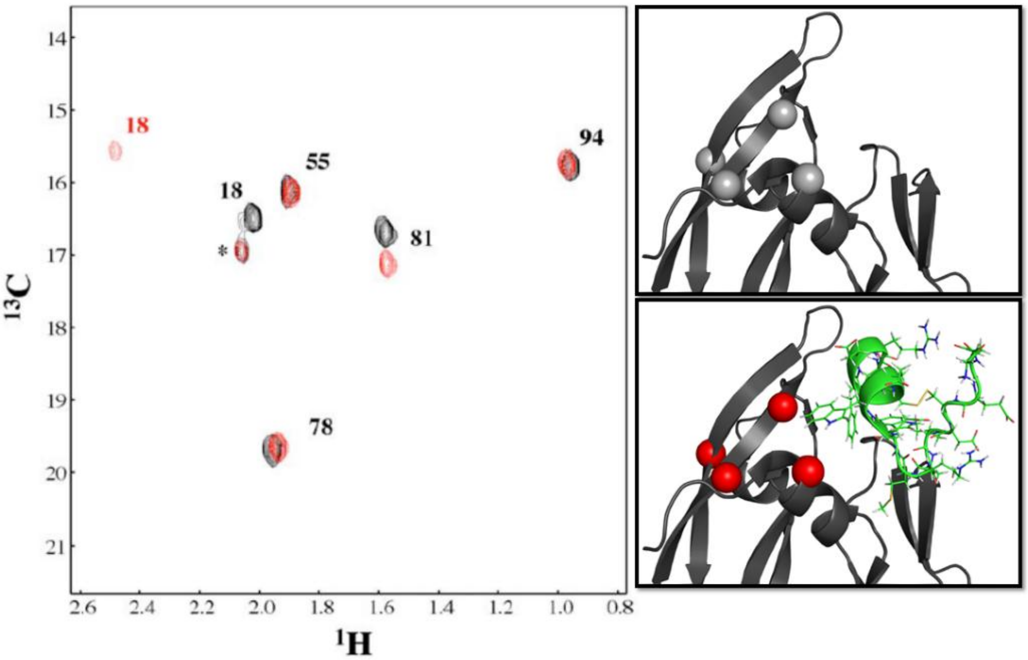


Figure 11. *^{1}H-^{13}C HSQC spectra of methyl-^{13}C-Met VEGF$_{11\text{-}109}$ showing the assignment of cross-peaks in the absence (black) and in the presence of 2 eq. of peptide v107 (red). Spectra were acquired at 298 K in 25 mM phosphate buffer (pH 7.4), 50 mM NaCl. (*) Cross-peak corresponds to the methionine residue encoded by the starting codon of (His)$_6$-VEGF construct from residual VEGF with uncleaved His-tag. On the right, free and v107-complexed VEGF cartoon structures are depicted. Methionine residues are represented as black (free) or red (complex) spheres. Peptide v107 is shown in green.*

Resonance assignment using published data[188] revealed that, in concordance with their participation in VEGF-v107 intermolecular contacts,[188] the VEGF most affected signals by peptide addition corresponded to methyl groups in Met18 and Met81. Met18 exhibited the largest shifts, probably as a result of important contacts with Trp13 in the v107 peptide and the effect of its aromatic ring currents. Changes on Met81's methyl chemical shift were mainly restricted to ^{13}C chemical shift. Finally, binding effects on the rest of methionine residues were practically negligible.

The evaluation of the EXORIS peptides as possible VEGF-binders was done comparing the ^{1}H-^{13}C HSQC spectra of 50 µM methyl-^{13}C-Met VEGF$_{11\text{-}109}$ in the absence and presence of 0.5 mM cyclic peptide as shown in **Figure 12**. The peptide induced-chemical shift perturbation (CSP) on Met18 and Met 81 resonance were quantified using **Equation 1**.

$$CSP = \sqrt{(\Delta\delta_H)^2 + \left(\frac{1}{4}\Delta\delta_C\right)^2} \quad \textbf{Equation 1}$$

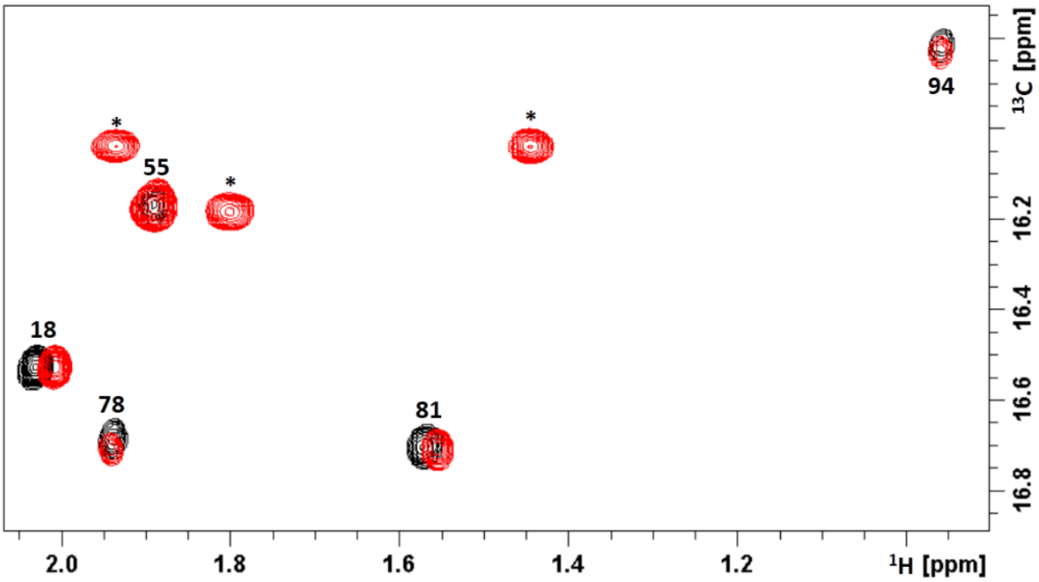


Figure 12. *Overlay of the ^{1}H-^{13}C HSQC spectra of 50 µM methyl-^{13}C-Met VEGF$_{11\text{-}109}$ in the absence (black) and in the presence (red) of 0.5 mM c(EpWEpW) peptide. Spectra were obtained at 308 K in 25 mM phosphate buffer (pH 7.4) 50 mM NaCl, 0.5% DMSO-d$_6$. Resonances assignment of VEGF are shown. *Cross-peaks corresponding to natural abundance ^{13}C peptide signals.*

Most EXORIS peptides caused almost negligible effects on the VEGF spectra. Only those peptides containing tryptophan residues (**Figure 12** and **Figure 13**), produced small but reproducible chemical shift changes on the methyl resonances of Met18 and Met81 of VEGF. These changes were mainly restricted to ^{1}H chemical shifts. Interestingly, no significant perturbations were observed in the remaining three VEGF methionine residues (**Figure 12**). This behavior suggests that only tryptophan-containing residues within the EXORIS library are able to bind VEGF, most probably in the receptor binding site, but with low affinity. These observations are in agreement with the key role of tryptophan or other aromatic residues in both the v107 peptide and the Flt1 receptor in the recognition of VEGF.[188]

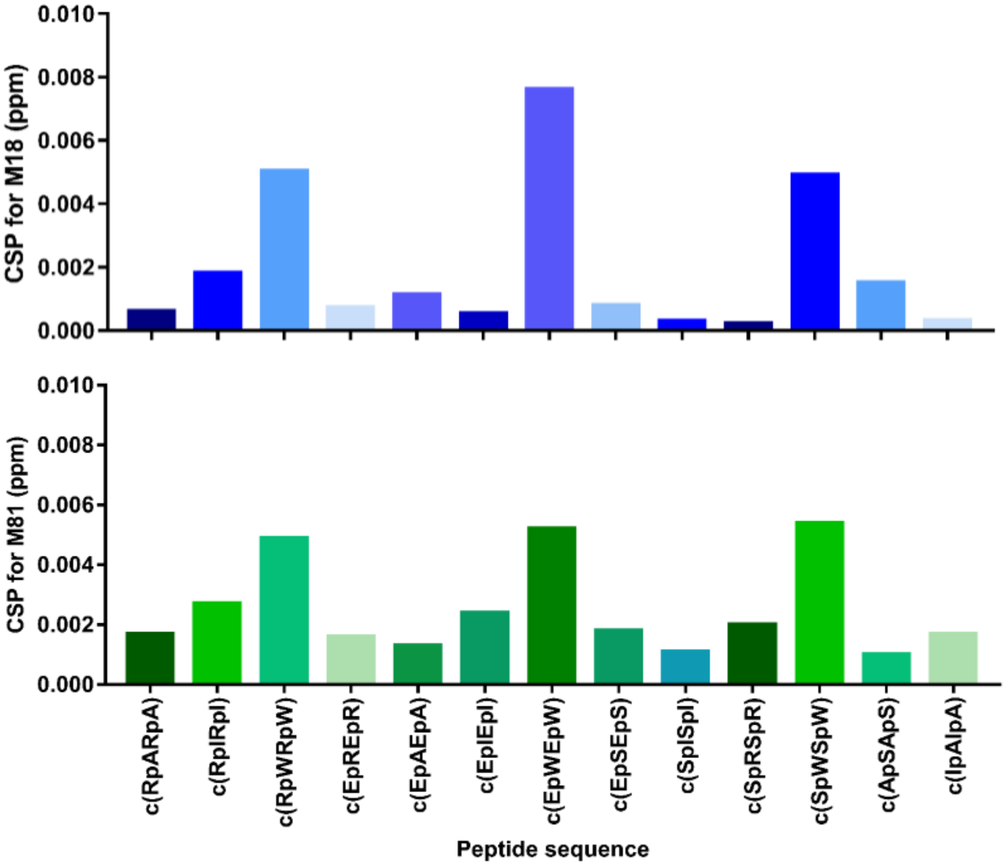


Figure 13. *Graphical representation of the CSP obtained for selectively labeled VEGF after the addition of 0.5 mM of the EXORIS library peptides. Results obtained for: Met18 (blue), and Met81 (green). Peptides IpWIpW and WpAWpA could not be evaluated due to solubility restrictions.*

In order to confirm the importance of tryptophan residues for VEGF binding, we selected a peptide with Trp (peptide 7 – c(EpWEpW)) and another that does not contain Trp (peptide 8 – c(EpSEpS)) to perform saturation transfer difference (STD) NMR. STD experiments were acquired at 278 K in D_2O on samples containing 1 mM of cyclic peptide and 10 µM VEGF. The on-resonance irradiation was set at 0 ppm, in a spectral region were only protein resonances are present. An identical STD was performed on a peptide sample without VEGF. In the presence of protein, peptide 7 displayed strong STD signals, especially in the aromatic region of the spectrum (**Figure 14b**). The observed STD signals are an indication of binding since are produced by the transfer of magnetization from the saturated protein to the peptide in its bound state. In the absence of VEGF no significant STD signals were observed (**Figure 14a**) confirming that the STD effects were produced by the interaction of the peptide with the protein.

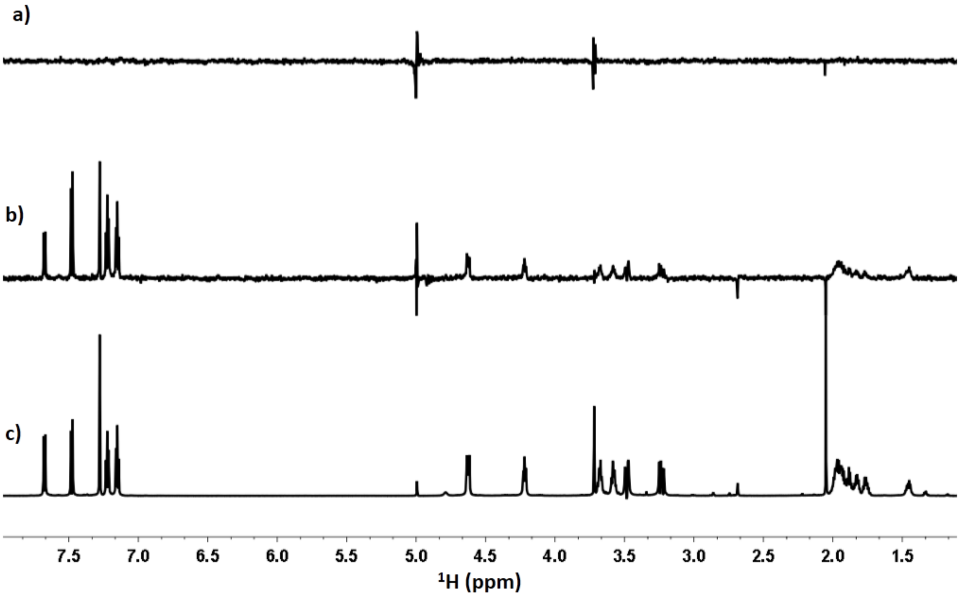


Figure 14. *STD experiments for peptide c(EpWEpW). Off-resonance (c) and STD spectra (b) of peptide 7 in the presence of VEGF. As a control, the STD spectrum of the peptide in the absence of VEGF (a) is shown. Compared to the STD spectra, the vertical scale of the off-resonance spectrum has been reduced by a factor of 5. The on-resonance spectra of the STD were irradiated 2 seconds at 0 ppm.*

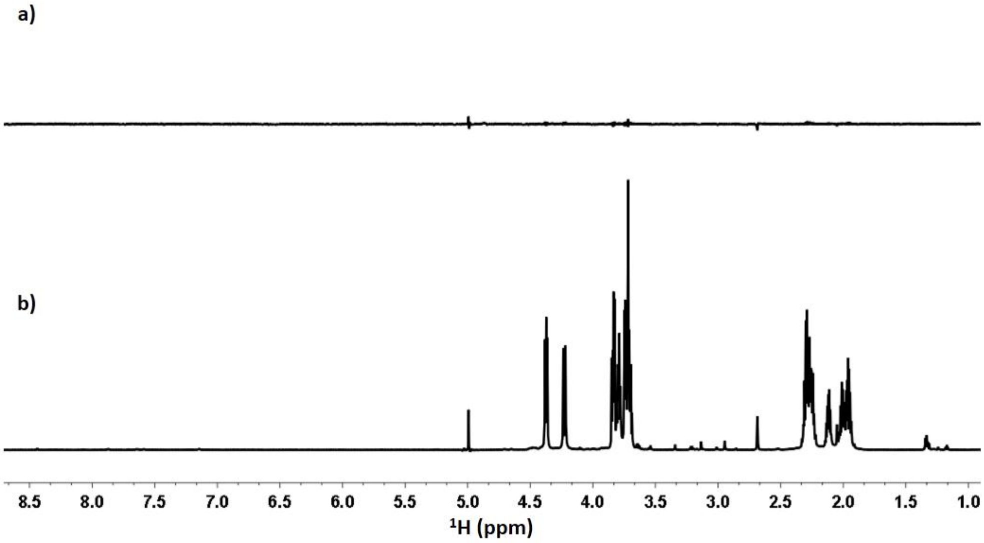


Figure 15. *STD experiments for peptide c(EpSEpS). Off-resonance (b) and STD spectra (a) of peptide 8 in the presence of VEGF. Compared to the STD spectrum, the vertical scale of the off-resonance spectrum has been reduced by a factor of 5. The on-resonance spectra of the STD were irradiated 2 seconds at 0 ppm.*

When the STD experiments were performed with peptide 8 we could not observe any STD signals in the presence of VEGF, confirming that, in this case, magnetization was not transferred from the protein to the peptide and, therefore, there was not binding interaction between them (**Figure 15**).

Once binding of some EXORIS peptides was confirmed by protein and ligand based NMR experiments, we decided to perform a more detailed analysis of the interaction of VEGF with a representative peptide from the EXORIS library. We selected peptide 7 (c(EpWEpW)) because it caused the largest CSP in the VEGF Met18 and 81 resonances, and also because peptide v107 contains the motif "WEWE" (similar to that of peptide 7), where tryptophan residues are important for binding VEGF while glutamates are completely solvent exposed.

Selective perturbation of Met18 and Met81 in the presence of peptide 7 (**Figure 12**) suggests that it binds to the VEGF receptor binding site. However, to further characterize its binding region on the protein surface, we monitored the perturbations induced by this peptide in the ^1H-^{15}N HSQC spectrum of a ^{15}N-uniformly labeled VEGF sample. ^1H-^{15}N HSQC spectra display a cross-peak for each amidic NH (for each non-proline VEGF residue). Comparison of the ^1H-^{15}N HSQC spectra of VEGF in the absence and presence of peptide allows to detect those residues that are selectively perturbed by the ligand. Mapping the most affected residues on a three dimensional structure of VEGF provides information about the peptide binding region. CSP were calculated using **Equation 2**, where γ_H and γ_N are the gyromagnetic constant of ^1H and ^{15}N nuclei, respectively.

$$\boldsymbol{CSP} = \sqrt{(\Delta\boldsymbol{\delta_H})^2 + \left(\frac{\mathbf{1}}{\mathbf{5}}\Delta\boldsymbol{\delta_N}\right)^2} \qquad \textbf{Equation 2}$$

Most affected residues after the addition of the cyclic peptide were observed near the N-terminal α-helix (α1) and in the vicinity of β5- and β6-strands, reported to be important in intermolecular interactions.[188] Amino acids close to β2-strand, and also those near β4- and β7-strands, were also affected by the presence of these peptides, and although they fall far away from β5- and β6-strands in the VEGF primary structure; on the three dimensional structure, we could see that they belong to the same β-sheet structural motif (**Figure 16a** and **b**). Consequentlly, it may be that the perturbations observed in these two clusters are coupled and that these effects may be transmitted at a relative long distance through the β-sheet motif. On the basis of these results, we decided to perform docking simulations by computational tools to predict the binding mode of peptide 7 on the VEGF surface. For this purpose, Dr. Jesús Seco conducted virtual docking experiments using the crystal structure of VEGF obtained in complex with VEGFR-1 (PBD: 1FLT).

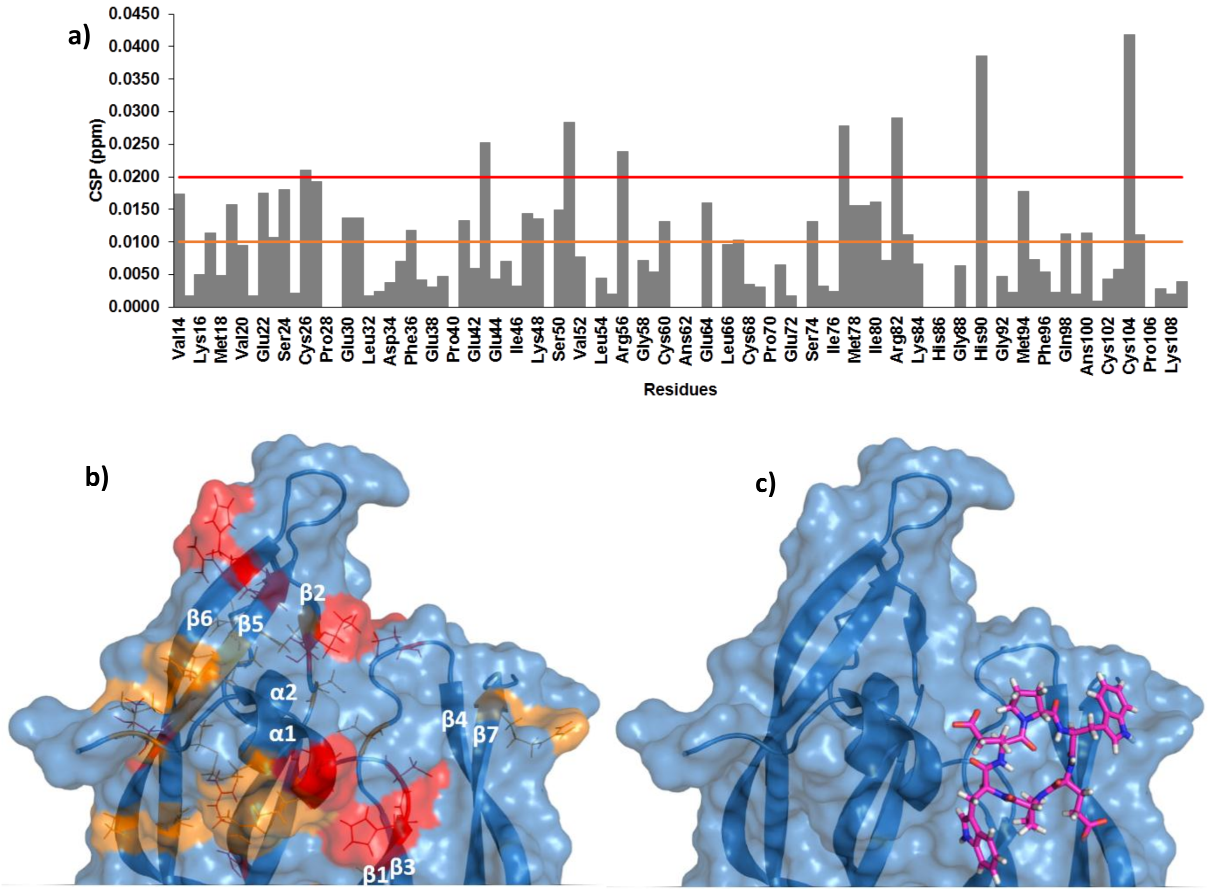


__Figure 16.__ a) Bar graphs of the average combined chemical shift differences ($\Delta\delta_{HNav}$) induced by peptide 7 as a function of VEGF sequence. The mean value (0.01 ppm) is shown as an orange line; the mean value plus standard deviation (0.02 ppm) is shown as a red line. b) Cartoon representation of the structure of the receptor binding epitope of VEGF. Most affected residues in the presence of peptide 7 are depicted in red ($\Delta\delta HN_{av}$ > 0.02 ppm) or orange $\Delta\delta HN_{av}$ > 0.01 ppm); c) Binding mode predicted for peptide 7. VEGF is shown as cartoon representation in blue. Peptide 7 is shown as sticks in pink.

The docking simulation showed that peptide 7 binds VEGF in a region of VEGF that is also known to interact with VEGFRs as it can be observed in **Figure 16c**. These results are in agreement with those obtained experimentally by CSP experiments (**Figure 16b**), as important chemical shift changes were observed in the region comprised by the N-terminal α-helix and β4 and β7 strands. Chemical shift changes observed in the regions close to β5 and β2 by NMR, which seemed to be far from the binding site predicted by virtual docking, can be explained due to the intrinsic flexibility of VEGF, which can place those residues close to peptide 7. Alternatively, as our NMR experiments were obtained in the presence of 10 equivalents of peptide, several peptide molecules may bind simultaneously on different regions of the VEGF surface.

All these results suggest that peptide 7 binds VEGF near the receptor binding epitope. Nonetheless, CSP observed both in the ^{1}H-^{13}C HSQC and ^{1}H-^{15}N HSQC experiments were relatively small, suggesting that this peptide may bind VEGF with low affinity.

In order to determine the binding affinity of this compound for VEGF we tried different methodologies. First attempts were done using a described competition fluorescence polarization (FP) assay.[190] This methodology can be used to identify molecules that bind to the VEGF dimer in the region that is recognized by receptors and to obtain a K_D value for this binding event. It requires a relatively small, fluorescent ligand ("tracer") for VEGF that could be displaced by candidate inhibitors. A previously described v107-derived tracer was used in these experiments.[190] The v107-derived fluorescein-labeled tracer peptide together with the unlabeled v107 precursor peptide were synthesized by Dr. Monica Varese.

We first confirmed binding of the tracer peptide to VEGF by a direct assay adding increasing concentrations of VEGF to a fixed concentration of tracer peptide (67 nM). This assay provided a K_D of 146.1 ± 15.9 nM, in agreement with the values reported for this tracer.[190] We also assessed the capacity of v107 to displace our tracer. For this purpose we added increasing amounts of unlabeled v107 to fixed concentrations of VEGF (1 µM) and tracer (67 nM). The obtained FP values were plotted against unlabeled v107 peptide concentration and the resulting competition curve was fitted to the "One site-Fit Ki" equation from GraphPad Prism (**Figure 17**). Unlabeled v107 displayed a Ki value that was in agreement with that previously reported.[190]

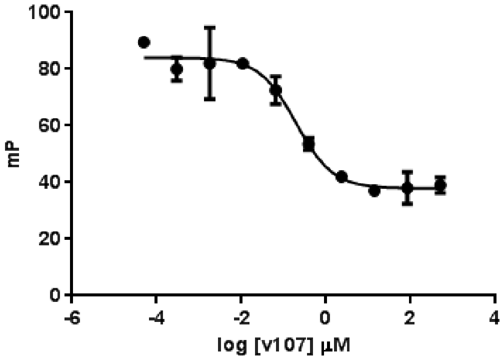


***Figure 17**. Competitive displacement of v107. Ki = 0.13 ± 0.04 µM*

However, when we tried to estimate the binding affinity of peptide 7 using the competition FP assay, we did not observed significant changes in fluorescence polarization upon addition of the EXORIS peptide, even at high concentrations (0.5 mM) meaning that the peptide was not able to displace the tracer. When higher peptide concentrations (2.5 mM) were used, we did not observe the typical behavior of a competition assay but that of a direct binding experiment, suggesting that the cyclic hexapeptide was aggregating with the tracer, which causes an increase in the fluorescence polarization. We concluded that competition assay by FP was not suited to obtain the affinity for the interaction between peptide 7 and VEGF. If we wanted to try a direct FP assay we should label the cyclic hexapeptide with a fluorescent dye, and taking into account the size of these peptides, this labeling could significantly affect the binding properties of these compounds.

As an alternative, we also tried microscale thermophoresis (MST). However, in our hands, this methodology did not provide reproducible results for this system. After various trials we hypothesized that the observed variations between experiments could be due to non-homogenous labeling of the VEGF protein. Most probably, this happened because dye introduction was done by means of Lys (primary amines) reaction. Looking at the three dimensional structure of VEGF we noticed that all lysine residues were close to the receptor binding epitope, so labeling with dyes in these positions could also affect the VEGF binding (**Figure 18**).

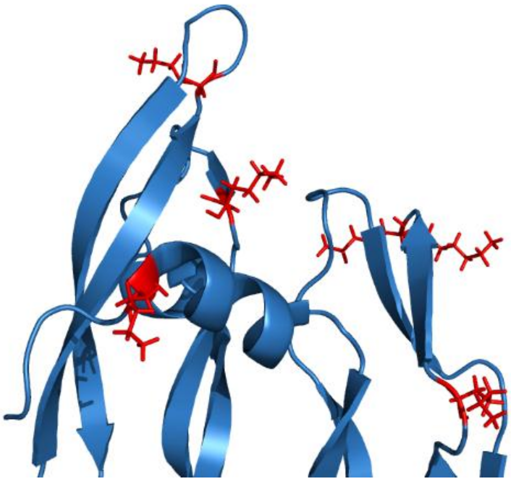


Figure 18. *Cartoon representation of the receptor binding site of VEGF. Lys residues are shown as sticks colored in red.*

We could also use non-label MST experiments taking advantage of the fact that VEGF does not contain tryptophan residues in its sequence, while the selected EXORIS peptides do have tryptophan residues. Moreover, these experiments need big amounts of protein and that was, in our case, a limitation as protein production was time-consuming and expensive.

At that point, we decided to use NMR to estimate the K_D value for the interaction of peptide 7 with VEGF. We performed a titration assay adding increasing concentrations of peptide stock in DMSO-d_6 to a sample of 50 µM methyl-^{13}C-methionine VEGF in 25 mM phosphate buffer (pH 7.4), 50 mM NaCl. We used DMSO-d_6 to improve peptides solubility as its interaction with VEGF was described by Dr. Ricard Rodriguez to be very low.[187] We worked below 5% DMSO (v/v) and in his thesis, he concluded that it should be safe to use DMSO concentrations below 10% (v/v).

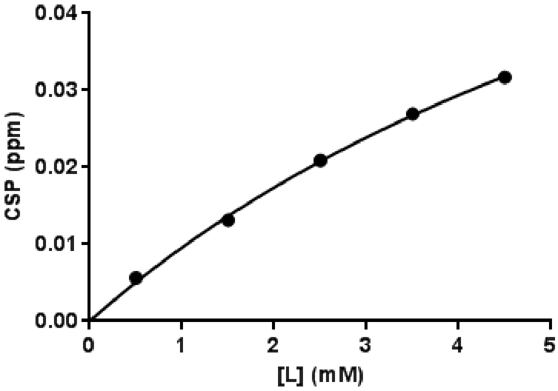


Figure 19. *Peptide induced chemical shift changes in the VEGF Met18 resonance. Titration of 50 µM VEGF in 25 mM phosphate buffer (pH 7.4), 50 mM NaCl, at 298 K; with increasing concentrations of EXORIS peptide c(EpWEpW).* K_D *= 9.0 ± 1.4 mM, Bmax = 0.09 ± 0.01 ppm,* R^2 *= 0.9984.*

Peptide-induced chemical shift changes in the VEGF Met18 resonance, which is the methionine that showed largest CSP, were plotted against ligand concentration and fitted to a “One site-specific binding” equation (GraphPad Prism software). As it can be noticed in **Figure 19**, we could not reach saturation even working at very high concentrations of peptide (4.5 mM). The K_D value calculated for this compound confirmed that peptide 7 was a VEGF binder with low mM affinity. Analogous titrations were performed for other EXORIS peptides that induced similar CSP on the Met18 signal in the ^{1}H-^{13}C HSQC spectra of VEGF, but again the estimated K_D values were in the low mM range.

Conformational analysis of EXORIS library peptides

In order to obtain a more complete characterization of the EXORIS peptides, we decided to perform a conformational study of a representative set of cyclic peptides, analyzing the effects of different amino on the conformation of these peptides in solution. Peptides showed in **Table 3** were selected for this study.

Table 3. *Selected peptides for its conformational study. Peptide 3-P1 was not included In the EXORIS library but was synthesized to study the effect of an aromatic residue previous to or following the D-Pro.*

Peptide code	Sequence	Observations
3	c(RpWRpW)	Aromatic + positive charge
6	c(EpIEpI)	Hydrophobic β-branched + negative charge
7	c(EpWEpW)	Aromatic + negative charge
11	c(SpWSpW)	Aromatic + polar
13	c(ApWApW)	Aromatic + hydrophobic
14	c(IpWIpW)	Aromatic + hydrophobic β-branched
3-P1	c(WpRWpR)	Changing aromatic position

1D ^{1}H NMR and standard 2D experiments such as TOCSY, DQF-COSY, NOESY and ^{1}H-^{13}C HSQC were acquired for peptide characterization. After complete assignment, we noticed that only three individual spin systems were obtained for all these compounds, suggesting a C2-symmetric conformation. Structure sensitive NMR parameters such as deviations from random coil chemical shift, temperature coefficient values of amide NH protons, $^3J_{\alpha NH}$ coupling constants and NOE patterns were evaluated to study the conformational behavior of these peptides. Peptides 13 and 14 could not be studied due to solubility restrictions.

An empirical relationship between protein backbone structure and deviations of chemical shifts from random coil values, so called secondary shifts, has long been recognized.[191] This magnitude is an indicator of non-random conformations and it is frequently used as a tool for the detection of secondary structure in biomolecules.[192]

As it can be noticed in **Figure 20**, the chemical shift deviation values do not show a clear trend, as they were very small, and the observed values were different for those reported for α-helix (< -0.4 and > 2 ppm, for proton and carbon respectively) or β-sheet (> 0.4 and < 2 ppm, for proton and carbon respectively). We could only conclude that all these peptides are pretty similar regarding their conformations.

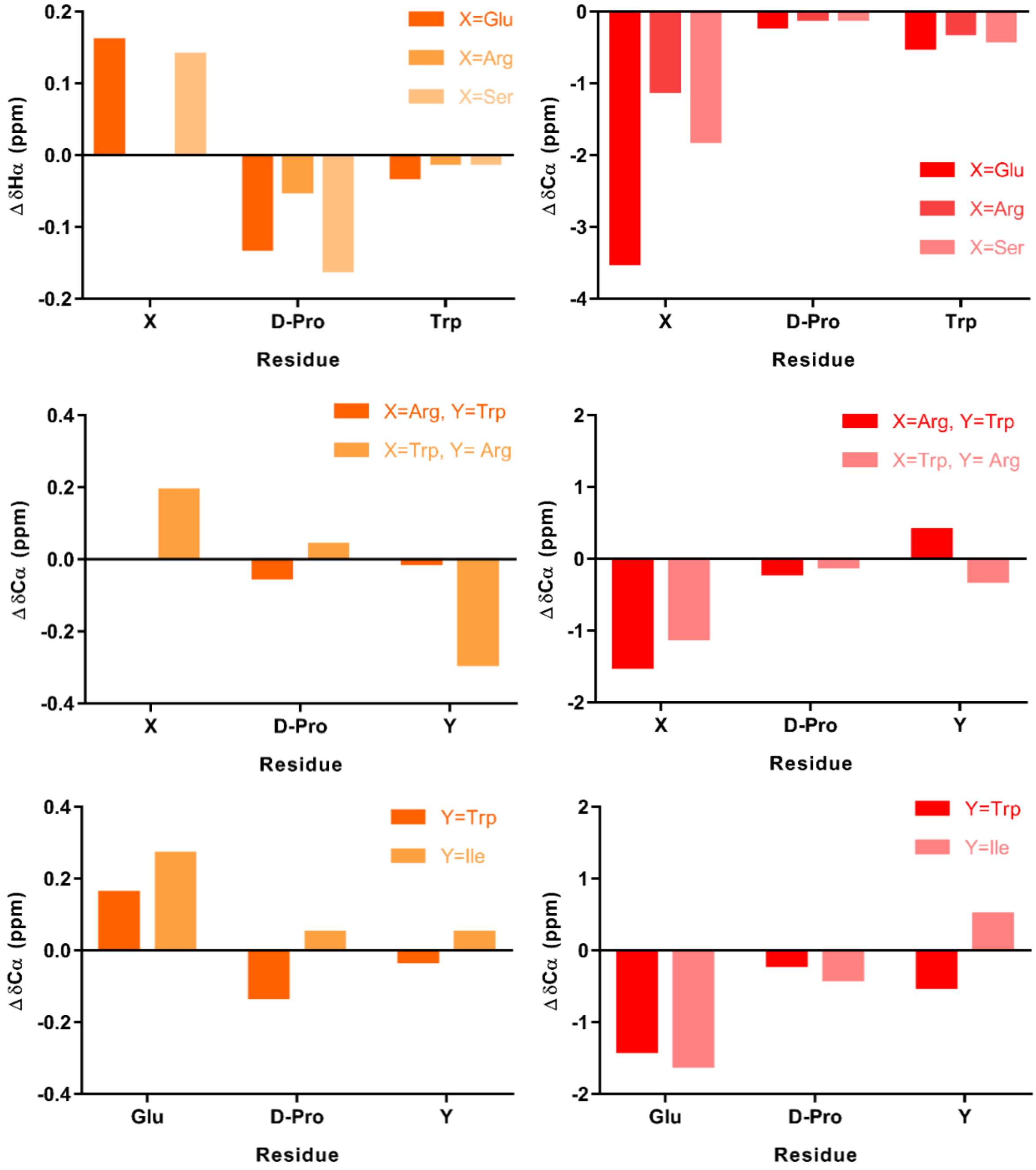


***Figure 20**. Histogram representations of ΔδHα (orange) and ΔδCα (red) comparing the effect of: Top) different amino acids preceding and following the D-Pro residue; Middle) cycloisomers, and; Bottom) The presence and absence of aromatic residue after the D-Pro. Random coil values were extracted from Wishart et al.*[193]

Unlike other peptide bonds, which exist almost exclusively in the *trans* conformation, Xaa-Pro bonds may also significantly populate the *cis* isomer. As EXORIS peptides contained two Xaa-D-Pro bonds, we decided to determine their conformation using a method based on ^{13}C chemical shift statistics.[194] This method analyzes the chemical shift difference between $^{13}C_{\beta}$ and $^{13}C_{\gamma}$ of the proline residue. It is known that variations in chemical shifts of proline $^{13}C_\beta$ and $^{13}C_\gamma$ within the region associated with a Xaa-Pro bond geometry can be correlated with *cis* or *trans* conformations of the imide bond. Results obtained for the studied peptides are shown in **Table 4**.

Table 4. *Chemical shift difference between $^{13}C_\beta$ and $^{13}C_\gamma$ of the proline residue for the studied peptides.*

Peptide code	Sequence	δC_β (ppm)	δC_γ (ppm)	$\Delta\delta C_{\beta\gamma}$ Pro (ppm)
3	c(RpWRpW)	31.0	27.2	3.9
6	c(EpIEpI)	31.4	26.4	4.0
7	c(EpWEpW)	30.7	27.1	3.7
11	c(SpWSpW)	30.8	27.2	3.6
3-P1	c(WpRWpR)	31.2	27.1	4.1

We observed that the $\Delta\delta C_{\beta\gamma}$ values obtained for the proline residue were around 4 ppm, which indicates that in these peptides the Xaa-Pro bonds are in the *trans* conformation. These results were also confirmed by NOESY experiments, where we could observe an intense cross-peak between the H_δ protons of proline and the H_α proton of the preceding glutamic residue (**Figure 21**).

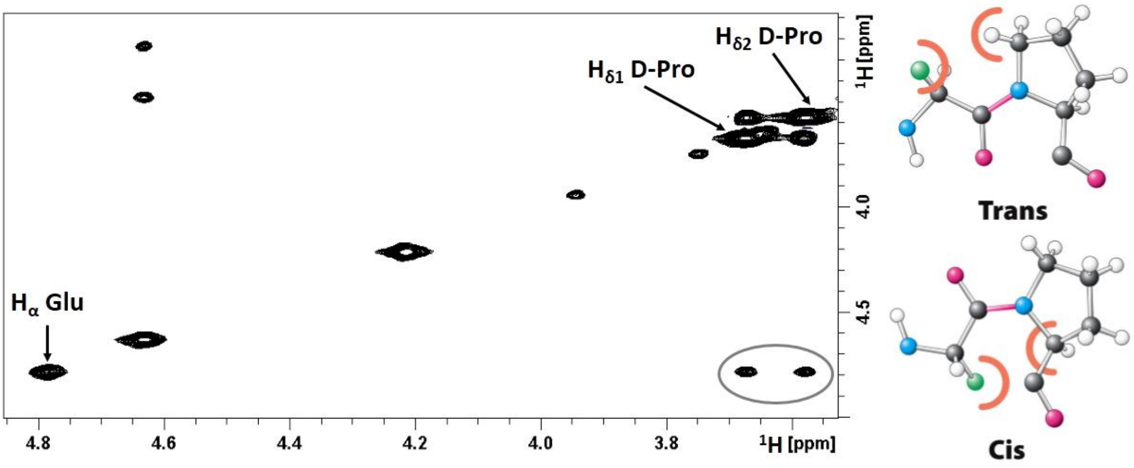


Figure 21. *Left) Expanded region of the 2D NOESY spectrum of peptide 7 at 278 K, in 25 mM phosphate buffer (pH 7.4), 50 mM NaCl. The NOE cross-peak between the H_δ of proline and the H_α of the preceding glutamic residue is indicated. Right) Schematic representation of the trans and cis conformations of the Xaa-Pro imide bond in peptides.*

To explore the presence of hydrogen bonds in these peptides, a series of 1D ^{1}H-NMR spectra were acquired at 5 or 10ºC intervals from 278 to 305 K, and temperature coefficient values ($-\Delta\delta_{NH}/\Delta T$) were obtained for each backbone amide proton. These values are widely used as an indicator of solvent accessibility to amide protons. Small temperature coefficients ($0 < -\Delta\delta_{NH}/\Delta T < 5$ ppb/K) are related to low accessibility and are often interpreted as evidence for intramolecular hydrogen bonds; compared to $6 < -\Delta\delta_{NH}/\Delta T > 10$ ppb/K, for solvent accessible amide protons.

Table 5. *Temperature coefficient values for the studied peptides (c(Xaa-D-Pro-Yaa)$_2$)*

Peptide code	Sequence	$-\Delta\delta_{NH}/\Delta T$ (ppb/K)	
		Xaa	Yaa
3	c(RpWRpW)	4.5	8.0
6	c(EpIEpI)	3.5	8.0
7	c(EpWEpW)	1.5	7.7
11	c(SpWSpW)	1.0	8.0
3-P1	c(WpRWpR)	2.8	8.5

As it can be noticed, in all the studied peptides, low temperature coefficient values were observed for residues preceding the D-proline residue, suggesting implication of these amide protons in hydrogen bonds, as they are solvent protected. On the contrary, residues following the D-proline showed high values indicating solvent exposure (**Table 5**). Finally, we also analyzed $^3J_{\alpha NH}$ coupling constants for these peptides.

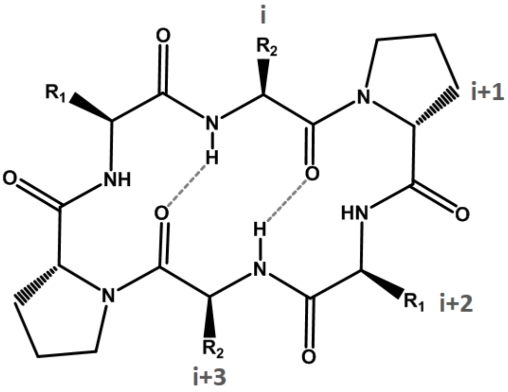
Table 6. *Coupling constants of Xaa and Yaa in the studied peptides c(Xaa-D-Pro-Yaa)$_2$. b.s = broad signal.*

Peptide code	Sequence	$^3J_{\alpha NH}$ (Hz)	
		Xaa	Yaa
3	c(RpWRpW)	8.5	b.s.
6	c(EpIEpI)	9.0	8.2
7	c(EpWEpW)	8.1	7.7
11	c(SpWSpW)	8.7	7.7
3-P1	c(WpRWpR)	9.6	b.s.

The observed coupling constant values fall in the range 7.7 - 9.6 Hz and would normally be considered to represent averages over several conformations (**Table 6**). In linear peptides, values around 6-7 Hz are normally interpreted as result of conformational average of very different conformations. Because of the constrained nature of cyclic peptides, in the current case, these coupling constants should more probably be the result of a well-defined conformation or average over a range of similar conformations.

As mentioned before, cyclic hexapeptides with no sequence symmetry and those of the type (A_1-A_2-A_3) where A_1, A_2 and A_3 account for various amino acid residues, adopt conformations featured by two β-turns, many of which are stabilized by their associated hydrogen bonds.[177]

β-turns are actually consistent with the results obtained in this study, as they involve four residues and are generally stabilized by an intra-turn hydrogen bond between the backbone CO(i) and the backbone NH(i+3). We observed in **Table 5** low temperature coefficient values for all the residues preceding the D-proline, which is consistent with these NHs forming hydrogen bonds and our peptides containing two β-turns in their conformations as shown in **Figure 22**. Depending on the ϕ and ψ torsion angles of their two central residues, β-turns are further subcategorized into types I, II, VIII, I', II', VIa1, VIa2 and VIb.



Type	Φ(i+1)	Ψ(i+1)	Φ(i+2)	Ψ(i+2)
I	-60	-30	-90	0
II	-60	120	80	0
VIII	-60	-30	-120	120
I'	60	30	90	0
II'	60	-120	-80	0
VIa1	-60	120	-90	0*
VIa2	-120	120	-60	0*
VIb	-135	135	-75	160*

***Figure 22**. Left) Proposed structure for the studied cyclic hexapeptides. Dashed lines represent hydrogen bonds. Right) Ideal angle values for different types of β-turns. *Types VIa1, VIa2 and VIb are subject to the additional condition that residue (i+2) must be a cis proline.*

Assuming that D-Prolines should be in the i+1 position of the β-turn, and considering that D-proline restricts Φ to 60 ± 20º, only two β-turn types (type I' or II') are possible in the studied peptides (**Figure 22**). Finally, the specific type of β-turn could be confirmed by analysis of observed NOE cross-peaks from the D-Pro-Yaa segment. A strong NOE cross-peak could be observed between the H_α of D-Pro and the NH of the following tryptophan residue (**Figure 23**). This is consistent with a D-Pro Ψ value of -120^o which is only compatible with type II'. In the case of type I', a strong NOE between the H_δ of D-Pro and the backbone NH of the following Trp should be expected.

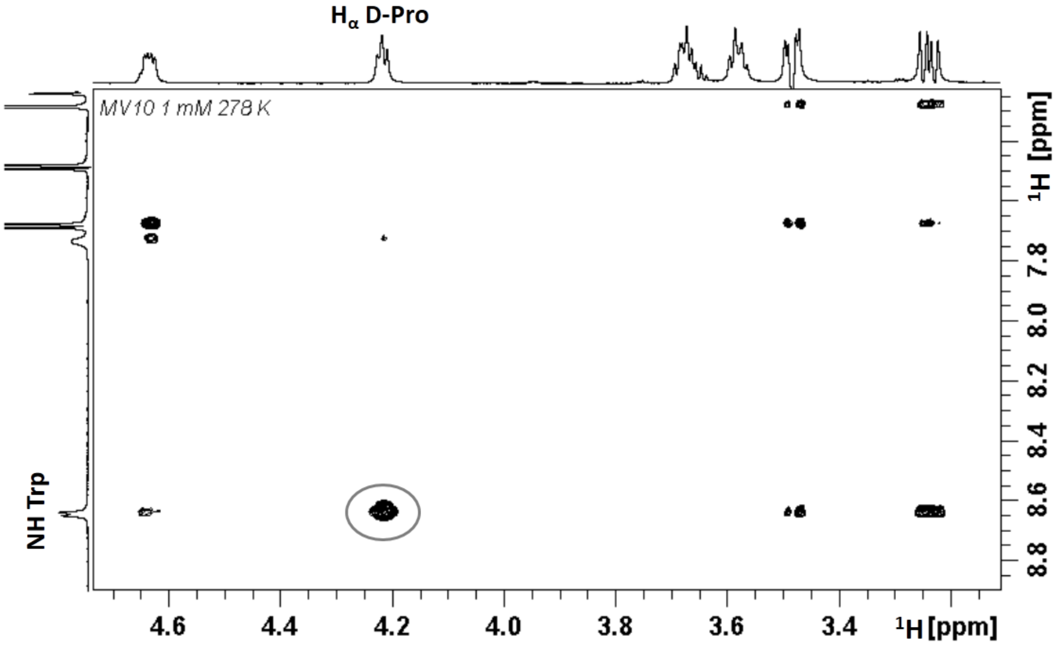


Figure 23. *NOESY spectrum at 278 K for peptide 7 in 25 mM phosphate buffer (pH 7.4), 50 mM NaCl.*

This observation, combined with all the other parameters, allowed us to confirm that those peptides were adopting a C2 symmetric conformation containing two type II' β-turns.

Evaluation of peptide 7 derivatives

The results obtained for peptide 7 in the previous section encouraged us to design and evaluate the binding capacity for VEGF of a series of related peptides. These peptides were designed exploring three different modifications of peptide 7: 1) breaking the symmetry, 2) amino acid replacement, and 3) changing residue's configuration (**Figure 24**).

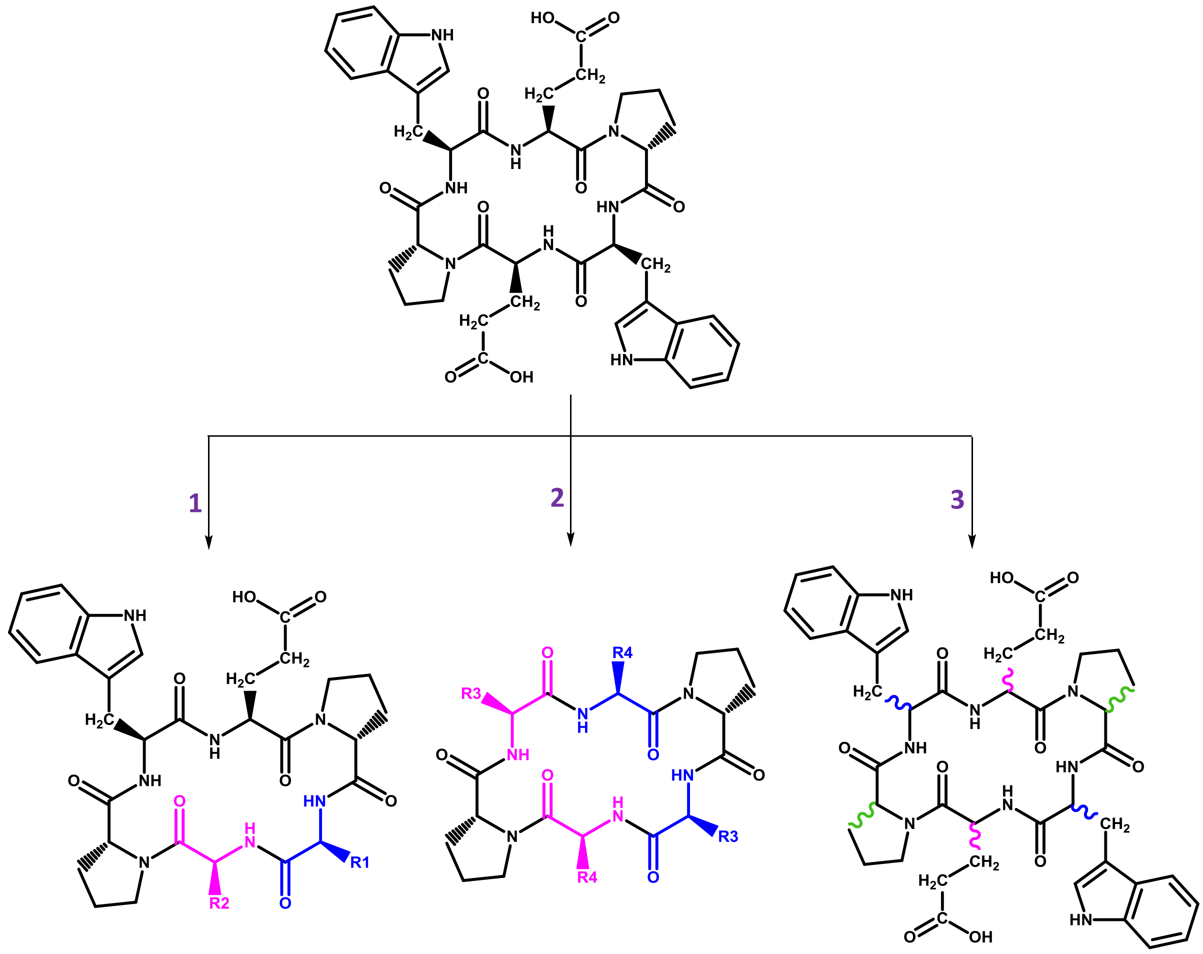


***Figure 24**. Schematic representation of the designed process for the optimization of the EXORIS peptide. R1 and R2 are the amino acids from the reduced base that were also used in the EXORIS library. R3 are natural and non-natural aromatic amino acids; and R4 are natural and non-natural negatively charged amino acids.*

When changing the configuration of the three different residues of the cyclic hexapeptides, while keeping the symmetry, we obtained seven possible diastereomers of peptide 7 (**Table 7**). I synthesized these cyclic hexapeptides by SPPS and then performed the screening of all of them by NMR as it was described in the previous section.

Table 7. *Symmetric diastereomers of peptide 7.*

Peptide code	Diastereomer	Residue configuration	Obtained peptide
7	c(EpWEpW)	LDL	c(EpWEpW)
7-C01	c(ePwePw)	DLD	c(ePwePw)
7-C02	c(EpwEpw)	LDD	c(EpwEpw)
7-C03	c(ePWePW)	DLL	c(ePWePW)
7-C04	c(epWepW)	DDL	c(epWepW)
7-C05	c(EPwEPw)	LLD	c(EPwEPw)
7-C06	c(EPWEPW)	LLL	c(EPWEPW)$_2$
7-C07	c(epwepw)	DDD	c(epwepw)$_2$

It was not possible to obtain the homochiral peptides 7-C06 and 7-C07 were not able to cyclize as cyclic hexapeptides. Instead, during the cyclization process, we observed exclusive formation of the cyclic dodecapeptides (**Table 7**).

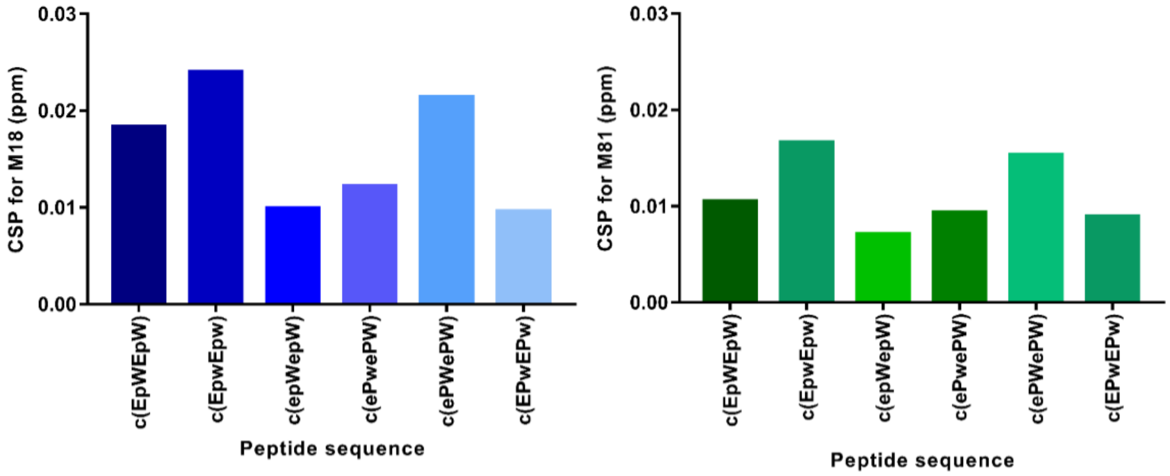


Figure 25. *Chemical shift changes induced on the methyl resonances of methionine 18 (blue) and methionine 81 (green) of VEGF by the presence of the indicated diastereomers of peptide7. c(EpWEpE) is the parent compound selected from the EXORIS library.*

The VEGF-binding ability of the peptides listed in **Table 7** was evaluated by methyl NMR experiments of methionine labeled VEGF. Modest variations on the magnitude of the peptide-induced chemical shift changes on VEGF were observed within the diastereomers listed in **Table 7**. Although the CSP seemed to be affected by residue configuration, none of the evaluated peptides showed significantly larger CSP than those obtained for peptide 7, suggesting that all of them may have similar affinities for VEGF.

Following the strategy of breaking the symmetry of cyclic hexapeptides by replacing two residues for amino acids included in the reduced base used for the synthesis of the EXORIS library, fifteen possible compounds could be obtained (**Table 8**). In the case of Trp and Glu amino acid replacement, many variations could be performed for peptide 7. Some of these cyclic hexapeptide are described in **Table 9**.

The screening of all these peptides by NMR experiments would have been very expensive and time consuming, as it required the synthesis of all the possible peptides by SPPS and great amounts of selectively labeled VEGF samples to perform the NMR assays. Taking advantage of the know-how acquired in our lab by Dr. Jesús Seco, we decided to follow a computer-aided peptide design strategy, checking the VEGF-binding capacity of these peptides by virtual docking. This in-silico methodology has the ability to predict interactions between two biomolecules, trying to find the binding site and the best binding orientation in a fast and cost effective way.[195] In typical virtual screening strategies, large libraries of drug-like compounds are computationally screened against targets of known structure, ranking them according to a scoring function, and only those molecules that have the best scores are experimentally tested.

Dr. Jesús Seco carried out in-silico docking for all the peptides described in **Table 8** and **Table 9** using the docking module implemented in MOE 2015.[196] The crystal structure of VEGF obtained in complex with VEGFR-1 (PBD: 1FLT) was used as docking template. To determine the binding mode of each cyclic peptide structure, a docking multilevel strategy was carried out. The best scored conformations after refinement protocol are detailed in **Table 8** and **Table 9**, expressed as kcal/mol. Low docking score values (negative), indicate a stable system, and thus a likely binding interaction.

Docking simulation results of all the non-symmetric studied compounds showed that none of the peptides analyzed by computational tools had significantly better values than the parent compound (peptide 7), suggesting that they should display equal or similar binding affinities (**Table 8**).

Table 8. *Docking score for the virtual screening performed for the selected non-symmetric cyclic hexapeptides.*

Peptide code	Sequence	Docking score (kcal/mol)
7	c(EpWEpW)	-8.04
7-S01	c(EpRApW)	-6.82
7-S02	c(EpEApW)	-8.42
7-S03	c(EpIApW)	-6.47
7-S04	c(EpWApW)	-7.61
7-S05	c(EpASpW)	-6.53
7-S06	c(EpRWpW)	-7.49
7-S07	c(EpIWpW)	-6.47
7-S08	c(EpSWpW)	-7.56
7-S09	c(EpRIpW)	-5.98
7-S10	c(EpEIpW)	-6.47
7-S11	c(EpSIpW)	-7.38
7-S12	c(EpERpW)	-8.07
7-S13	c(EpSRpW)	-7.59
7-S14	c(EpSEpW)	-6.61
7-S15	c(EpEWpW)	-8.44

Table 9. *Docking score for the virtual screening performed for the selected cyclic hexapeptides where amino acids replacement was done.*

Peptide code	Sequence	Docking score (kcal/mol)
7	c(EpWEpW)	-8.04
7-F01	c(DpWDpW)	-7.03
7-F02	c(XpWXpW) X =Arg(NO_2)	-7.38
7-F03	c(EpYEpY)	-5.80
7-F04	c(EpXEpX) X = 2Nal (*2-naphtyl alanine*)	-7.65
7-F05	c(EpXEpX) X = Tyr(PO_3H_2)	-7.07
7-F06	c(EpXEpX) X = Cha (*cyclohexylalanine*)	-6.88
7-F07	c(EpXEpX) X = Phe(4CF_3)	-6.21
7-F08	c(EpXEpX) X = Tyr(Me)	-6.65
7-F09	c(EpXEpX) X = Trp(5F)	-6.56
7-F10	c(EpXEpX) X = Trp(3F)	-6.71
7-F11	c(XpWXpW) X = Gla (*gamma carboxy glutamic acid*)	-6.90
7-F12	c(XpWXpW) X = Api (*pimelic acid*)	-7.65
7-F13	c(XpWXpW) X = Aad (*adipic acid*)	-7.57
7-F14	c(XpWXpW) X = Asu (*suberic acid*)	-6.00

The same happened when we did amino acid replacement (**Table 9**). Again similar docking score values were obtain indicating similar VEGF binding affinities. Nevertheless, we selected a reduced set of peptides to perform an NMR-based screening. Non-symmetric and amino acid replacement peptides were synthesized by Dr. Núria Bayó. Again we used the VEGF Met18 and Met81 NMR resonances to monitor binding as these are the two methionine residues known to be in the receptor binding interface of VEGF.

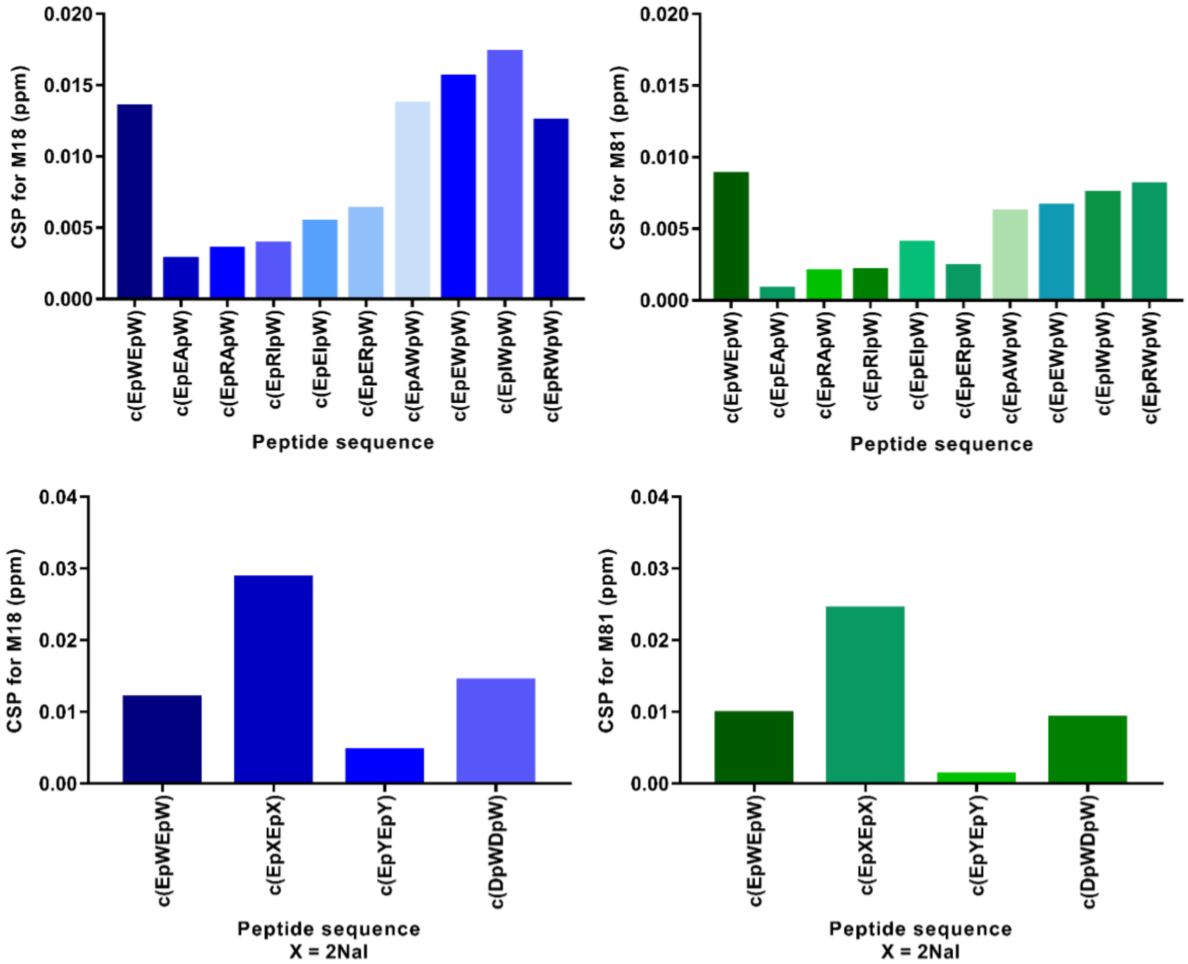


Figure 26. *Chemical shift changes induced on methyl resonances of methionine 18 (blue) and methionine 81 (green) in the presence of non-symmetric peptides (top); peptides with amino acid replacement (bottom). c(EpWEpE) is the parent compound selected from the EXORIS library.*

As shown in **Figure 26**, the chemical shift changes caused by all evaluated non-symmetric peptides in the VEGF Met18 and Met81 resonances were similar to those caused by the parent compound. These results were in agreement with those obtained by virtual screening. We noticed that, compared with peptides containing two tryptophan residues, peptides containing only one tryptophan residue induced smaller CSP both in Met18 and Met81. This can be due to the fact that both tryptophan residues may be involved in binding or because the effective concentration of tryptophan is doubled in peptides containing two aromatic amino acids.

In the case of amino acid replacement, we confirmed that tryptophan was actually important for binding VEGF. Replacement of glutamic acid by aspartic acid, which shortens the side chain in a methylene group, had no significant effects. Substitution of tryptophan by tyrosine, an aromatic and polar amino acid significantly reduced binding. On the contrary, introduction of 2NaI instead of tryptophan caused higher chemical shift changes both in Met18 and Met81. These results make sense since the receptor binding site in VEGF has a large exposed hydrophobic patch and more hydrophobic residues may enhance the interaction with this hydrophobic region. Taking this into account, we performed a titration with the compound containing 2NaI (2-naphtylalanine) to obtain an estimated a K_D value that could be compared with the one obtained for the parent compound.

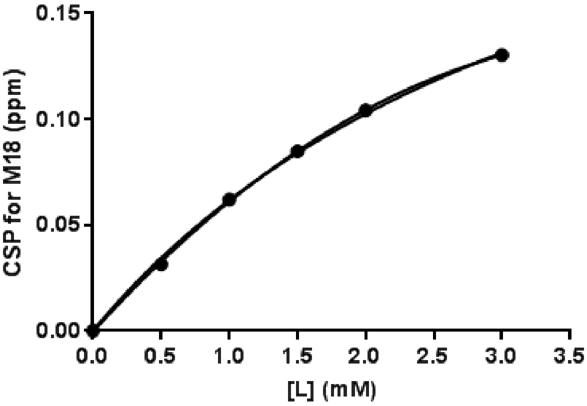


Figure 27. *Titration of 50 µM VEGF in 25 mM phosphate buffer (pH 7.4), 50 mM NaCl, at 298K; with increasing concentrations of c(EpXEpX), where X=2NaI.* K_D *= 3.9 ± 0.4 mM, Bmax = 0.30 ± 0.02 ppm,* R^2 *= 0.9988*

Again, chemical shift changes in Met18 after the addition of increasing amounts of cyclic peptide were plotted against peptide concentration and fitted using the "One site-specific binding" equation in GraphPad Prism software (**Figure 27**). The affinity of this compound for VEGF seemed to be slightly better than the one obtained for the parent compound (K_D = 3.9 ± 0.4 mM and 9.0 ± 1.4 mM; respectively), but was still in the mM range which is considered too low affinity for a reasonable good binder.

Finally, as peptides 7-C06 and 7-C07 cyclized as dodecapeptides instead of hexapeptides, we decided to study how ring size could affect VEGF binding. In this regard, we evaluated the two dodecameric peptides (7-C06D and 7-C07D), and also a nonapeptide (7-T01) that was synthesized by Dr. Núria Bayó, as possible VEGF binders (**Table 10**).

Table 10. *Peptides for the evaluation of ring expansion effect on VEGF binding.*

Peptide code	Sequence
7	c(EpWEpW)
7-C06D	c(EPWEPW)$_2$
7-C07D	c(epwepw)$_2$
7-T01	c(EpWEpWEpW)

Peptide 7-T01 caused significant chemical shift changes in the VEGF Met18 and Met81 resonances, while the other three VEGF methionine residues were not affected by the addition of 1 mM peptide. A titration assay was also performed for this compound obtaining slightly better K$_D$ than the one obtained for peptide 7. However, the affinity of this peptide for VEGF (K$_D$ = 1.1 mM) was still in the mM range (**Figure 28**).

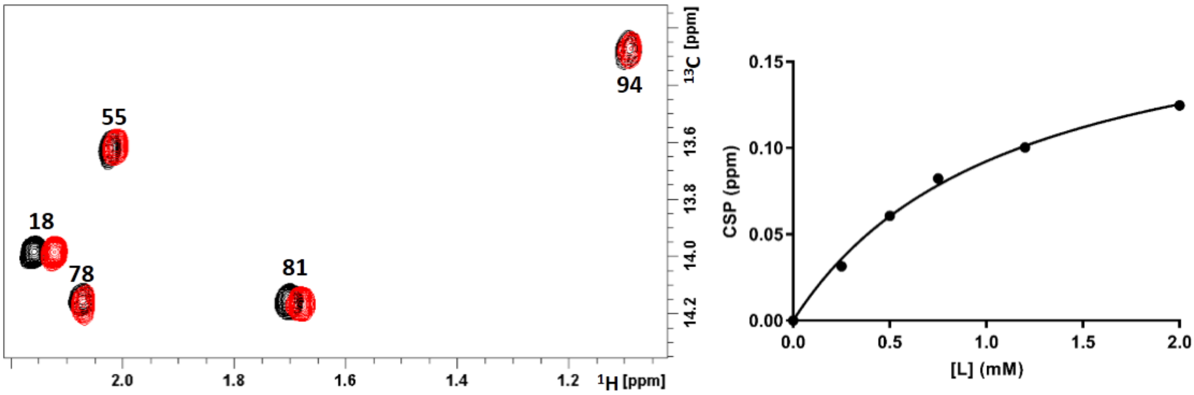


Figure 28. *Left) Overlay of the ^{1}H-^{13}C HSQC spectra of methyl-^{13}C-Met VEGF in the absence (black) and presence (grey) of 1 mM c(EpWEpWEpW) peptide. Spectra were acquired on samples containing 50 µM VEGF in 25 mM phosphate buffer (pH 7.4),50 mM NaCl. Right) Titration of 50 µM VEGF in 50 mM NaCl, 25 mM phosphate buffer pH 7.4 at 298K; with increasing concentrations of c(EpWEpWEpW). K$_D$ = 1.1 ± 0.1 mM, Bmax = 0.20 ± 0.01 ppm, R^2 = 0.9968.*

Interesting results were observed for the cyclic dodecapeptides. While peptide 7-C06D showed small chemical shift changes in Met18 and Met81, peptide 7-C07D showed significant changes in both methionine residues. Moreover, while all the previous tested peptides always showed ^{1}H chemical shift changes, and small or no ^{13}C chemical shift perturbations; peptide 7-C06D showed significant changes both in ^{1}H and ^{13}C chemical shifts. These results suggested a higher VEGF binding affinity, which was confirmed by the titration experiment performed for peptide 7-C07D. We obtained a K$_D$ value in the µM range (K$_D$ = 160 µM), one order of magnitude lower than that obtained for peptide 7 (**Figure 29**).

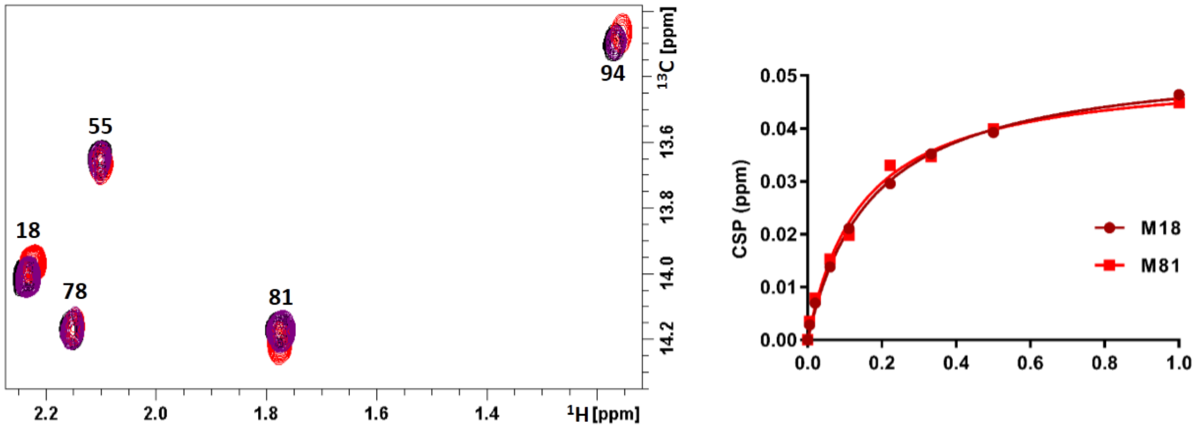


Figure 29. *Left) Overlay of the ^{1}H-^{13}C HSQC spectra of methyl-^{13}C-Met VEGF in the absence (black) and presence of 1 mM* c(EPWEPW)$_2$ *peptide (purple) or 1 mM* c(epwepw)$_2$ *peptide (red). Spectra were acquired on samples containing 50 µM VEGF in 25 mM phosphate buffer (pH 7.4), 50 mM NaCl at 298K. Right)* Titration of 50 µM VEGF in *50 mM NaCl, 25 mM phosphate buffer pH 7.4 at 298K; with increasing concentrations of c(epwepw)$_2$. K_D values were obtained fitting independently the CSP of Met18 and Met81. Met18: K_D = 0.17 ± 0.01 mM, Bmax = 0.054 ± 0.001 ppm, R^2 = 0.9978. Met81: K_D = 0.15 ± 0.02 mM, Bmax = 0.051 ± 0.002 ppm, R^2 = 0.9919.*

We were finally able to obtain an all-D cyclic peptide that binds VEGF with a µM affinity in the VEGF receptor binding interface, as it could be deduced by the large CSP in Met18 and Met81, while small or no changes are observed in the other VEGF methionine residues. To further map the peptide-binding region on the protein, we decided to perform ^{1}H-^{15}N HSQC experiments of VEGF in the presence of peptide 7-C07D.

Binding of peptide 7-C07D to the VEGF receptor binding epitope was confirmed by NMR as it is observed in **Figure 30**. Most affected residues were placed near β5 and β6 strands and in the vicinity of α1 helix and β2 strand, which are known to be important for receptors binding. With these results we could conclude that after an extensive screening of different cyclic peptides we were able to find a cyclic dodecapeptide able to bind VEGF in the receptor binding region with a µM affinity.

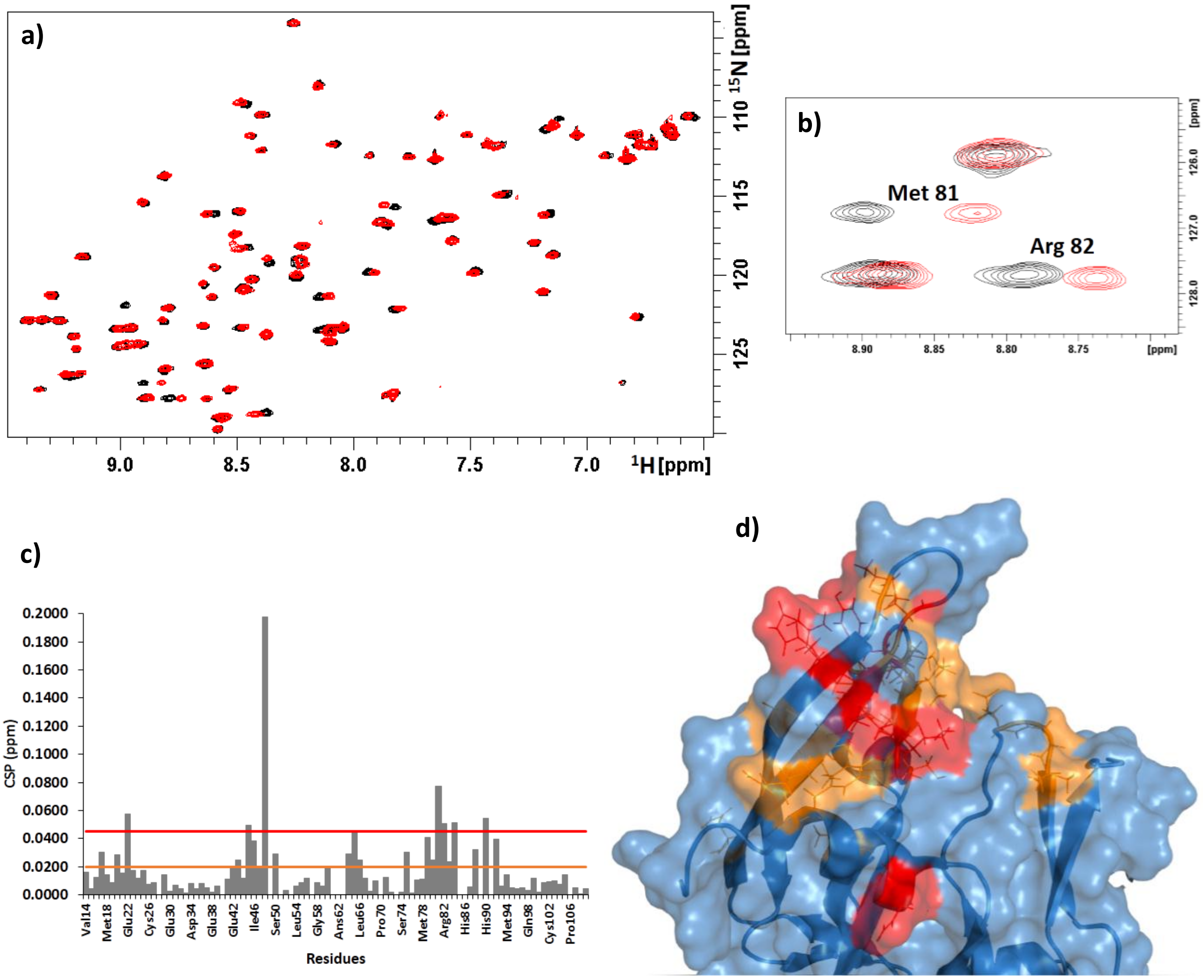


Figure 30. *NMR characterization of the VEGF region that interacts with peptide 7-C07D. a) Overlay of the* ^{1}*H-*^{15}*N HSQC spectra of* ^{15}*N-labeled VEGF in the absence (black) and in the presence of 0.15 mM of peptide 7-C07D (red). Spectra were acquired on samples containing 50 µM VEGF in 25 mM phosphate buffer (pH 7.4), 50 mM NaCl at 318K. b) Zoomed region of the* ^{1}*H-*^{15}*N HSQC spectra represented in (a) showing the chemical shift changes of two of the residues that experienced largest shifts (Met 81 and Arg82), which are located in the receptor binding epitope. c) Plots of the average combined chemical shift differences (*$\Delta\delta_{HNav}$*) induced by peptide 7-C07D as a function of VEGF sequence. The mean value (0.020 ppm) and the mean value plus standard deviation (0.045 ppm) are shown as an orange and red lines, respectively. b) Cartoon representation of the structure of the receptor binding epitope of VEGF. Most affected residues in the presence of peptide 7-C07D are depicted in red (*$\Delta\delta HN_{av}$ *> 0.045 ppm) or orange* $\Delta\delta HN_{av}$ *> 0.020 ppm).*

Conformational analysis of different diastereomers of peptide 7

As it was previously described in this chapter, we also performed a conformational analysis of different diastereomers of peptide 7 to gain insights on how the residue configuration affect the final conformation of cyclic hexapeptides (**Table 11**).

Table 11. *Peptides used to study how residue's configuration affects the conformation of cyclic hexapeptides*

Peptide code	Sequence	Configuration
7	c(EpWEpE)	LDL
7-C01	c(ePwePw)	DLD
7-C02	c(EpwEpw)	LDD
7-C03	c(ePWePW)	DLL
7-C04	c(epWepW)	DDL
7-C05	c(EPwEPw)	LLD

First, circular dichroism was performed for peptides listed in **Table 11**. As we can observe in **Figure 31**, peptides 7-C01, 7-C03 and 7-C05 behave, as expected, as enantiomers of peptides 7, 7-C02 and 7-C04, respectively. We could also notice that spectra a) and b) on **Figure 31** were pretty similar. Both present broad spectra lacking of a long-wavelength minimum and are termed type C' by comparison with Woody's classification of calculated spectra of β-turns.[197] On the contrary, the spectrum of peptide 7-C05 (**Figure 31c**) has a small negative band at ca. 225 nm and a larger positive band at ca. 200 nm (the inverse for its enantiomer 7-C04) which corresponds with type B in Woody's classification. It is known that both types B and C' are representative of type II β-turns, or type II' in the case of the enantiomer, having hydrogen bonds formed by the backbone CO of residue (i) and the backbone NH of the residue (i+3).[198] This conformation was confirmed for the previously studied EXORIS peptides, including peptide 7. The spectrum of peptide 7-C02 suggest a similar conformation than that observed for peptide 7.

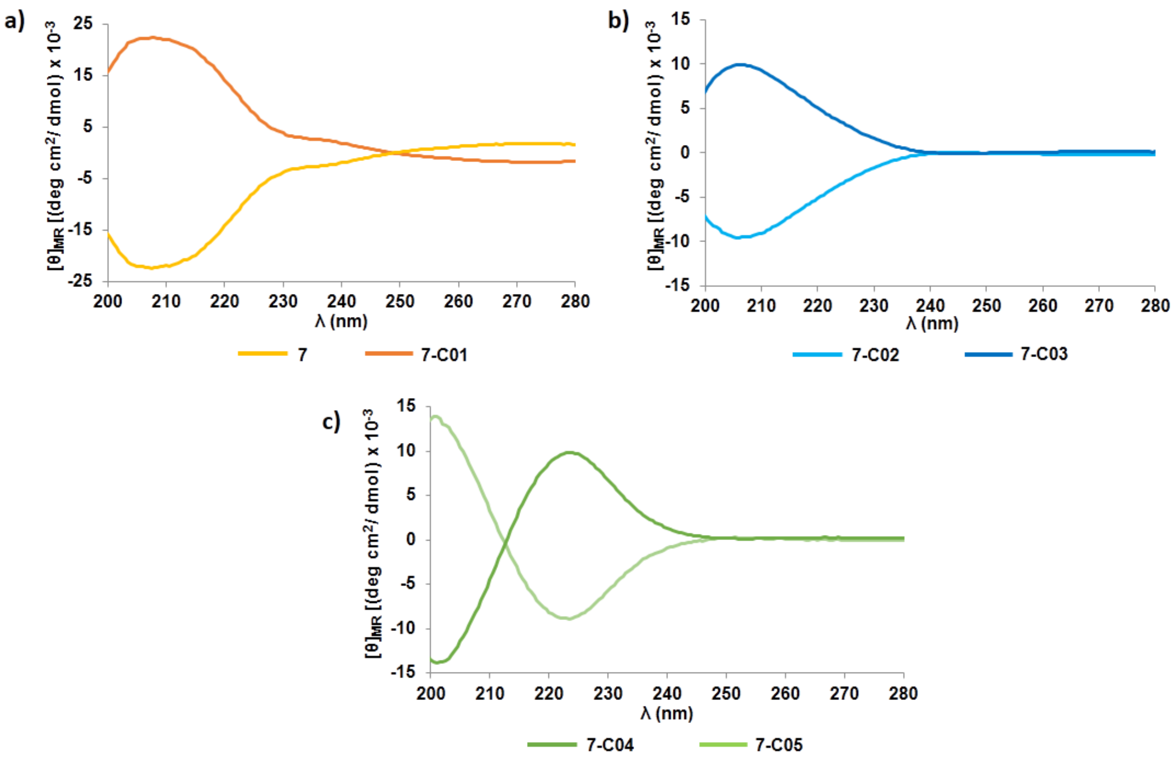


Figure 31. *Circular dichroism spectra of cyclic hexapeptides, 50 µM in 50 mM NaCl, 25 mM phosphate buffer at pH 7.4, at 298K.*

The circular dichroism spectra of these peptides suggested that we had different types of β-turns present in these compounds. Therefore, we decided to further study them by NMR. As confirmed by CD, peptides 7-C01, 7-C03 and 7-C05 were enantiomers and should have the enantiomeric conformation of peptides 7, 7-C02 and 7-C04, respectively. Therefore, we only studied peptides 7-C02 and 7-C04 to compare them with the results obtained for peptide 7. The remaining three peptides should yield identical results. 1D ^{1}H NMR as well as a series of 2D standard NMR experiments (TOCSY, COSY, NOESY and ^{1}H-^{13}C HSQC) were acquired for those peptides.

The first interesting observation, detected by visual inspection of the 1D ^{1}H-NMR spectra, was that for peptides 7 and 7-C02 we only observed one set of signals corresponding to three individual spin systems, suggesting either the presence of one conformation or a conformational ensemble in fast exchange in the NMR timescale. On the contrary, for peptide 7-C04 we could observe more than two sets of signals with different peak intensities, suggesting the presence of more than one species that are in slow exchange in the NMR timescale (**Figure 32**).

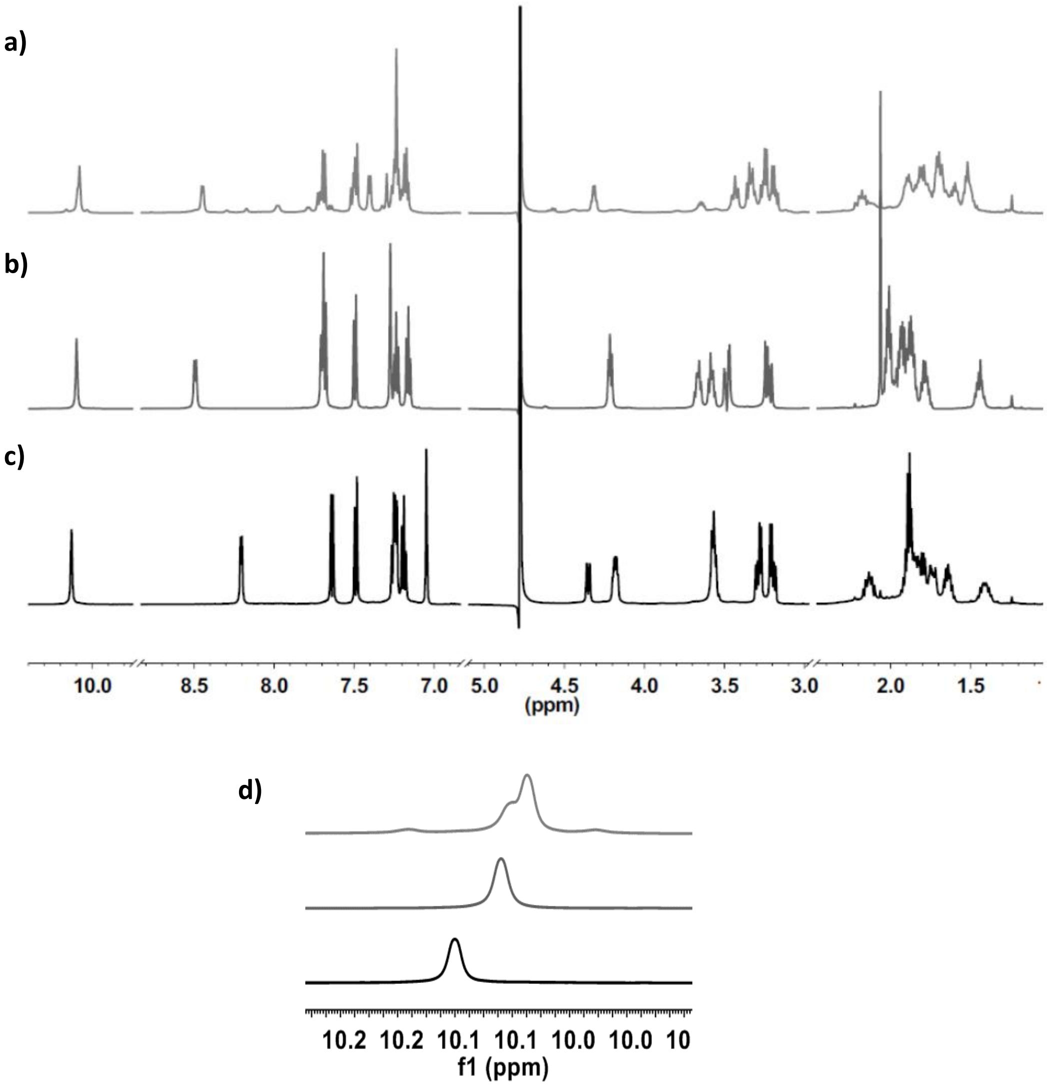


Figure 32. *1D ^{1}H NMR spectrum for: a) peptide 7-C04, b) peptide 7 and c) peptide 7-C02. d) Zoom for the ^{1}H NMR spectra signal corresponding to the NH indole where it can be clearly seen that only one signal is present for peptides 7 and 7-C02 while more than two signals are observed for peptide 7-C04.*

We hypothesized that the different species observed for peptide 7-C04 could be due to the *cis* and *trans* isomerization of the Xaa-Pro bonds. To evaluate the conformation of the Xaa-Pro bond, we estimated the differences in the ^{13}C chemical shifts of the C_β and C_γ of the Pro residues of the observed species, as it was previously described in this chapter.

Table 12. *Chemical shift difference between $^{13}C_{\beta}$ and $^{13}C_{\gamma}$ of the proline residue for the indicated peptides.*

Peptide code	δC_{β} (ppm)	δC_{γ} (ppm)	$\Delta\delta C_{\beta\gamma}$ (ppm)	Xaa-Pro bond conformation
7	30.7	27.1	3.7	*trans/trans*
7-C02	32.0	26.2	5.8	*trans/trans*
7-C04 (major)	34.4	24.2	10.3	*cis/cis*
7-C04 (minor)	30.8	27.7	3.4	*trans/trans*

Comparison of the results obtained for these peptides (**Table 12**) revealed a distinct trend of the $\Delta\delta C_{\beta\gamma}$. The higher $\Delta\delta C_{\beta\gamma}$ value obtained for peptide 7-C02 ($\Delta\delta C_{\beta\gamma}$ = 5.8 ppm), when compared with that of peptide 7 ($\Delta\delta C_{\beta\gamma}$ = 3.7 ppm), may be due to the coexistence of several species in fast exchange in the NMR time scale.

Peptide 7-C04 presented two different species that correlated with all-*trans* and all-*cis* Xaa-Pro bonds, being the all-*cis* the most abundant one. Other minor conformations were also present in the 1D spectrum of this peptide but could not be unambiguously assigned due to severe signal overlap.

We also analyzed the dependence of the amide NH protons chemical shifts (**Table 13**), $^{3}J_{HNH\alpha}$ coupling constants (**Table 14**), and H_{α} and C_{α} chemical shifts (**Figure 33**) for these peptides.

Table 13. *Temperature coefficient values for the indicated peptides (c(Xaa-D-Pro-Yaa)$_2$).*

Peptide code	Xaa-Pro bond conformation	Sequence	$-\Delta\delta_{NH}/\Delta T$ (ppb/K)	
			Xaa	Yaa
7	*trans/trans*	c(EpWEpE)	1.5	7.7
7-C02	*trans/trans*	c(EpwEpw)	8.7	2.7
7-C04 (major)	*cis/cis*	c(epWepW)	5.0	8.7
7-C04 (minor)	*trans/trans*	c(epWepW)	8.0	5.7

Based on the temperature coefficients (**Table 13**), only peptide 7-C02 seemed to present a hydrogen bond between the backbone (CO) of (i) residue and the NH of the (i+3) residue of the turn. Nonetheless, these NHs seemed to be more solvent exposed that the ones present in peptide 7, as they showed higher temperature coefficient values. Taking these into account it seemed that, while peptide 7 showed a conformation with the D-proline residues in the i+1 position of the type II' β-turn; for peptide 7-C02, D-proline residues seemed to be in position i+2 of the turn. For 7-C02, the β-turn subtype could not be unequivocally determined because residue i+2 in this peptide is a proline without NH amide proton to detect NOE correlations. As proline had to be in the i+2 position of the turn and, taking into account that D-proline restricts Φ to 60 ± 20°, we only had one possibility which corresponds to type II β-turn. Furthermore, it was previously described by Lila Gierasch[198] that peptides with the sequence c(Gly-Pro-Val)$_2$ and c(D-Phe-Pro-Val)$_2$ tend to form type II' β-turns in solution with proline residues in positions i+2 of the turn.

Table 14. *$^3J_{HN\text{-}H\alpha}$ coupling constants for the indicated peptides c(Xaa-D-Pro-Yaa).*

Peptide code	**Xaa-Pro bond conformation**	**Sequence**	**$^3J_{\alpha NH}$ (Hz)**	
			Xaa	**Yaa**
7	*trans/trans*	c(EpWEpE)	8.1	7.8
7-C02	*trans/trans*	c(EpwEpw)	5.3	7.7
7-C04 (major)	*cis/cis*	c(epWepW)	7.4	6.5
7-C04 (minor)	*trans/trans*	c(epWepW)	8.6	6.4

There is a well-known correlation between spin-spin coupling constants and dihedral angles. However, the $^3J_{\alpha NH}$ measured for these peptides showed intermediate values that could not be used for an unambiguous estimation of dihedral angles (**Table 14**).

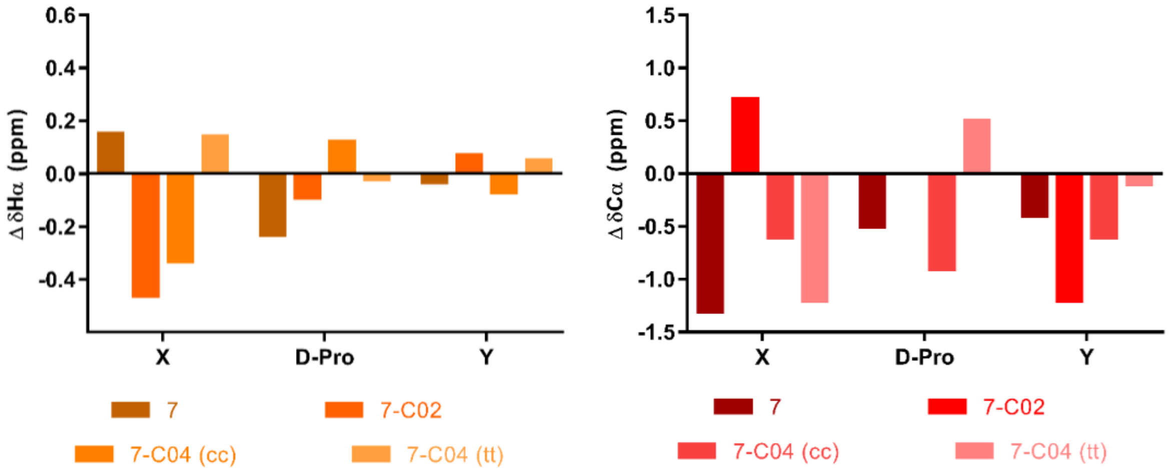


Figure 33. *Plots of Hα (orange) and Cα (red) secondary chemical shifts ($\delta_{observed} - \delta_{random}$) comparing the indicated peptides. Random coil values were extracted from Wishart et al.*[193]

Despite secondary chemical shifts showed substantial differences between peptides, no significant information regarding the conformation of these peptides could be obtained, since a clear trend was not observed for any of them.

A feature to be emphasized when comparing these three peptides is the differential behavior of the methylene H_β protons of tryptophan residues in peptide 7. In this peptide, the tryptophan H_β protons displayed a large diastereotopic splitting ($\delta_{H\beta}$ = 3.48 and 3.22 ppm) and a remarkable difference in $^3J_{H\alpha H\beta}$ coupling constants (4.3 and 9.5 Hz) (**Figure 34**). In contrast, there are not great differences in chemical shift and $^3J_{H\alpha H\beta}$ values between diastereotopic H_β protons in the tryptophan residues of peptide 7-C02 ($\delta_{H\beta1}$ = 3.29 ppm, $^3J_{H\alpha H\beta1}$ = 7.2 Hz; $\delta H\beta2$ = 3.21 ppm, $^3J_{H\alpha H\beta2}$ = 7.0 Hz) and the major conformer of 7-C04 ($\delta_{H\beta1}$ = 3.24 ppm, $^3J_{H\alpha H\beta1}$ =8.3 Hz; $\delta_{H\beta2}$ = 3.19 ppm, $^3J_{H\alpha H\beta1}$ = 7.8 Hz). The pronounced differences observed between diasterotopic H_β protons in the tryptophan residues of peptide 7 suggest restricted rotation of the aromatic side chain in this peptide in comparison with that of peptides 7-C02 and 7-C04. Most probably, rotation of the tryptophan side chain in peptide 7 is restricted by the presence of π-CH interactions between the aromatic ring and the proline methylene groups. If present, the π-CH interaction is expected to cause ring current induced shifts in the proline resonances. This is in agreement with the upfield shift of the H_β resonance of the proline residue in peptide 7 ($\delta_{H\beta}$ = 1.93, 1.44 ppm), compared with peptides 7-C02 ($\delta_{H\beta}$ = 2.14, 1.75 ppm) and 7-C04 (2.18, 1.70 and 2.10, 1.71 ppm, for the cis and trans, respectively). These π-CH interactions may enhance the stability of a reduced set of peptide conformations. On the contrary, the NMR-derived parameters of peptides 7-C02 and 7-C04 reflected a larger conformational ensamble (**Table 13** and **Table 14**)

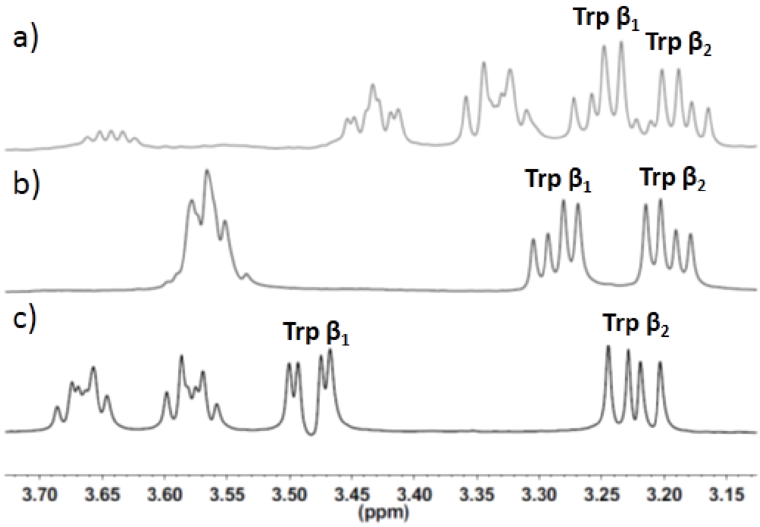


Figure 34. *^{1}H-NMR spectra regions showing signals corresponding to Trp H_β resonances for peptides: 7-C04 (a), 7-C02 (b) and 7 (c).*

Having analyzed the above mentioned NMR parameters, we could hypothesize that peptide 7 had a stable symmetric conformation containing two type II' β-turns with proline in position i+1 of these turns. Peptide 7-C02 seemed to present a symmetric conformation containing two type II β-turns with D-proline in position i+2 of the turn. Even so, the results for this peptide appeared to be less definitive than those obtained for peptide 7.

NMR analysis of peptide 7-C04 was difficult since it presented different conformers in slow exchange in the NMR timescale, which prevented a detailed analysis of the NMR parameters for the different species observed. However, in a recent work,[199] the conformational analysis of peptides with the sequence c(X-Pro-Y-Lys-Pro-Y), where X = Tyr, Ser or Ser(PO_3H_2) and Y = D-Ala or D-Phe, presenting different Xaa-Pro *cis/cis* amide bond contents was described. The authors of the study claimed that peptide c(Ser-Pro-D-Phe-Lys-Pro-D-Phe) adopts a conformation with two pseudo-βVIb turns. This βVIb turn is characterized by *cis* proline in position i+2 and no hydrogen bonds; which perfectly correlates with the NMR data obtained for peptide 7-C04.

As peptide 07-C04 has the enantiomeric sequence configuration of that described by Malešević et al,[199] we hypothesized that peptide 7-C04 may had a conformation containing two inverse pseudo-βVIb turns with D-proline in the i+2 position.

Nowadays, rational design of a βVI-turn conformation still represents a considerable challenge as it is characterized by a *cis* imide bond preceding a proline residue in the i+2 position. The βVIa subtype is additionally stabilized by an intramolecular hydrogen bond between the carbonyl oxygen of residue i and the amide group of residue i+3, whereas this hydrogen bond is absent in the βVIb-turn. The Xaa-Pro segment of these βVI-turn has been proposed as a potential binding motif for different peptidyl prolyl *cis/trans* isomerases (PPIases); which are emerging as important targets related with different diseases such as cancer, Alzheimer, asthma, etc.[200-202]

Knowing that peptide 7-C04 can adopt, to some extent, two pseudo-βVIb turns; we decided to perform different modifications of this peptide with the aim of designing a cyclic hexapeptide with the highest possible population of the *cis* Xaa-proline bond that adopts a βVIb turn conformation. In this regard, in peptide 7-C04 (c(epW)$_2$) we replaced residues X and Y (c(X-D-Pro-Y)$_2$) by alanine while retaining the configuration of the parent compound to evaluate how the *cis/cis* content of the peptide was affected. We also replaced the Trp residue by Tyr to compare the effect on peptide conformation of different aromatic residues in this position.

Table 15. *Peptide analogs of compound 7-C04. *Peptide 7-C04-A1 could not be evaluated due to solubility restrictions. Quantification was done by peak integration in 1D ^{1}H-NMR spectra. More precise quantification could not be perform due to signal overlap.*

Peptide code	Sequence	cis/cis content	Other conformers observed
7-C04	c(epWepW)	~85%	trans/trans
7-C04-A1	c(apWapW)	*	*
7-C04-A2	c(epAepA)	~60%	trans/trans
7-C04-A3	c(apAapA)	< 5%	trans/trans
7-C04-A4	c(epYepY)	~90%	trans/trans

Table 15 shows the percentage of conformers observed in the 1D ^{1}H NMR spectra of these peptides at 298 K. None of the evaluated peptides present higher *cis/cis* content than peptide 7-C04. Removing the aromatic side chain of the residue following the D-Pro reduced the *cis/cis* content from 85% in peptide 7-C04 to 60% in peptide 7-C04-A2. Interestingly, when both X and Y residues were replaced by alanine (peptide 7-C04-A3) the *cis/cis* content was drastically reduced to <5%. These observations confirm that the nature of the amino acid preceding the D-Pro has a higher impact on the conformation of the Xaa-Pro bond. When changing Trp by Tyr, no significant changes were observed on the conformation of this compound.

It is known that aromatic residues preceding the D-proline residues favor the *cis* conformation of the imide bonds,[203] so we decided to switch the positions of residues X and Y in peptide 7-C04, while keeping the same configuration. Remarkably, the *cis/cis* content increased from 85% in 7-C04 to >95% in the resulting peptide (7-C04-P1) (**Table 16**).

After these promising results, we decided to further study this peptide by substituting the X and Y residues for other proteinogenic amino acids. We replaced the aromatic residue (D-Trp) by a bulky hydrophobic but not aromatic residue (D-isoleucine) and by a small hydrophobic residue (D-Alanine). Finally, we replaced the glutamic residue (negatively charged) by glutamine (polar residue), arginine (positively charged residue) and alanine (hydrophobic residue).

Table 16. *Peptide 7-04-P1 and analogs. *Peptide 7-C04-P1-A4 could not be evaluated due to solubility restrictions. ** Minor conformation could not be unambiguously assigned due to very low signal intensity. Quantification was done by peak integration in 1D* ^{1}*H-NMR spectra. More precise quantification could not be perform due to signal overlap.*

Peptide code	Sequence	cis/cis content	Other conformers observed
7-C04-P1	c(wpEwpE)	> 95%	**
7-C04-P1-A1	c(ipEipE)	~80%	cis/trans
7-C04-P1-A2	c(apEapE)	~10%	trans/trans
7-C04-P1-A3	c(wpQwpQ)	> 95%	**
7-C04-P1-A4	c(wpRwpR)	> 95%	**
7-C04-P1-A5	c(wpAwpA)	*	*

As it can be observed in **Table 16**, replacement of the aromatic residue by isoleucine reduced to some extent (from >95% to 80%) the *cis/cis* content in this cyclic hexapeptide. Nevertheless, as isoleucine has a bulky side chain, we still observed a high content of the *cis/cis* conformer. Furthermore, in this peptide it can be noticed that the second conformer observed still contains one Xaa-Pro *cis* imide bond, while the second conformation observed for all the other peptides had both Xaa-Pro imide bonds in the *trans* conformation. Replacement of the aromatic residue by alanine had a drastic effect. In peptide 7-C04-P1-A2 we observed only a 10% of *cis/cis* content, being the all-*trans* the major conformation present in this peptide. We could conclude that aromaticity of the residue preceding the D-proline residue seemed to significantly increase the *cis/cis* content; however, steric effects are also important to determine the conformation of these imide bonds.

Peptides 1-C04-P1-A3 and 1-C04-P1-A4 displayed very similar ^{1}H-NMR spectra when compared to that of peptide 1-C04-P1, suggesting that the position following the D-Pro residue has modest or no effects in the conformation of the imide bonds when there is an aromatic residue preceding the D-Pro. A complete characterization of these peptides is included in the experimental section.

Comparison between peptide 7-C04 and 7-C04-P1 revealed significant differences between the two all-*cis* conformations (**Table 17** and **Figure 35**).

Table 17. *Chemical shift difference between $^{13}C_{\beta}$ and $^{13}C_{\gamma}$ of the proline residue; temperature coefficient values and $^{3}J_{HN\text{-}H\alpha}$ coupling constants of Xaa and Yaa in the indicated c(Xaa-D-Pro-Yaa)$_2$ peptides.*

Peptide code	Sequence	$-\Delta\delta_{NH}/\Delta T$ (ppb/K)		$^{3}J_{\alpha NH}$ (Hz)		$\Delta\delta C_{\beta\gamma}$ (ppm)
		Xaa	Yaa	Xaa	Yaa	
7-C04 (all-cis)	c(epWepE)	5.0	8.7	7.4	6.5	10.3
7-C04-P1	c(wpEwpE)	7.0	7.5	6.6	6.8	9.6

Similar $^{3}J_{HN\text{-}H\alpha}$ coupling constants and amide temperature coefficient values were observed for these two peptides. Nonetheless, an important upfield shift of the H_{α} of the proline residue was noticed in peptide 7-C04-P1.

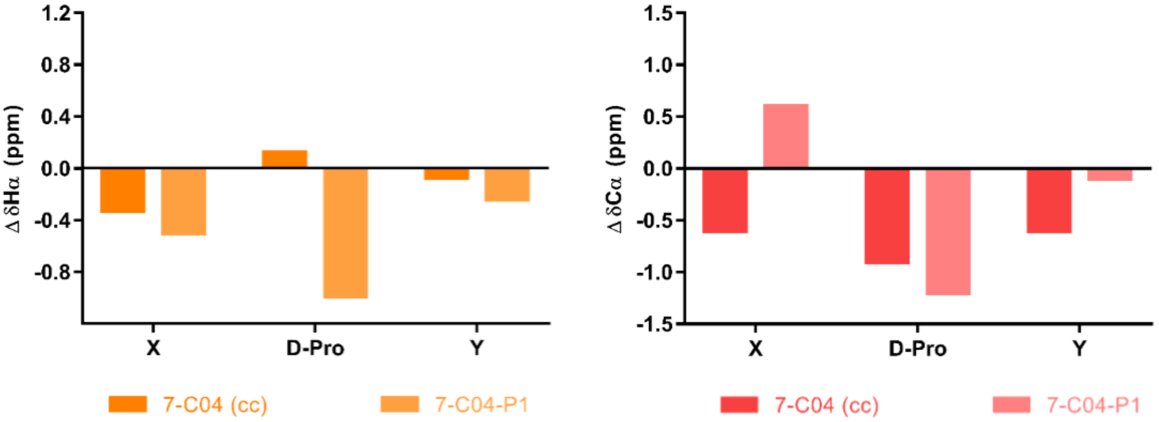


Figure 35. *Plots of Hα (orange) and Cα (red) secondary chemical shifts ($\delta_{observed} - \delta_{random}$) comparing the indicated peptides. Random coil values were extracted from Wishart et al.*[193]

Similarly to what was observed for peptide 7, the tryptophan H_{β} for peptide 7-C04-P1 had widely different chemical shifts (3.39 and 3.03 ppm) and $^{3}J_{H\alpha H\beta}$ coupling constants (4.4 and 10.9 Hz), as opposed to that of the *cis* conformer of peptide 7-C04 ($\delta_{H\beta}$ = 3.24, 3.19 ppm and $^{3}J_{\alpha\beta}$ = 8.3 and 7.8 Hz). As it was previously described for peptide 7, this indicates restricted bond rotation in peptide7-C04-P1, suggesting the presence of π-CH interactions (**Figure 36**). Consistent with this observation, the current effects of the spatially proximal aromatic ring may cause the significant upfield shift (0.98 ppm) of the proline H_{α} proton in peptide7-C04-P1. These π-CH interactions may be responsible for the enhanced stability, compared to peptides 7-C04 and 7-C04-P1-A1, of the *cis* imide bond in peptide 7-C04-P1. These π-CH interactions were also observed in peptides 7-C04-P1-A3 and 7-C04-P1-A4.

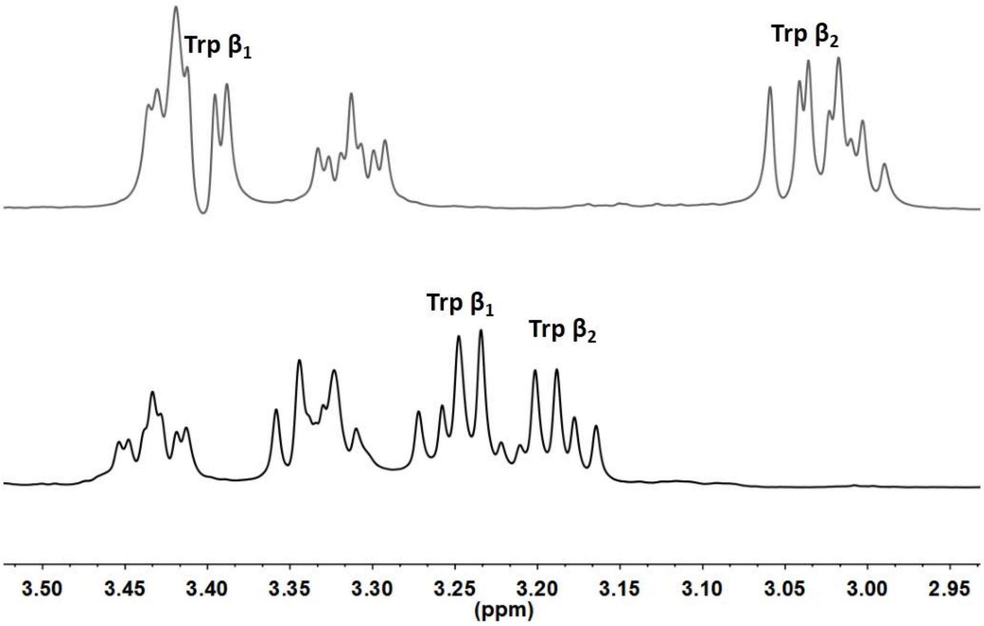


Figure 36. *^{1}H-NMR spectra regions showing signals corresponding to Trp H_β resonances for: top) 7-C04-P1, bottom) 7-C04.*

We observed that cyclic hexapeptides could adopt different β-turn conformations depending on the configuration of their residues. In order to gain more insight into different conformational ensembles sampled by these peptides, we decided to perform molecular dynamics in collaboration with Dr. Jesús Seco. Peptides 7, 7-C02, 7-C04 and 7-C04-P1 were selected for that purpose.

Molecular simulation began with a structure of the peptide built with the XLEaP program from the AMBER14 molecular mechanics package.[204] The cyclic peptides were constructed by linking the N- and C-terminal residues of the linear peptide followed by energy minimization. Corresponding chiralities were assigned to different residues. For all the systems, the Amber ff99SB force field[205] was used, together with the re-optimized omega-bond angle parameters.[206] The simulated snapshots were clustered using a heuristic method in which a snapshot was included into its closest cluster if the main-chain root mean square distance RMSD is smaller than 2.5 Å after rigid body alignment. Representative structures were further compared and the clusters were merged together if the representative structures were closer than 2.5 Å.

The three more commonly populated states for peptides 7, 7-C02, 7-C04 and 7-C04-P1 are shown in **Figure 37**, **Figure 39**, **Figure 41** and **Figure 44**.

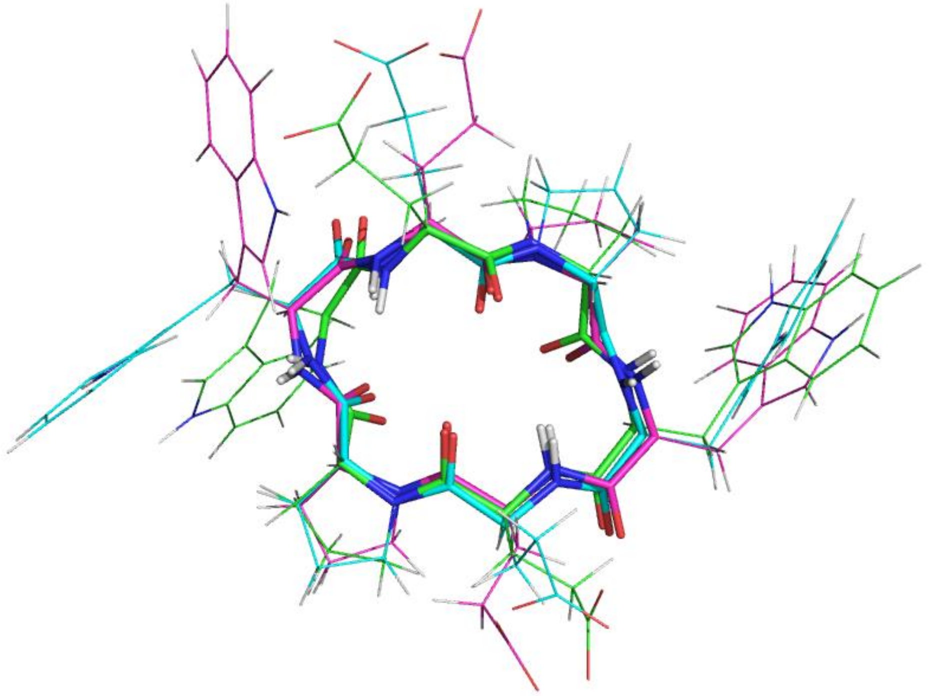
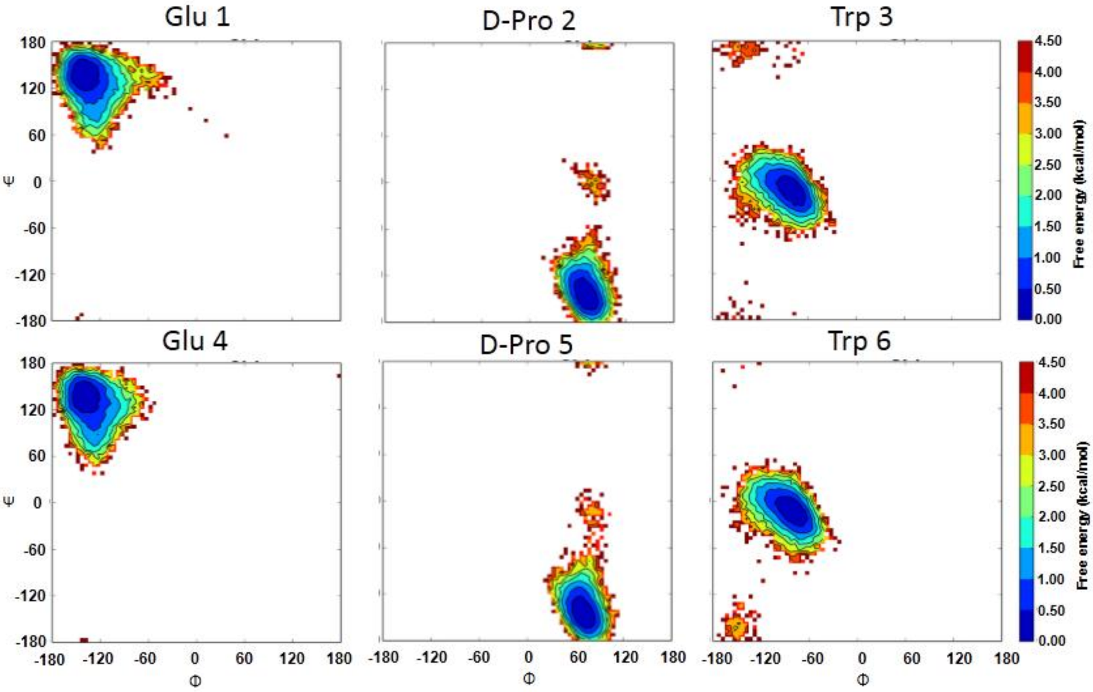


Figure 37. *Three most commonly populated states obtained from REMD for peptide 7 (RMSD = 1.15 ± 0.37).*

As it can be observed in **Figure 37**, the three most populated states for peptide 7, which account for ~45% of the total population, are in perfect agreement with the NMR results; suggesting a conformation with two β-turns, with proline residues occupying i+1 positions of these turns, and hydrogen bonds between backbone NH and CO of both glutamic residues.



Glu 1

D-Pro 2

Trp 3

Glu 4

D-Pro 5

Trp 6

Figure 38*. Ramachandran plots obtained from REMD for each residue of peptide 7.*

The Ramachandran plots obtained from REMD for peptide 7 (**Figure 38**), perfectly correlates with those of a βII’ turn (Φ(i+1)= 60, Ψ(i+1)= -120; Φ(i+2)= -80, Ψ(i+1)= 0). This results, together with those obtained by NMR confirm that peptide 7 in solutions adopts a very stable symmetric conformation containing two βII’ turns.

In contrast, the most abundant conformations obtained for peptide 7-C02 by REMD, which account for ~70%, did not completely correlate with the results obtained by NMR.

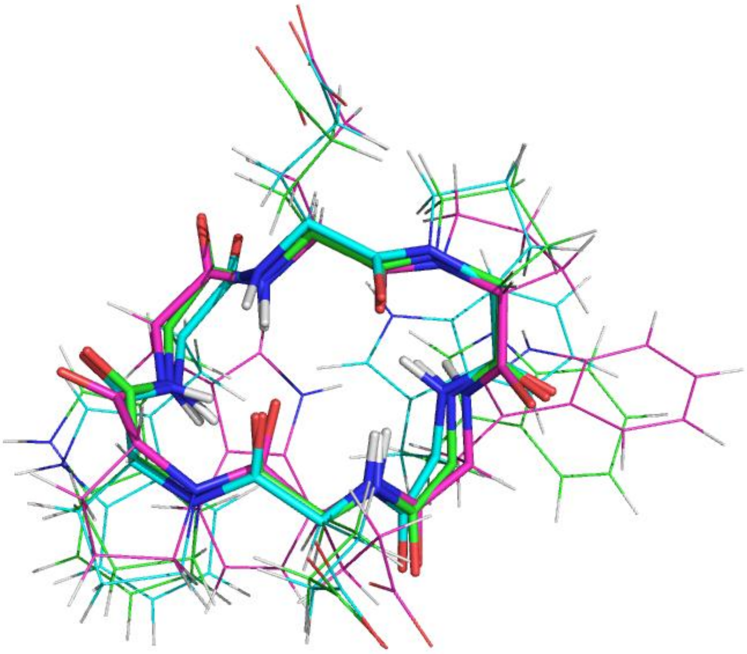


Figure 39. *Three most commonly populated states obtained from REMD for peptide 7-C02 (RMSD = 1.71 ± 0.74).*

As it can be noticed in **Figure 39**, this peptide tends to adopt a conformation similar to the one observed for peptide 7 with two β-turns with the proline residues in the i+1 position and with two hydrogen bonds between backbone NH and CO of glutamic residues. Even so, NMR results suggested the presence of two β-turns with the proline residues in the i+2 position and two hydrogen bonds between backbone NH and CO groups of tryptophan residues.

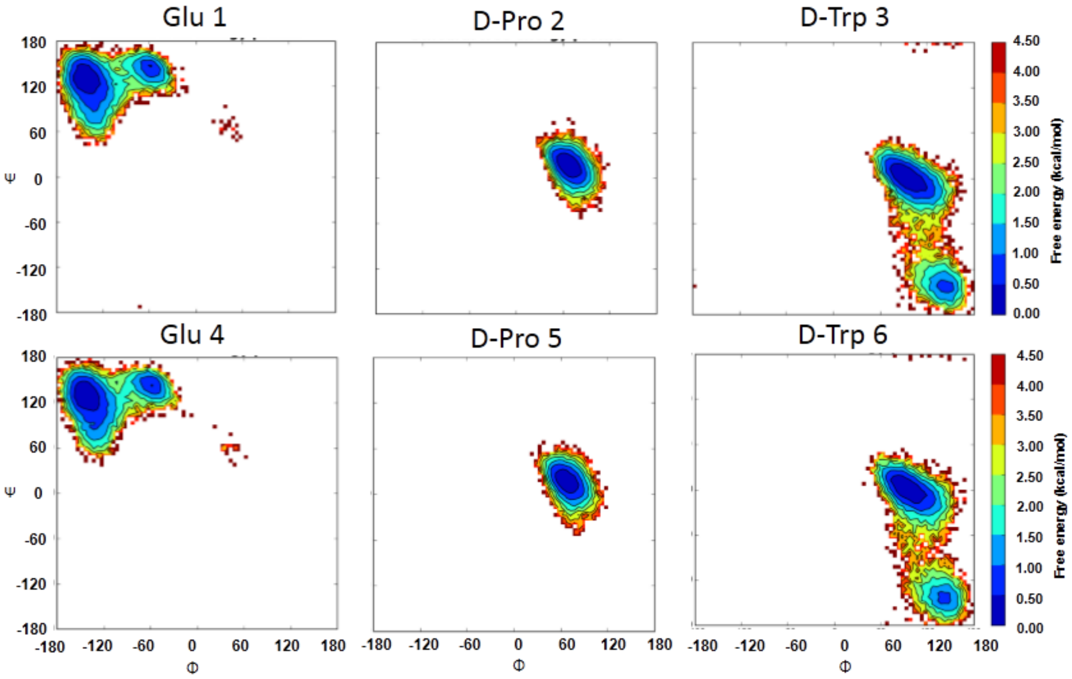


Figure 40. *Ramachandran plots obtained from REMD for each residue of peptide 7-C02.*

The dissimilarities between the results obtained by NMR and by REMD can be explained after a close observation of the Ramachandran plots obtained for each residue (**Figure 40**). As it can be noticed, dihedral angles for proline residues are around Φ= 60±20 and Ψ= 0±30, suggesting either type βI' (Φ(i+1)= 60, Ψ(i+1)=30) or type βII (Φ(i+2)= 80, Ψ(i+2)=0). If βI' turns are present in this peptide's conformation, i+2 position may be occupied by tryptophan residues and may present typical βI' dihedral angles (Φ(i+2)= 90, Ψ(i+2)=0) which perfectly correlates with the most abundant dihedral angle values for tryptophan observed in the Ramachandran plot. In this conformation, glutamic residues may be in an extended conformation. In the case of βII turns glutamic residues may occupy i+1 positions presenting typical βII dihedral angles (Φ(i+2)= -60, Ψ(i+2)=120) which are in agreement with one of the regions sampled by glutamic in the Ramachandran plot. In this conformation tryptophan residues are supposed to adopt an extended conformation. This βII conformation is in agreement with the NMR results obtained for this peptide, but is not the most abundant conformation by REMD.

These observations may suggest the presence of two different conformational ensembles containing two βII turns or two βI' turns that are in equilibrium exchange. This type I' ↔ II inter-conversions in cyclic hexapeptides was recently reported by McHugh et al.[207] In this work they suggested a pathway through which a conformation with a type II turn at residues 1 and

2 + a type II turn at residues 4 and 5 changes into a conformation with a type I′ at residues 2 and 3 + a type I′ turn at residues 5 and 6.

REMD can explain the fact that the results obtained by NMR were not as unambiguous as the ones obtained for peptide 7. As we hypothesized this can be explained by the presence of the two different conformations that are in fast equilibrium exchange in the NMR time scale.

As when analyzing peptide 7-C04 by NMR we observed two different conformers, corresponding to all-*trans* and all-*cis* imide bonds, we sampled both conformations by REMD (**Figure 41**). For the all-*trans* conformer, the presence of two β-turns with proline residues in position i+2 can be noticed. In this case, tryptophan residues are present in an extended conformation forming two backbone hydrogen bonds, while glutamic residues occupy the i+1 positions of this β-turns. For the all-*cis* conformer no hydrogen bonds were observed in the structures observed for the most commonly populated states, which may suggest the presence of pseudo βVIb turns in this peptide.

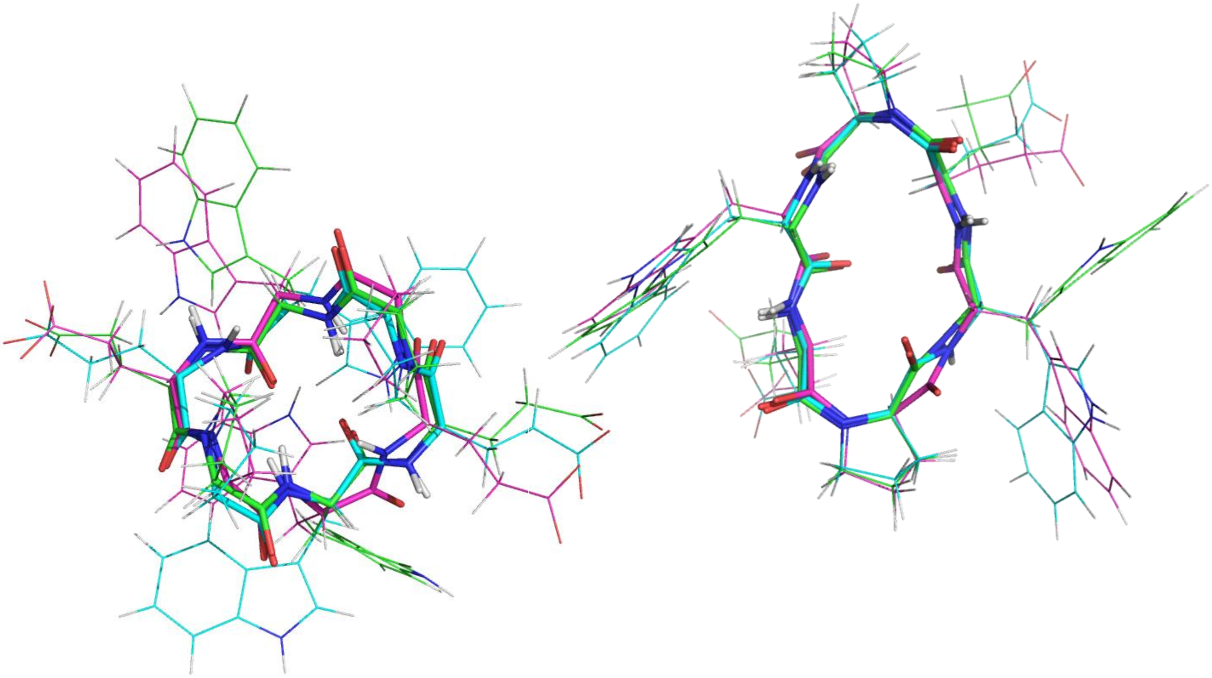


Figure 41. *Three most commonly populated states obtained from REMD for peptide 7-C04: left) all trans (RMSD = 2.09 ± 0.94), right) all cis (RMSD = 1.37 ± 0.56).*

A close observation of the Ramachandran plots obtained by REMD may provide more insights into the different conformations adopted by peptide 7-C04 (**Figure 42** and **Figure 43**).

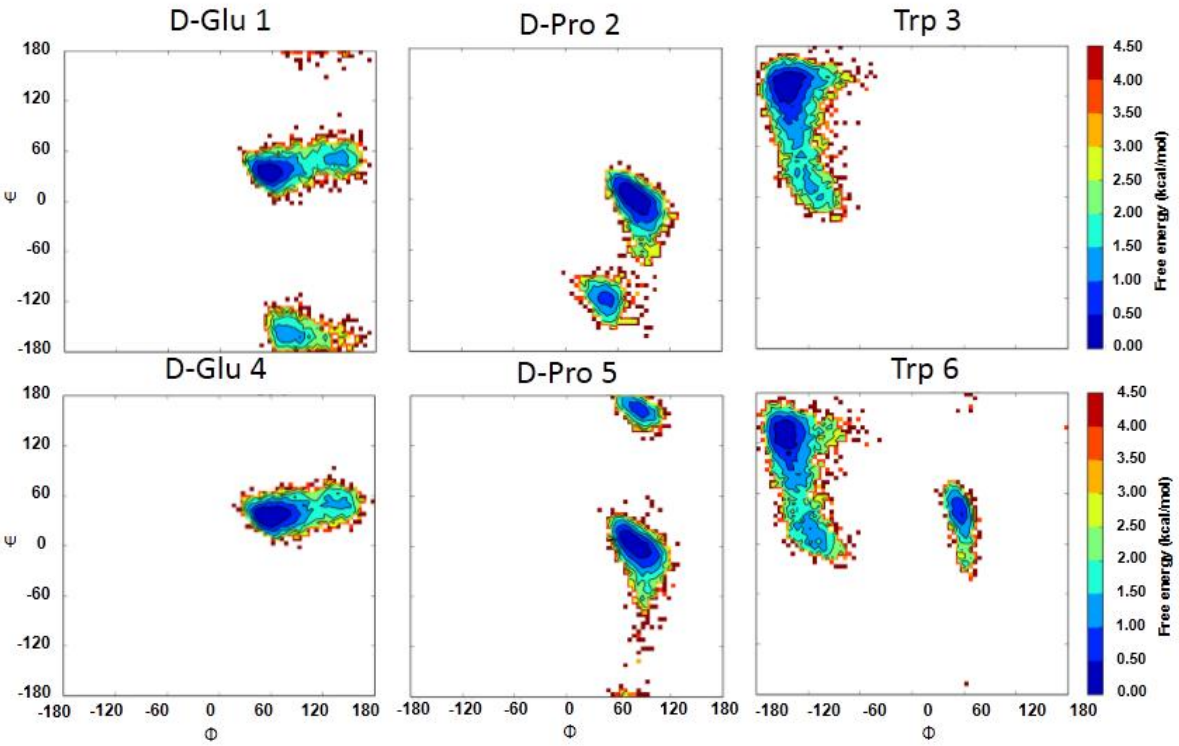


Figure 42. *Ramachandran plots obtained from REMD for each residue of peptide 7-C04 all trans conformer.*

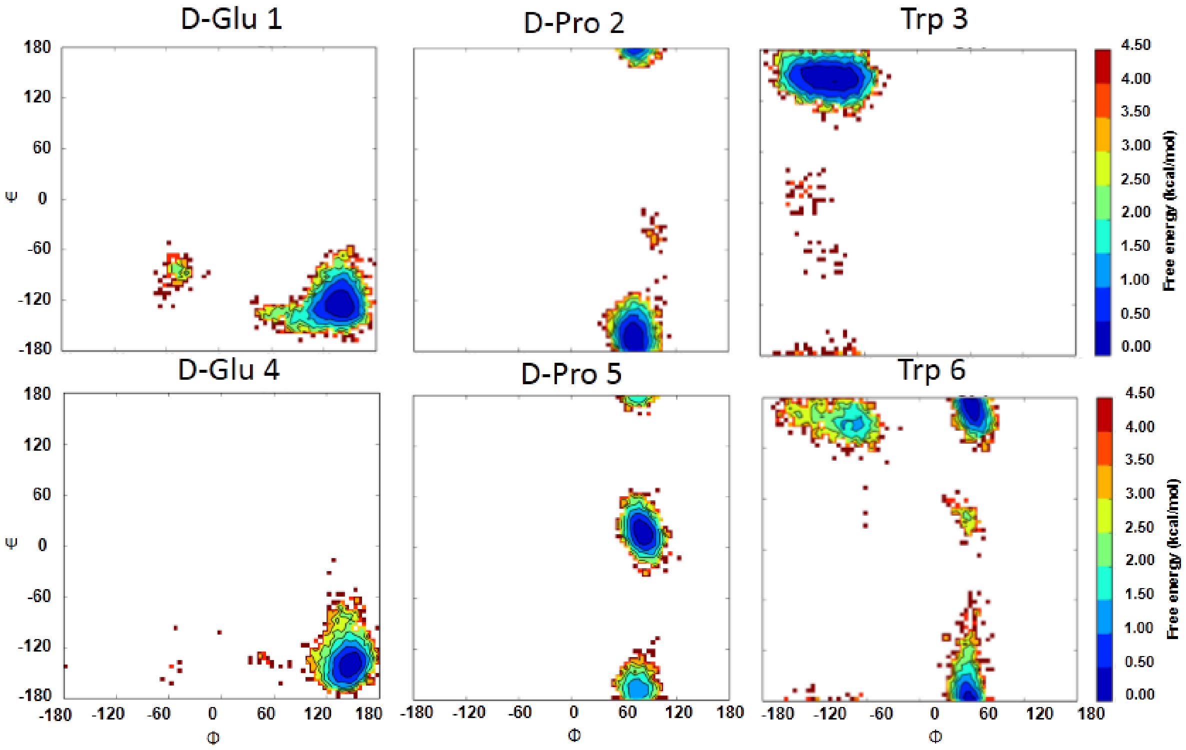


Figure 43. *Ramachandran plots obtained from REMD for each residue of peptide 7-C04 all cis conformer.*

The first surprising observation is that different Ramachandran plots were obtained for the same type of residue, for example between D-Pro 2 and D-Pro 5. This suggests the presence of non-symmetric conformations that can correspond to the minor species observed in the 1D NMR spectrum of this peptide. When analyzing the Ramachandran plots, we observed two possible conformations corresponding to βI' turns with glutamic residues in i+1 positions (Φ= 60, Ψ= 30) and proline in the i+2 positions (Φ= 90, Ψ= 0). This correlates with the conformation observed for the three most populated states obtained by REMD. Nevertheless, other conformations are also sampled in the Ramachandran plot; such as a pseudo βII' turn with proline in position i+1 (Φ= 60, Ψ= -120), tryptophan in position i+2 (Φ= -80, Ψ= 0) and glutamic in an extended conformation. If we observe the structure obtained by REMD for the all-*trans* conformer we noticed that, in this conformation, proline suffers from significant steric hindrance due to the proximity of the tryptophan side chain. This feature can explain the preference of peptide 7-C04 to adopt an all-*cis* conformation where this steric hindrance is alleviated by setting tryptophan side chains in an equatorial position.

In the Ramachandran plots obtained for the all *cis* conformation of peptide 7-C04, again we observed different Ramachandran plots for the same type of residue, suggesting the presence of non-symmetric conformations. The presence of inverse βVIb can be confirmed with these plots, having the proline residues in the i+2 position (Φ= 75, Ψ= -160) and glutamic residues in the i+1 position (Φ= 135, Ψ= -135). However, in the Ramachandran plot obtained for D-Pro 5 residue we also observed another sampled region corresponding to Φ= 60, Ψ= 0, suggesting the presence of pseudo βVIa2 turns.

The results obtained by REMD for peptide 7-C04, are in agreement with those from the NMR analysis, suggesting the presence of different conformational ensembles that are in slow equilibrium exchange in the NMR timescale.

Finally, as peptide 7-C04-P1 showed a major conformational ensemble with *cis* Xaa-Pro bonds, we decided to analyze this one to compare it with the results obtained for peptide 7-C04.

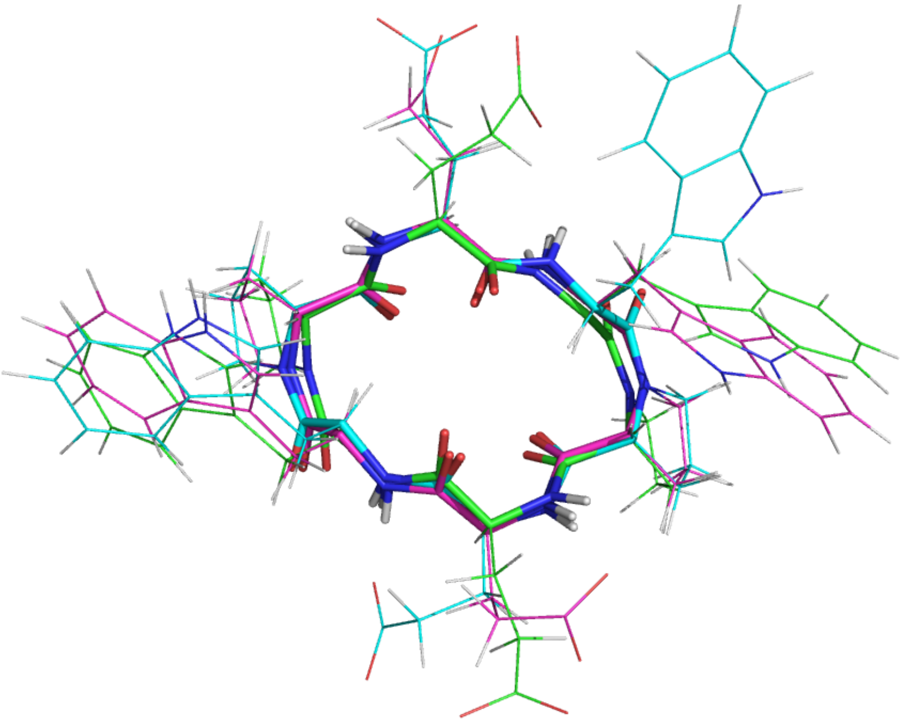


Figure 44. *Three most commonly populated states obtained from REMD for peptide 7-C04-P1 (RMSD = 1.56 ± 0.84).*

We observed a high consensus between the three most commonly populated states which account for almost 70% of the total population. Again no hydrogen bond was observed in these structures suggesting the presence of type βVIb turns. In order to confirm the presence of this type of β-turns, we analyzed the Ramachandran plots obtained by REMD. We observed a conserved symmetry for these peptides in contrast with the results obtained for 7-C04.

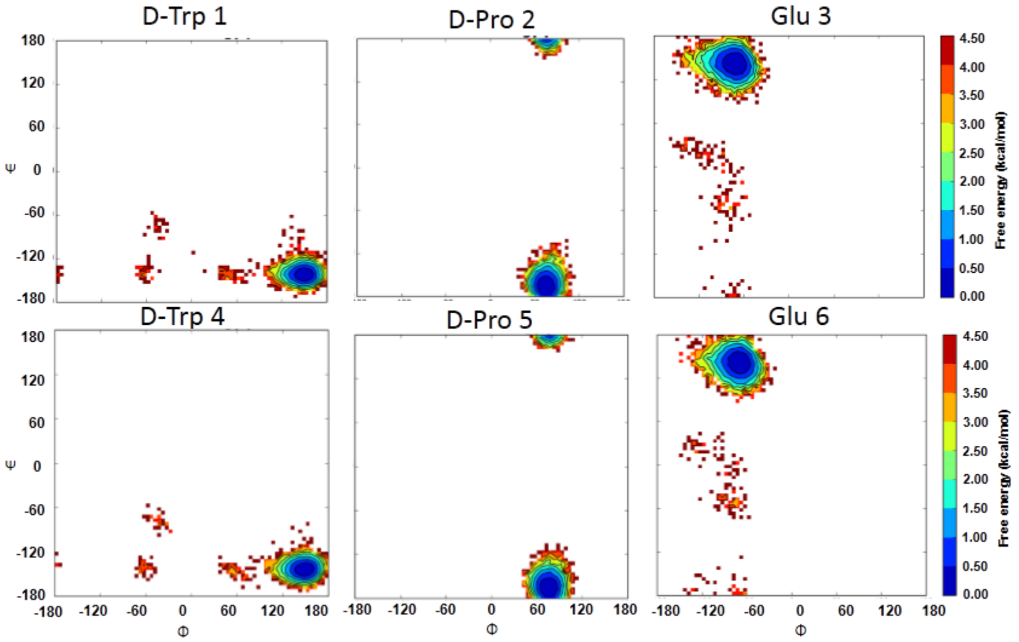


Figure 45. *Ramachandran plots obtained from REMD for each residue of peptide 7-C04-P1.*

The presence of inverse pseudo-βVIb turns was unambiguously confirmed by the Ramachandran plots. The proline residue, which may be in the i+2 position of the β-turns, sampled the region corresponding to Φ= 75, Ψ= -160, that are exactly the opposite of those described for βVIb turn (Φ= -75, Ψ= 160). Tryptophan residues sampled the region corresponding to Φ~ 150, Ψ~ -150, which are close to those described for residue i+1 in βVIb turns (Φ= -135, Ψ= 135).

REMD combined with NMR helped us to define the different β-turn types present in cyclic hexapeptides. To modulate PPIs, this can be a very useful tool to design desired types of β-turns that adopt a bioactive conformation for specific targets.

Conformational analysis of apamin analogues

This chapter is partially based on the following article:

Oller-Salvia, Benjamí; Sánchez-Navarro, Macarena; **Ciudad, Sonia**; Guiu, Marc; Arranz-Gibert, Pol; Garcia, Cristina; Gomis, Roger R.; Cecchelli, Roméo; García, Jesús; Giralt, Ernest; Teixidó, Meritxell. *MiniAp-4: A Venom-Inspired Peptidomimetic for Brain Delivery*. Angew Chem Int Ed Engl, 2016, 55(2), 572-5.

Venoms have evolved to attain high affinity and selectivity for a wide range of biological targets. The recent miniaturization of bio-assays and the development of high-throughput techniques have renewed the interest in these products in drug discovery programs.[208]

One of the most relevant properties of some venoms is their capacity to reach the central nervous system (CNS) through blood, which allows noninvasive administration of drugs and provides the most homogenous distribution in the brain.[209] However, to penetrate the brain parenchyma, substances have to cross the blood–brain barrier (BBB). The BBB is a highly complex cerebrovascular system comprised of approximately 100 billion capillaries, whose endothelial cells are strictly joined with tight junctions[210] and surrounded by astrocytes, pericytes, and macrophages, found throughout the dense and complex barrier that protects the CNS. Designing small molecules that are able to pass through this barrier is a challenge and one that is unfeasible for most large molecules: more than 98% of drugs intended for the therapeutic treatment of CNS diseases are never commercialized because of their inability to reach their targets.[209]

One of the most promising non-invasive strategies under investigation for drug delivery to the CNS is the use of BBB shuttles.[211] These vectors are able to enhance the passage of compounds across the BBB without altering its integrity. However, an important drawback of existing sequences is their high lability to serum proteases. Our lab and others have recently shown that peptides with reversed sequences and D-amino acids (retro–enantio approach) can be more efficient carriers than their L-counterparts.[44, 212, 213] Nonetheless, this strategy is relatively costly and may decrease the affinity of the peptide for the receptor that mediates its transport. A way to circumvent these drawbacks could be the use of natural cyclic peptides targeting the CNS, such as those found in venoms. Still, an inconvenience of using these compounds as BBB shuttles is their high toxicity.

The bicyclic peptide apamin, found in bee venoms, crosses the BBB and block calcium-mediated potassium channels.[214] Despite the proven CNS-targeting capacity of apamin,[215] the extended application of this molecule has been limited by its toxicity and high immunogenicity. Dr. Benjamí Oller-Salvia, recently demonstrated the ability of apamin and its non-toxic analogue (ApOO) to pass through an *in vitro* BBB model.[143] In his work, he excluded passage of substances by simple diffusion (PAMPA test) as the mechanism of action and suggested active transport through receptor mediated transcytosis. He also hypothesized a specific cell passage, as the peptides did not permeate through a monolayer Caco-2 cell model. Due to the resistance of the two peptides to the various pHs, temperatures, serum proteases, and coefficients of permeation, similar to other shuttle peptides, apamin and ApOO were proposed as potential carriers for drug delivery to the CNS. However, a specific passage mechanism could be specified.

A recent study was conducted in our lab, by Dr. Benjamí Oller-Salvia and Dr. Macarena Sánchez, with the aim of obtaining all L-protease-resistant shuttles by minimizing apamin, while reducing its toxicity and immunogenicity. To simplify the structure of apamin, they synthesized the native peptide and a non-reported analog that does not contain the main residues responsible for its toxicity (MiniAp-1). They assayed both compounds in a bovine-cell-based BBB model and found that permeability of MiniAp-1 was 60% higher than that of apamin (**Figure 46b**). The strong inhibition of MiniAp-1 transcytosis by a change in temperature or the addition of sodium azide indicated that the peptide was transported mainly through an active mechanism. These results encourage them to dissect the structure of MiniAp-1 to prepare minimized derivatives based on the loop between Cys3 and Cys11. They first synthesized a linear version (MiniAp-2), which was transported less efficiently than MiniAp-1 and readily degraded by serum proteases (**Figure 46**). They then prepared two monocyclic analogues, one by connecting Cys3 and Cys11 (MiniAp-3) and the other by switching the disulfide to a lactam bridge (MiniAp-4). MiniAp-3 showed permeability values similar to those of MiniAp-1, and was over 30 times more resistant to serum proteases than the linear version. Remarkably, the permeability of MiniAp-4 was 50% higher than that of MiniAp-3, which could be attributed to its impressive resistant to serum proteases (**Figure 46**).

Aiming to further explain the differences in BBB transport between apamin derivatives, they decided to study their conformational preferences. The α-helical structure of MiniAp-1, typical of apamin[216] was confirmed by circular dichroism (CD; **Figure 46d**). While the CD spectrum of the linear analogue (MiniAp-2) showed a clear random coil behavior, no significant structural information could be concluded for MiniAp-3 and MiniAp-4 from their CD spectra (**Figure 46d**).

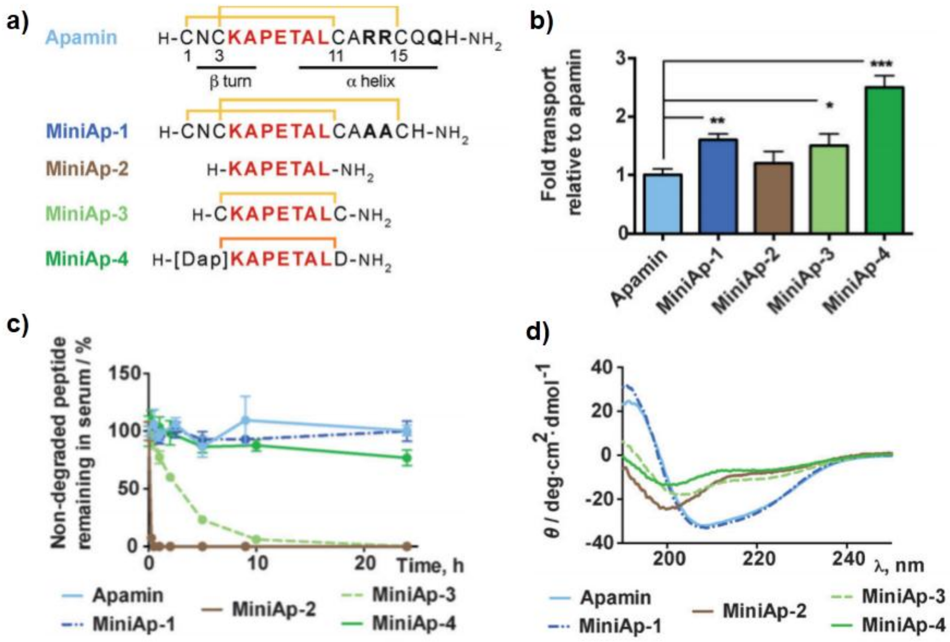


Figure 46. *Comparison of mini-apamin shuttle candidates. a) Peptide sequences. The disulfides are depicted in yellow and the lactam bridge in orange. b) Relative transport of peptides in the bovine-cell-based model. c) Stability of the peptides in human serum. d) CD spectra.*

After seeing these results we decided to conduct an NMR structural analysis of these apamin analogues. In this chapter the results obtained for the conformational analysis of the four apamin analogues by NMR are shown. Our aim was to establish a structure-activity relationship for these molecules to better understand the differences in BBB transport between them.

Conformational analysis of MiniAp-1 by NMR

The interest of apamin lies not only in its biological activity but also in its characteristic structure. The shape of this peptide is defined by a β-turn at the N-terminus and an α-helix, which are tightly held together by two disulfide bonds. This compact structure confers very high stability in a wide range of pH values and temperatures.

In an attempt to explain the differences observed in the permeability of apamin analogues across the BBB by Dr. Oller-Salvia, we decided to perform a detailed conformational study of these peptides by nuclear magnetic resonance (NMR).

In order to compare the results obtained for apamin analogues with those of apamin, all the experiments were performed at the same conditions previously described[216], working at pH 2-3. As the *in-vitro* evaluation of these peptides to assess their BBB permeability was performed at physiological pH, NMR experiments were also acquired in 10 mM sodium phosphate buffer at pH 7.4 to check if their conformational behavior may be influenced by pH changes. 1D ^{1}H NMR and standard 2D experiments such as TOCSY, DQF-COSY, NOESY and ^{1}H-^{13}C HSQC were acquired for peptide characterization. Structure based NMR parameters such as chemical shift deviations, temperature coefficient values of amide NH protons, $^{3}J_{\alpha NH}$ coupling constants and NOE patterns; were evaluated to study the conformation of these peptides.

First, we evaluated if the bicyclic apamin analogue (MiniAp-1), synthesized by Dr. Oller-Salvia, maintains the highly compact structure of apamin.

As it was previously described for apamin, two distinct sets of resonances were observed in the ^{1}H NMR spectrum of MiniAp-1, with relative populations of 90% and 10% (**Figure 47**). Despite complete assignment was not possible for the minor species, due to severe signal overlapping, it was identified as the *cis* Xaa-Pro bond conformer on the basis of strong NOE cross-peak between the H_{α} of Ala5 and the H_{α} of Pro6 (**Figure 47d**), and the Pro C_{β} - C_{γ} chemical shift differences ($\Delta\delta$ = 9.2 ppm).[194] The chemical shift differences between the Pro C_{β} and C_{γ} atoms ($\Delta\delta$ = 4.4 ppm) and the characteristic sequential NOE observed between the H_{α} of Ala5 and the H_{δ} of Pro6 (**Figure 47c**) confirmed that the *trans* rotamer was the major MiniAp-1 species.

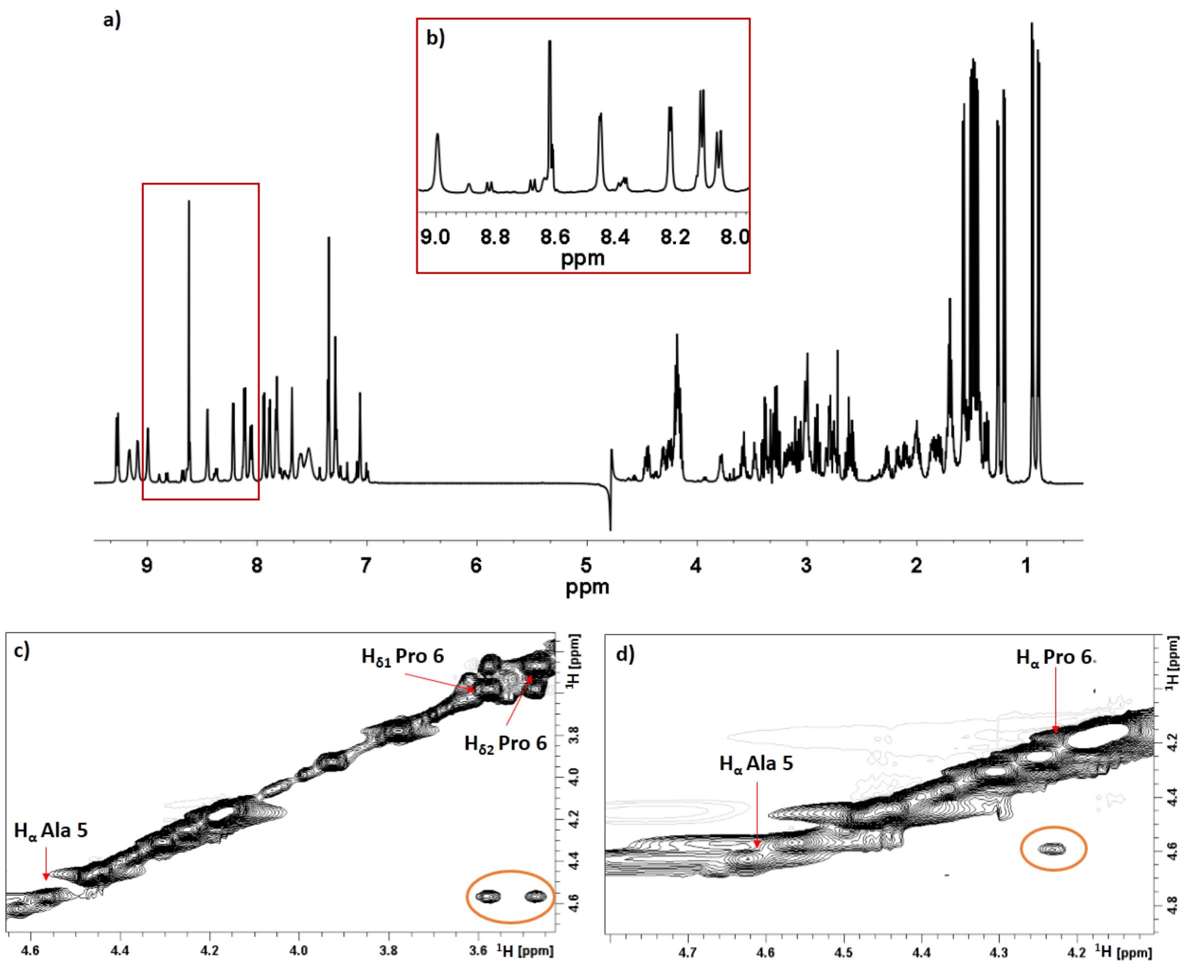


Figure 47. *a) ^{1}H NMR spectrum of MiniAp-1 acquired at pH 2-3 and 298 K. b) Zoomed region showing the presence of a minor species. c) Expanded regions of the NOESY spectrum of MiniAp-1 showing the cross-peaks between the H_α of Ala5 and the H_δ of Pro6. d) Expanded regions of the NOESY spectrum of MiniAp-1 showing the cross-peak between the two H_α of Ala5 and Pro6.*

Structural characterization of the *trans* conformer was done on the basis of NMR parameters such as H_α and C_α secondary chemical shifts, temperature coefficients of NH amide protons and NOEs connectivities.

Chemical shift deviations of both H_α and C_α showed significant information regarding the structure of this peptide. Consecutive deviations from random coil values, negative for H_α and positive for C_α for the Ala9 - His16 segment clearly indicated the presence of an α-helical fold in this region (**Figure 48**).

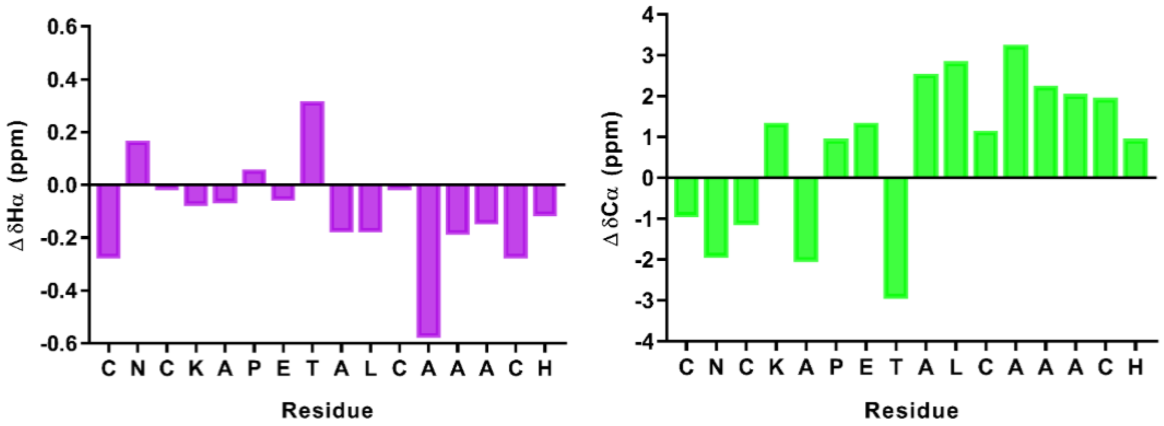
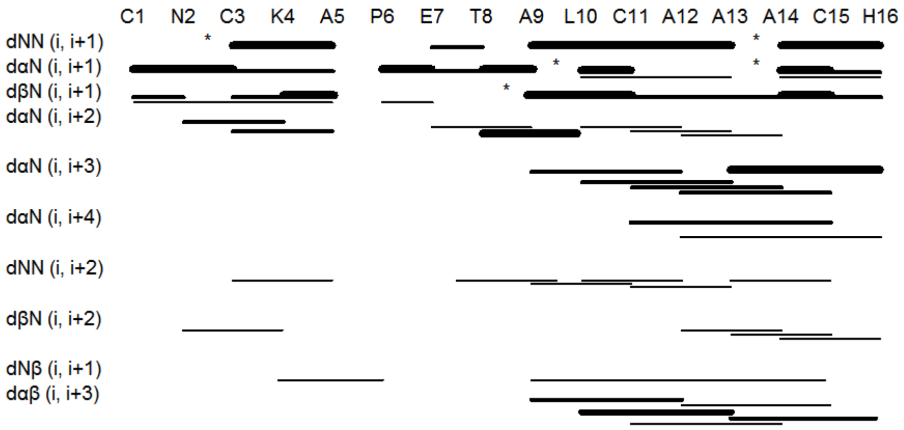


***Figure 48**. Histograms of chemical shift deviations from random coil, for ΔH_{α} (purple) and ΔC_{α} (green) in MiniAp-1. Random coil values were extracted from Wishart et al.*[193]

Inspection of the $^{3}J_{\alpha NH}$ coupling constants together with the temperature coefficients of NH amides (**Table 18**) and NOE connectivities (**Figure 49**), suggested that MiniAp-1 adopted a conformation similar to the one described for apamin. The presence of d_{NN} (i, i+1), $d_{\alpha N}$ (i, i+3), $d_{\alpha N}$ (i, i+4) and $d_{\alpha\beta}$ (i, i+3) NOEs in the Thr8 - Cys15 segment confirmed the presence of a regular α-helix. The observation of several non-sequential NOEs at the N-terminus and of medium-strong d_{NN} (i, i+1) NOEs between Cys3 - Lys4 and between Lys4 - Ala5 is consistent with a change of backbone direction, indicative of a turn-like structure, as previously described for apamin.



α C1 - βC11	*	α C3 - β C15	w	β P6 - NH A12	w	NH T8 - β C11	m
β N2 - NH2 N2	s	NH A5 - α P6	w	β P6 - α A12	w	γ T8 - NH A9	w
β A5- NH2 N2	w	NH A5 - δ P6	s	γ P6 - α A12	w	γ T8 - NH L10	w
β N2 - NH A5	w	α A5 - δ P6	s	NH E7 - β C11	w	γ T8 - NH C11	w

***Figure 49.** Summary of NOE connectivities observed for MiniAp-1 (mixing time 250 ms). The thickness of the bar indicates the strength of the NOE (weak, medium or strong). The stars indicate signal overlap.*

$^3J_{\alpha NH}$ coupling constants bigger than 8 Hz in the region Cys1 - Ala5 corroborates the presence of a turn in this segment. In contrast, small $^3J_{\alpha NH}$ coupling constants in the segment Thr8 - Ala15 are in agreement with those residues adopting an α-helical structure. Finally, small temperature coefficient values in both Cys1 - Ala5, and Thr8 - His16 also support the chemical shift- and NOE-based conformational analysis.

Table 18. *$^3J_{\alpha NH}$ coupling constants and temperature coefficient values ($-\Delta\delta_{NH}/\Delta T$) for MiniAp-1.*

Residue	$^3J_{\alpha NH}$ (Hz)	$-\Delta\delta_{NH}/\Delta T$ (ppb/K)
Cys 1	9.2	-
Asn 2	5.3	4.0
Cys 3	8.1	8.5
Lys 4	7.5	1.5
Ala 5	-	2.0
Pro 6	6.5	-
Glu 7	broad signal	7.5
Thr 8	3.5	3.0
Ala 9	4.5	5.5
Leu 10	6.1	8.0
Cys 11	3.8	4.5
Ala 13	5.0	4.0
Ala 14	5.0	4.0
Ala 15	6.1	3.0
His 16	7.2	4.5

The results obtained by NMR for MiniAp-1 clearly indicated that this peptide in solution adopted a similar conformation than the one described for apamin (**Figure 50**).

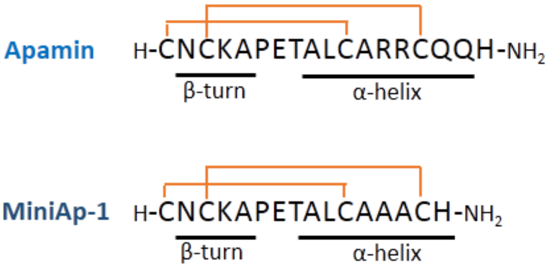
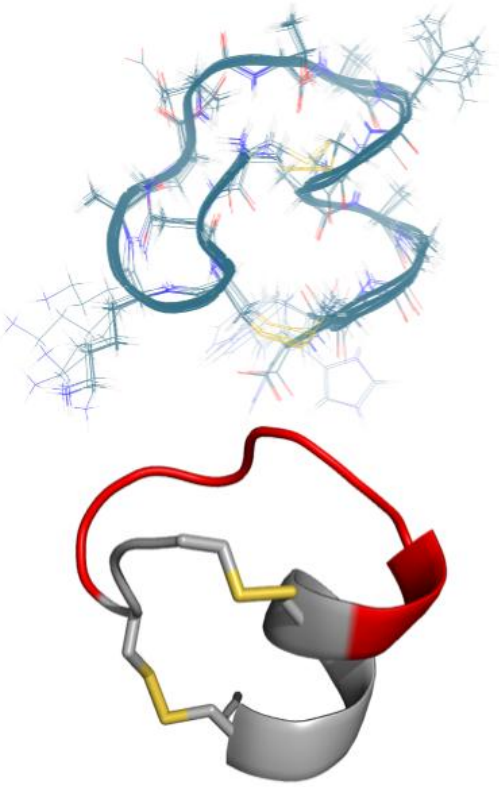


Figure 50. *Structural scheme comparing apamin with MiniAp-1.*

NMR conformational analysis of MiniAp-1 was also performed at pH 7.4 (results in Section: Product characterization). Structure-based NMR parameters such as secondary chemical shifts, $^3J_{\alpha NH}$ coupling constants, NOE connectivities and amide temperature dependence were perfectly in agreement with those obtained at acidic pH.

To determine the three dimensional structure of MiniAp-1, Dr. Jesús García conducted a simulated annealing calculation by applying distance and dihedral angle restraints. An overlay of the ten lowest-energy structures from the calculation is shown in **Figure 51**. The structure obtained for MiniAp-1 consisted of an N-terminal loop (residues 1-8) and a C-terminal α-helix (residues 9-15) which is very similar to that of apamin,[216] with an RMSD value of 0.72 Å for the backbone superposition of both structures. However, they display some local differences in the loop orientation and in the C-terminal residue which is less defined in the MiniAp-1 structure.



Restraints	
NOE	
Sequential ($\|i-j\| = 1$)	21
Medium range ($1 < \|i-j\| \leq 4$)	25
Long range ($\|i-j\| > 4$)	6
Dihedral angle restraints	
ϕ	8
ψ	8
Quality	
Ramachandran Analysis	
Most favoured regions	12 (92.3%)
Additional allowed regions	1 (7.7%)
Generously allowed regions	0 (0%)
Disallowed regions	0 (0%)

Figure 51*. NMR structure of MiniAp-1. Left) Ensemble of the best 10 structures of MiniAp-1. Root-mean-square deviation (RMSD) values of 0.24 and 1.31 Å were calculated for the backbone and heavy atom superimposition respectively. Cartoon representation of MiniAp-1. The loop on which MiniAps-2, -3 and -4 are based is highlighted in red and the disulfide bonds are shown in yellow. Right) Structural statistics for MiniAp-1.*

Conformational analysis of MiniAp-2, MiniAp-3 and MiniAp-4 by NMR

NMR experiments were performed both at pH 2-3, for comparison reasons, and at pH 7.4 since the BBB permeability assay of these peptides was performed at physiological pH. As for MiniAp-1, 1D ^{1}H NMR and standard 2D experiments such as TOCSY, DQF-COSY, NOESY and ^{1}H-^{13}C HSQC were acquired for peptide characterization.

MiniAp-2 random coil conformation observed by circular dichroism was confirmed by NMR as this peptide did not showed any evidence of secondary structure elements in terms of H_{α} and C_{α} chemical shift deviations (**Figure 52**), NOE connectivities, $^{3}J_{\alpha NH}$ coupling constants (**Table 19**) and temperature coefficients of amide protons (**Table 20**).

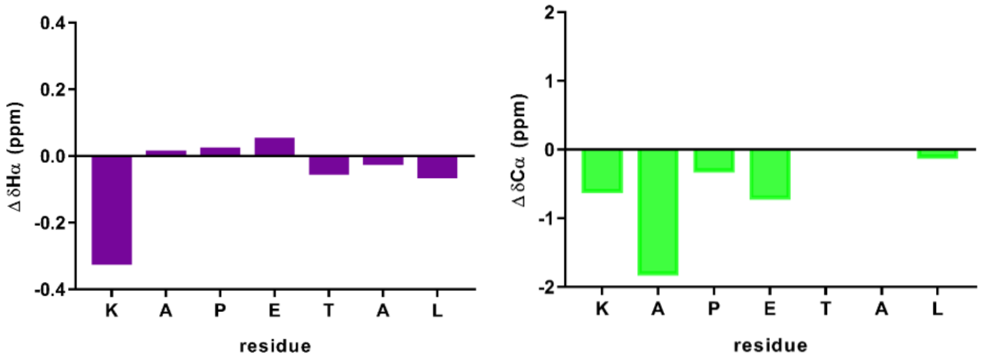


Figure 52. *Histograms showing the chemical shift deviations from random coil observed in H_{α} (purple) and C_{α} (green) for MiniAp-2. Random coil values were extracted from Wishart et al.*[193]

Notably, the NMR spectra of both MiniAp-3 and MiniAp-4 showed equally populated *cis* and *trans* conformers; while the linear analogue showed only 10% of the *cis* rotamer (**Figure 53**).

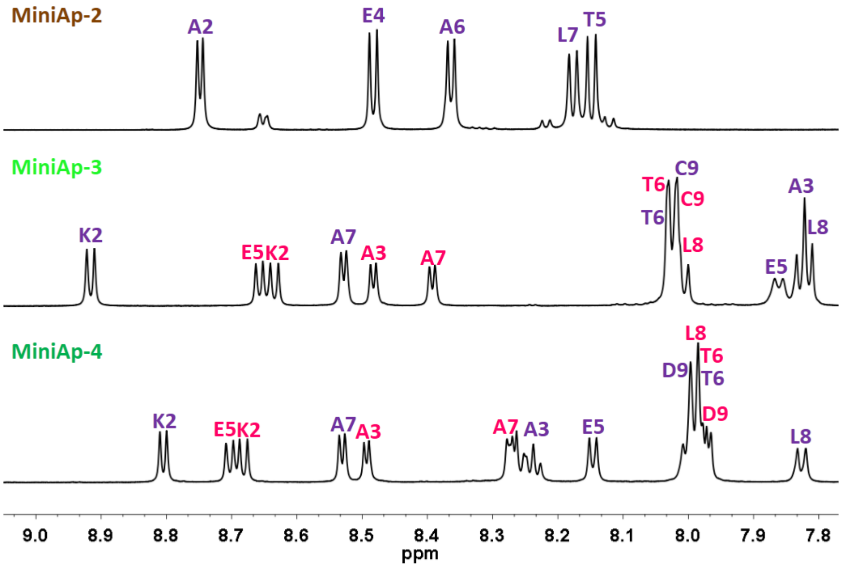


Figure 53. *Comparison of the amide regions of the 1D ^{1}H NMR spectra of MiniAp-2, MiniAp-3 and MiniAp-4. The amide resonances of the trans (purple) and the cis (pink) conformers are labeled.*

The *cis* and *trans* species were identified on the basis of strong NOE cross-peaks between the H_α of Ala2 (in MiniAp-2) or Ala3 (in MiniAp-3 and -4) and the H_δ of the Pro, in the case of the *trans* conformer; and between the H_α of Ala2 (in MiniAp-2) or Ala3 (in MiniAp-3 and -4) and the H_α of the Pro (**Figure 54**) for the *cis* conformer. This observation was further confirmed by the chemical shift difference between the proline C_β and C_γ atoms: *trans*-Pro C_β - C_γ ($\Delta\delta$ = 4.7 ppm for MiniAp-2 and 4.5 ppm for both MiniAp-3 and MiniAp-4) and *cis*-Pro C_β - C_γ ($\Delta\delta$ = 9.6 ppm for MiniAp-2 and 9.2 for both MiniAp-3 and MiniAp-4).

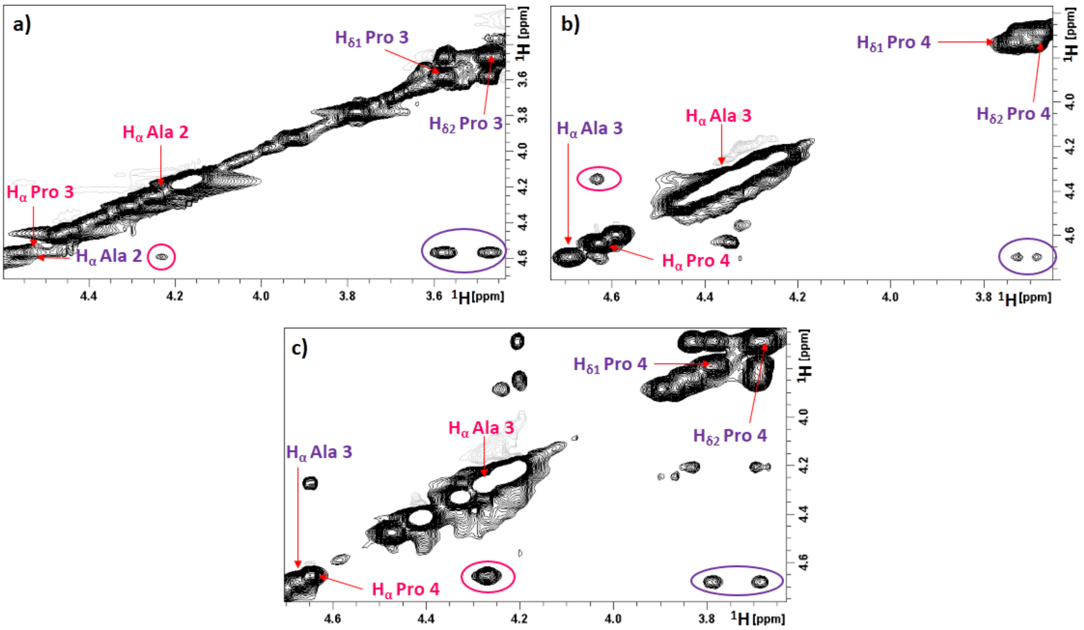


Figure 54. *Regions of the NOESY spectra of MiniAp-2 (a), MiniAp-3 (b) and MiniAp-4 (c) showing the cross-peaks between the H_α of Ala and the H_δ of Pro for the trans conformer (purple) and the cross-peak between the H_α protons of Ala and Pro for the cis conformer (pink).*

The close resemblance of the $^{13}C_\alpha$ chemical shifts (**Figure 55**), $^{3}J_{\alpha NH}$ coupling constants (**Table 19**) and NH temperature coefficients (**Table 20**) between the *trans* conformers of both peptides on one side and between the *cis* conformers on the other side; suggested that the backbone conformation adopted for MiniAp-3 and MiniAp-4 were very similar.

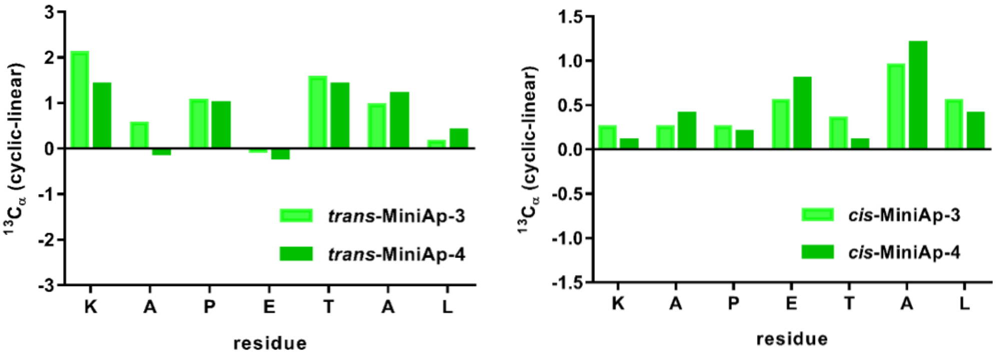


Figure 55. *Comparison of the $^{13}C_\alpha$ chemical shift values of the linear derivative (MiniAp-2) and the trans (left) and cis (right) conformers of the cyclic analogues (MiniAp-3 and MiniAp-4).*

Nonetheless, the above mentioned NMR parameters did not suggested a defined secondary structure for either the *cis* or *trans* conformers of both monocyclic derivatives. In addition, in the NOESY spectra of both analogues only sequential NOE connectivities were observed, indicating that these peptides do not exhibit any dominant secondary structure.

Table 19. *$^{3}J_{\alpha NH}$ coupling constants for peptides MiniAp-2, MiniAp-3 and MiniAp-4.*

Residue	**$^{3}J_{\alpha NH}$ (Hz)**				
	MiniAp-2 (trans)	**MiniAp-3 (trans)**	**MiniAp-3 (cis)**	**MiniAp-4 (trans)**	**MiniAp-4 (cis)**
Cys/Dapa		-	-	-	-
Lys	-	6.9	7.5	5.9	7.3
Ala	5.1	7.1	4.9	6.5	4.5
Pro	-	-	-	-	-
Glu	6.7	7.7	6.5	overlapped	6.7
Thr	7.5	7.1	7.9	5.9	7.8
Ala	6.0	4.9	5.0	4.8	overlapped
Leu	7.1	7.5	overlapped	7.4	overlapped
Cys/Asp		7.7	7.4	7.5	overlapped

Table 20. *Temperature coefficient values ($-\Delta\delta_{NH}/\Delta T$) of backbone amide protons for peptides MiniAp-2, MiniAp-3 and MiniAp-4.*

Residue	**$-\Delta\delta_{NH}/\Delta T$ (ppb/K)**				
	MiniAp-2 (trans)	**MiniAp-3 (trans)**	**MiniAp-3 (cis)**	**MiniAp-4 (trans)**	**MiniAp-4 (cis)**
Cys/Dapa		-	-	-	-
Lys	-	5.0	6.0	5.0	6.0
Ala	7.0	7.0	8.0	6.5	8.5
Pro	-	-	-	-	-
Glu	9.0	13.5	5.0	8.5	4.0
Thr	8.0	0.5	9.0	2.5	7.5
Ala	9.0	5.5	6.0	7.0	4.5
Leu	9.0	8.0	7.5	7.5	9.5
Cys/Asp		6.5	5.0	5.0	8.5

These results were in agreement with the CD spectra obtained for these analogues which were very similar to that of the MiniAp-2. However, when compared to the linear peptide, the *trans* conformer of both monocyclic derivatives (MiniAp-3 and MiniAp-4) showed significantly larger amide chemical shift dispersion (**Figure 53**), suggesting less conformational flexibility. Significant differences in $^{13}C_{\alpha}$ chemical shifts were also observed between each conformer of both monocyclic derivatives and the linear peptide, further suggesting that cyclization constrains the peptide (**Figure 55**).

Interestingly, the temperature coefficients ($-\Delta\delta_{NH}/\Delta T$) of the Thr 6 NH were 0.5 and 2.5 ppb/K for the *trans* MiniAp-3 and the *trans* MiniAp-4, respectively (**Table 20**); suggesting that these amide protons may be involved in an intramolecular hydrogen bond in the conformation adopted by the *trans* conformers. In contrast, the higher temperature dependence of the Thr 6 NH chemical shift in the *cis* conformers ($-\Delta\delta_{NH}/\Delta T$ = 9 and 7.5 ppb/K for MiniAp-3 and MiniAp-4, respectively) was indicative of a solvent exposed amide.

NMR conformational analysis of MiniAp-3 and MiniAp-4 was also performed at pH 7.4 (results in Section: Product characterization). Structure-based NMR parameters such as secondary chemical shifts, $^{3}J_{\alpha NH}$ coupling constants, NOE connectivities and amide temperature dependence were perfectly in agreement with those obtained at acidic pH.

For these monocyclic analogues (MiniAp-3 and MiniAp-4) we did not obtained a three-dimensional structure, as it was done for MiniAp-1, as they were too flexible to provide a reliable model.

We hypothesized that the higher preference for the *cis* Ala-Pro rotamer in MiniAp-4 with respect to MiniAp-1 could partly account for the differences in permeability between the two analogues with highest resistance to proteases.

Development of fluorescence-based biosensors for protein detection

The work described in this chapter was done, as part of a short stay, in Prof. Imperiali's lab in collaboration with Prof. Wittrup's lab at the Massachusetts Institute of Technology (MIT) in Boston.

The field of biosensors has seen in the last decade a multitude of new approaches for the application of reagentless sensors. An ideal sensor would report the concentration and localization of a target analyte in real-time without interfering with its function. The overall strategy is the combination of a recognition unit and a signal-transducing unit into one molecular entity. Sensors such as those based on Förster Resonance Energy Transfer (FRET) couple the analyte recognition and signal transduction event into a single step and provide temporal and spatial resolution.[217] However, such sensors require the introduction of two unique fluorophores into either two binders or into proteins, which undergo large conformational changes upon binding. Solvatochromism is a more recent development, but is becoming more widely used as new scaffolds (affinity molecules) and dyes (environmentally-sensitive fluorophores) are developed and become available. In solvatochromism, an environment-sensitive fluorophore is coupled to a single analyte-specific binder.[218] Changes in the environment of the fluorophore (such as solvent shielding) upon analyte binding transduce the binding signal into a fluorescence read-out (**Figure 56**). Several groups have successfully developed solvatochromic-based biosensors using DNA aptamers,[219-221] native protein receptors,[145, 222] peptides[223, 224] or engineered binders using protein scaffolds.[225-230]

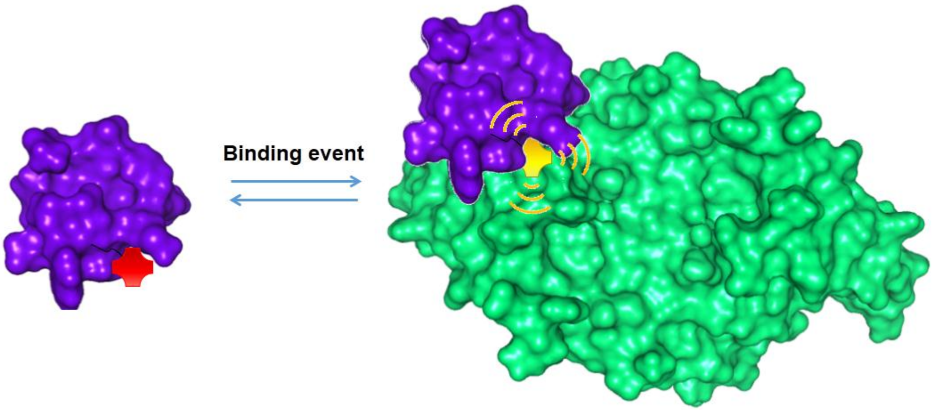
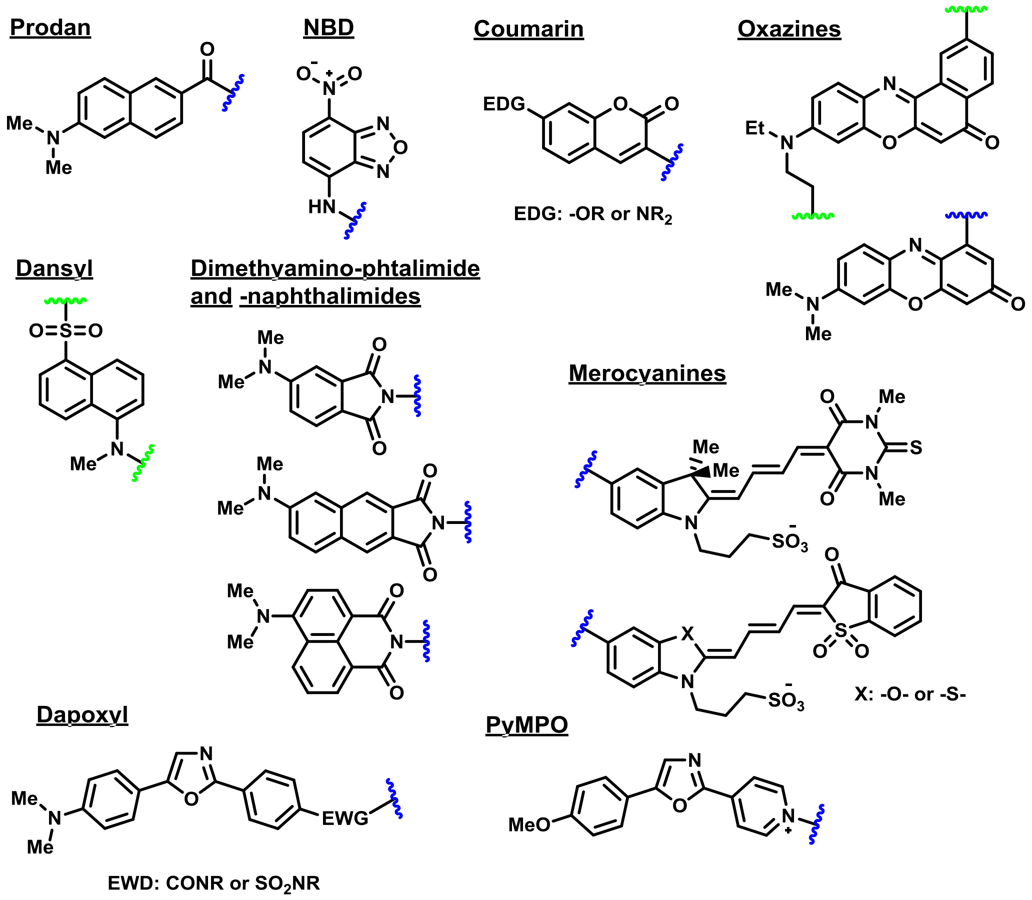


***Figure 56**. Applications of solvatochromism to report protein-protein interactions*

Protein studies based on solvatochromism have evolved with the use of both intrinsic and extrinsic fluorescence species. For example, the fluorescence of the natural amino acid tryptophan has long been known to be environment-dependent[231] and has been widely used in both folding and ligand-binding studies. However, the short wavelengths required for indole excitation and the relative abundance of tryptophan in nature strongly limits its potential for applications in complex systems. This has prompted the design and application of extrinsic synthetic fluorophores with improved photophysical properties. These efforts have led to the emergence of a host of solvatochromic probes with diverse properties (**Figure 57**).



Fluorophore family	ε M^{-1} cm^{-1}	λ_{excmax} (nm)	MW range g·mol^{-1}	Charge (+/-)
Prodan	2.0×10^4	390	200	Neutral
NBD	2.5×10^4	480	180	Neutral
Coumarin	$2.4 \times 10^4 – 5.0 \times 10^4$	376 – 419	160 – 215	Neutral
Oxazines	$1.4 \times 10^4 – 5.2 \times 10^4$	530 – 598	240 – 335	Neutral
Dansyl	$4.1 \times 10^3 – 5.7 \times 10^3$	336 – 340	235 – 265	-/neutral
Dimethylamino-phtalimide/naphthalimides	$6.5 \times 10^3 – 8.8 \times 10^3$	378 – 408	190 – 240	neutral
Merocyanines	$1.1 \times 10^5 – 2.2 \times 10^5$	570 – 605	440 – 500	+/-/neutral
Dapoxil	2.7×10^4	366 – 370	280 – 330	Neutral
PyMPO	2.3×10^4	412	250	+

Figure 57. *Representative structures of the different families of solvatochromic fluorophores and summary table of the solvatochromic fluorophores properties. The position of attachment to biomolecules, either directly or via a reactive group, are indicated by wavy lines (blue when the position is unique and green when multiple alternative anchoring positions have been developed). Points of substituent variability within the structures are indicated by EDG (electron donating group), EWG (electron withdrawing group) and X (heteroatom). Adapted from Trends in Biotechnology.*[232]

Currently, fluorescence-based biosensors are designed in three steps. First, a binder is engineered against the intended target using established display technologies[233-236] and multiple scaffolds,[237-239] ideally with tailored binding kinetics and affinity.[240] The second step consists on carefully choosing the labeling site, taking into account that the residue must be: 1) exposed to solvent, 2) non-perturbing to the analyte/scaffold interaction, and 3) close to the binding epitope. Finally, a fluorophore is conjugated to a specific site on the binding protein using established chemical[241] or enzymatic[242] techniques. The methods for introducing solvatochromic fluorophores into peptides and proteins are based on similar approaches for incorporating other unique functionalities, such as reactive crosslinking groups and affinity tags. However, in contrast to many other species, the insertion of solvatochromic fluorophores is topologically restricted to sites in the protein that preserve function and activity while permitting the dye to make necessary contacts that will result in measurable fluorescence changes. This consideration necessitates the use of methods that offer precise control over dye placement within the peptide or protein structure of interest with minimal perturbation.

Insertion of solvatochromic fluorophores into peptides can generally be accomplished through standard solid phase synthesis (SPPS) approaches to mutate a single native residue. For proteins, the most common methods can broadly be divided into three categories: (i) direct covalent modification; (ii) incorporation of fluorescent amino acids via semi-synthesis (expressed protein ligation); and (iii) incorporation of fluorescent amino acids via suppression of the amber (TAG) stop codon. Cysteine and lysine reactive agents offer a convenient and direct method for labeling proteins and have been used extensively to conjugate a myriad of auxiliary groups into biomolecules. Most common among these are the thiol-selective electrophiles, such as maleimides and α-halocarbonyl compounds, along with the amine-selective acylating agents such as the O-succinimidyl esters. In general, cysteine residues are ideal since this amino acid occurs relatively infrequently in proteins and possesses excellent nucleophilic properties under most physiological conditions. Proteins with unique cysteines can be readily prepared using standard molecular biology techniques.

The development of new solvatochromic fluorophores and methods to insert them into proteins of interest has greatly expanded the scope of potential applications for these unique chemical tools. Since the initial *in vitro* studies of proteins containing a single tryptophan, advances in this field have enabled researchers to conduct experiments in living cells to investigate the dynamics of protein activities in their native environment. Nowadays, many applications are known for these solvatochromic fluorophores such as: protein folding,[243, 244] structural information,[245-247] sensors for small molecules,[248-252] reporting conformational states,[253-255] direct reporting of post-translational modifications[256] or reporting binding interactions.[225, 257-259]

In solvatochromic biosensors, undesirable fluorophore pre-activation should be attenuated in a rigid protein scaffold. A rigid conformation (characteristic of the β-strands) should reduce the likelihood of fluorophore interactions with the scaffold itself, thereby reducing unwanted background fluorescence (**Figure 58**).

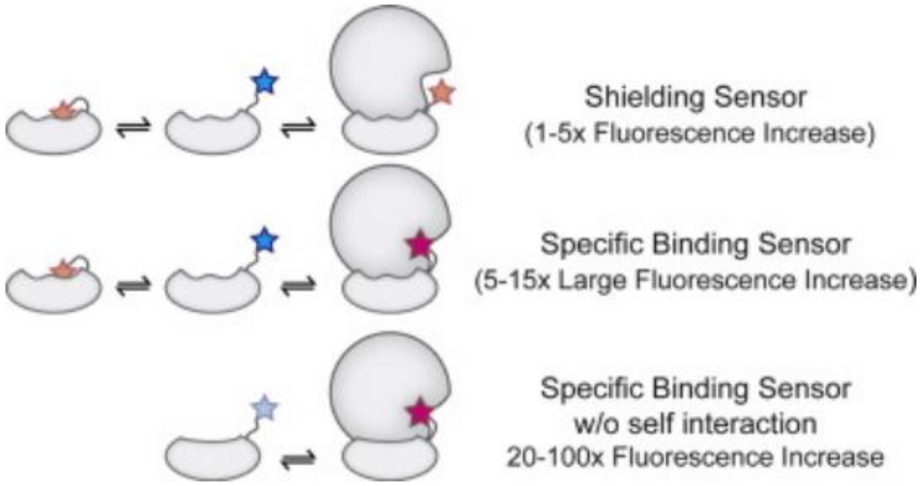


Figure 58. *Mechanistic drawing of the three types of reagentless biosensors. The top one describes a sensors that suffers from pre-activation and the fluorophore undergoes shielding upon binding, yielding a small F/F0 ratio. The middle one describes a sensor that may have some pre-activation but that alleviates this drawback by having a strong fluorescence enhancement due to a fluorophore binding pocket. The bottom scenario describes a sensor that possesses both ideal properties: little self-interaction and specific binding pocket interaction on the analyte. Extracted from Journal of Molecular Biology.*[260]

Hyperthermophilic proteins are excellent candidates for use as protein scaffolds to engineer molecular recognition. The ultrastable Sso7d protein from the hyperthermophilic archaeon *Sulfolobus solfataricus* can be used as a versatile scaffold to generate highly stable binding proteins for a wide range of target species. The Sso7d protein is a small (∼7 kDa, 63 amino acids) DNA-binding protein containing an SH3-like fold consisting of an incomplete β-barrel with five β strands and a C-terminal α helix.[261-263] Sso7d is highly stable with a melting temperature of nearly 100 °C and lacks cysteine.[264]

Yeast surface display (YSD) has been recently used for engineering proteins based on the Sso7d scaffold.[265, 266] The yeast display technique was first published by the laboratory of Professor K. Dane Wittrup. In this methodology, a protein of interest is displayed as a fusion to the Aga2p protein on the surface of yeast. The Aga2p protein is naturally used by yeast to mediate cell-cell contacts during yeast cell mating. As such, display of a protein via Aga2p projects the protein away from the cell surface, minimizing potential interactions with other molecules on the yeast cell wall. The use of magnetic separation and flow cytometry in conjunction with a yeast display library is a highly effective method to isolate high affinity protein ligands against nearly any receptor through directed evolution.[267] The major advantage of yeast display over other in vitro evolution methods is the ability of quantitative library screening through fluorescent-activated cell sorting (FACS).

Flow cytometry is a technology that simultaneously measures and then analyzes multiple physical characteristics of single particles, usually cells, as they flow in a fluid stream through a beam of light. Any suspended particle or cell from 0.2 to 150 micrometers in size is suitable for analysis. The properties measured include a particle’s relative size, relative granularity or internal complexity, and relative fluorescence intensity. These characteristics are determined using an optical-to-electronic coupling system that records how the cell or particle scatters incident laser light and emits fluorescence. A flow cytometer is made up of three main systems: fluidics (which transports particles in a stream to the laser beam for interrogation), optics (which consists of lasers to illuminate the particles in the sample stream and optical filters to direct the resulting light signals to the appropriate detectors), and electronics (which converts the detected light signals into electronic signals that can be processed by the computer. For some instruments equipped with a sorting feature, the electronics system is also capable of initiating sorting decisions to charge and deflect particles.

In the flow cytometer, particles are carried to the laser intercept in a fluid stream. Any suspended particle or cell from 0.2 to 150 micrometers in size is suitable for analysis. Cells from solid tissue must be disaggregated before analysis. The portion of the fluid stream where particles are located is called the sample core. When particles pass through the laser intercept, they scatter laser light (**Figure 59a**). Factors that affect light scattering are the cell's membrane, nucleus, and any granular material inside the cell. Cell shape and surface topography also contribute to the total light scatter.

Forward-scattered light (FSC) is proportional to cell-surface area or size and provides a suitable method of detecting particles greater than a given size independent of their fluorescence.

Side-scattered light (SSC) is proportional to cell granularity or internal complexity. It is collected at approximately 90 degrees to the laser beam by a collection lens and then redirected by a beam splitter to the appropriate detector.

Correlated measurements of FSC and SSC can allow for differentiation of cell types in a heterogeneous cell population (**Figure 59b**).

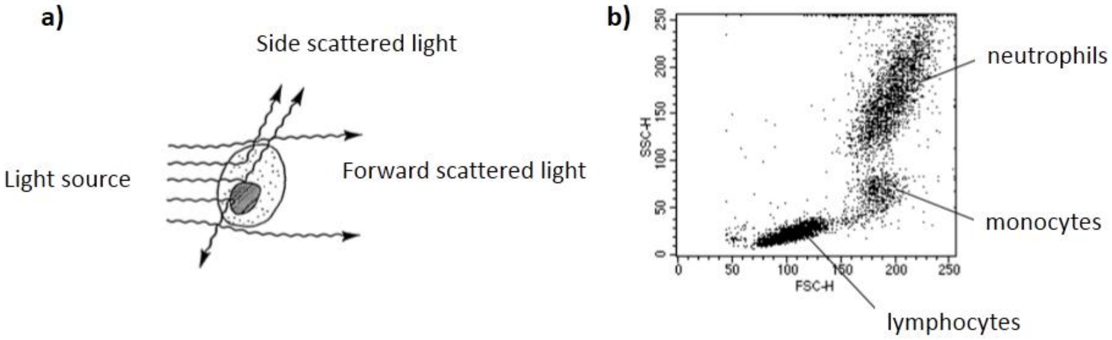


Figure 59. *Light scattering in flow cytometry. a) Light-scattering properties of a cell. b) Cell subpopulations based on FSC vs SSC.*

When particles are labeled with fluorophores, the amount of fluorescent signal detected is proportional to the number of fluorochrome molecules on the particle. When a fluorescent dye is conjugated to a monoclonal antibody, it can be used to identify a particular cell type based on the individual antigenic surface markers of the cell (**Figure 60**). In a mixed population of cells, different fluorochromes can be used to distinguish separate subpopulations. The staining pattern of each subpopulation, combined with FSC and SSC data, can be used to identify which cells are present in a sample and to count their relative percentages. The cells can also be sorted if desired.

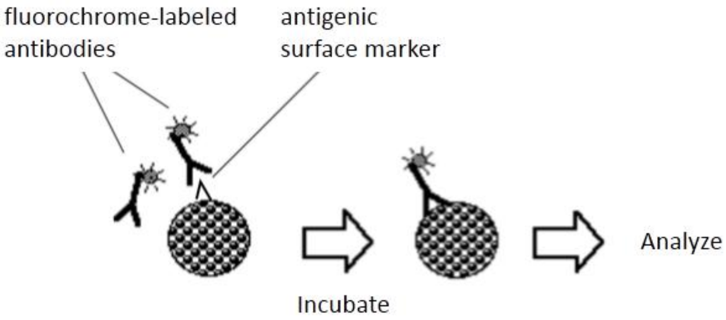


Figure 60. *Specific binding of fluorochrome-labeled antibodies to cell surface antigens.*

The scattered and fluorescent light is collected by appropriately positioned lenses. A combination of beam splitters and filters steers the scattered and fluorescent light to the appropriate detectors. The detectors produce electronic signals proportional to the optical signals striking them. List mode data are collected on each particle or event. A voltage pulse is created when a particle enters the laser beam and starts to scatter light or fluoresce. Once the light signals, or photons, strike one side of the photomultiplier tube or the photodiode, they are converted into a proportional number of electrons that are multiplied, creating a greater electrical current. The electrical current travels to the amplifier and is converted to a voltage pulse. The highest point of the pulse occurs when the particle is in the center of the beam and the maximum amount of scatter or fluorescence is achieved. As the particle leaves the beam, the pulse comes back down to the baseline (**Figure 61**). The characteristics or parameters of each event are based on its light scattering and fluorescent properties. The data are collected and stored in the computer. This data can be analyzed to provide information about subpopulations within the sample.

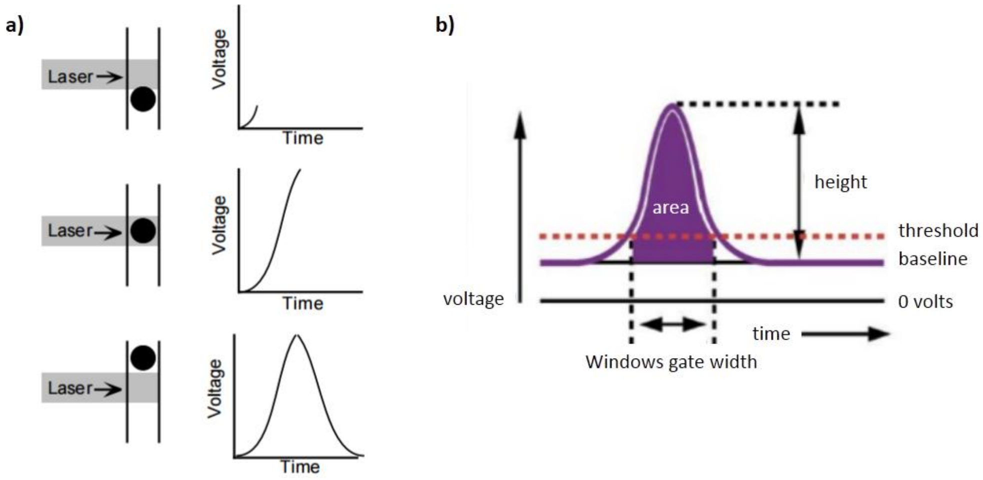


***Figure 61**. a) Creation of a voltage pulse in flow cytometry. b) Parameters measured in a voltage pulse: area (A), weight (W) and height (H).*

Once light signals have been converted to electronic pulses and then converted to channel numbers by the ADC, the data must be stored by the computer system. Once a data file has been saved, cell populations can be displayed in several different formats. Two parameters can be displayed simultaneously in a plot. One parameter is displayed on the x-axis and the other parameter is displayed on the y-axis. Plotting measured parameters one can gate different subpopulations based on size, granularity, etc.

In most applications, after a particle exits the laser beam, it is sent to waste. Sorting allows us to capture and collect cells of interest for further analysis. Once collected, the cells can be analyzed microscopically, biochemically, or functionally. To sort particles or cells, the cytometer first needs to identify the cells of interest, then separate out the individual cells. Once the population of interest has been identified on a data acquisition plot, a region is drawn around that population. A logical gate is created from the regions. This gate is then loaded into the cytometer's software as the sort gate. The sort gate identifies cells of interest to be sorted out of the stream.

Recently, De Picciotto et al. used an Sso7d-based murine serum albumin (MSA) binder (M11.1.3) obtained by YSD, and labeled it at seven sites of the engineered planar binding surface, as well as three sites in the loops connecting the strands, the C-terminus, and one near the N- terminus.[260] When labeled with the NBD fluorophore, they observed that the maximum emission wavelength for unbound biosensors was significantly longer than those obtained with other tested scaffolds (Off7 or Fn3), directly demonstrating reduced background activation with the Sso7d scaffold. They found that NBD-labeled M11.1.3(I23C), M11.1.3(W25C) and M11.1.3(G26C) showed greater than 50-fold increases in fluorescence upon addition of 10 µM MSA (**Figure 62**).

The solvatochromic fluorophore 4-N,N-dimethylamino-1,8-naphthalimide (4-DMN), developed in Prof. Imperiali's lab, possesses extremely sensitive emission properties due largely to the low intrinsic fluorescence that it exhibits in polar protic solvents such as water.[145] This greatly reduces the background signal thereby creating the effect of on-off or switch-like changes in the observed emission intensity with the potential to exceed ratios of 1000-fold. Furthermore, as it was previously reported, it has been determined that this fluorophore possesses significantly greater chemical stability than the other dimethylaminophthalimide dyes investigated.[145] They showed that the dimethylaminophthalimide series exhibits very little fluorescence in aqueous environments, which is very interesting as even a modest shift in the emission wavelength will typically be accompanied by a significant change in fluorescence intensity.

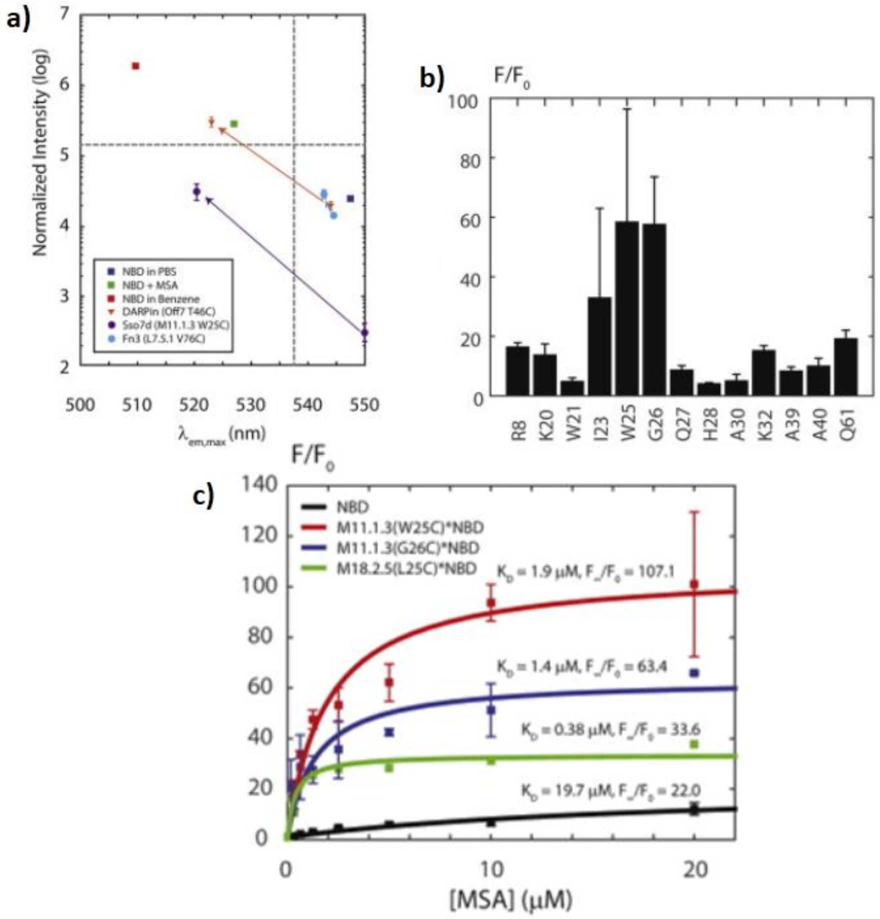


***Figure 62**. a) Maximum emission wavelength vs intensity for the best Fn3, DARPin, and Sso7d sensors. b) F/F$_0$ ratio for M11.1.3*NBD conjugates. c) Titrations of NBD and NBD labeled constructs with MSA. Extracted from Journal of Molecular Biology.*[260]

In a more recent work, also conducted in Prof. Imperiali's lab, they evaluated the performance of 4-DMN cysteine modifying agents with different linker lengths. These agents showed very promising results when compared with other five well-established commercially available solvatochromic fluorophores, making them a highly sensitive toolset for the detection of biomolecular interactions.[268]

In this chapter we will discuss the work that I performed during my stay in Prof. Imperiali's lab (Massachusetts Institute of Technology) in collaboration with Dr. Silvano Sanchini and Dr. Cristina Zamora. This project was based on the development of new biosensors based on the Sso7d scaffold. Our first aim was to re-express, purify and label with 4-DMN derivatives the best M11.1.3 mutants (M11.1.3(I23C), M11.1.3(W25C) and M11.1.3(G26C)) obtained previously by Dr. Seymour De Picciotto. The second objective was to assess their increase in fluorescence upon MSA binding and to compare them with the results obtained with the NBD-labeled M11.1.3 variants. Once this was accomplished, in collaboration with Prof. Wittrup's lab, we developed new Sso7d-based binders for a relevant target by yeast surface display (YSD). The selected target was the epidermal growth factor (EGF), a small protein (∼6 kDa, 53 amino acids) that binds to the extracellular domain of its receptor (EGFR) and activates the cytoplasmic catalytic function by promoting receptor dimerization and self-phosphorylation on tyrosine residues. Dysregulation of EGF levels has been linked to several diseases including but not limited to ovarian cancer and diabetes. We were able to obtain different Sso7d-based hEGF binders with µM affinities to start the development of new fluorescence-based sensors.

Expression and purification of M11.1.3 cysteine mutants

The Sso7d-based MSA binder (M11.1.3), identified by Dr. De Piccioto, has the following protein sequence (in bold are indicated the positions mutated in this work):

MGHHHHHHGGATVKSTYRGEEKQVDISKIKWV**I**R**WG**QHLAFKYDEGGGAAGYGWVSEKDAPKELLQ MLEKQ

Our first aim was to re-express and purify the M11.1.3(I23C), M11.1.3(W25C) and M11.1.3(G26C) mutants (**Figure 63**) and to label them with 4-DMN derivatives with different linker lengths.

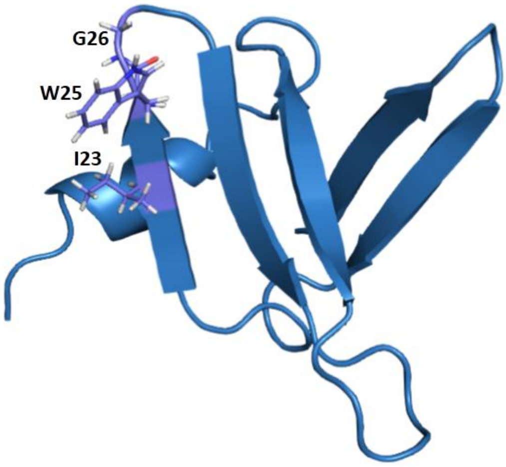


Figure 63. *Cartoon representation of the M11.1.3. Residues mutated to cysteine are labeled and represented as sticks.*

Sso7d-based mutants (M11.1.3(I23C), M11.1.3(W25C) and M11.1.3(G26C)) were expressed in *E. coli* from glycerol stocks and purified using Ni-NTA resin. Protein purity was assessed by electrophoresis as shown in **Figure 64**.

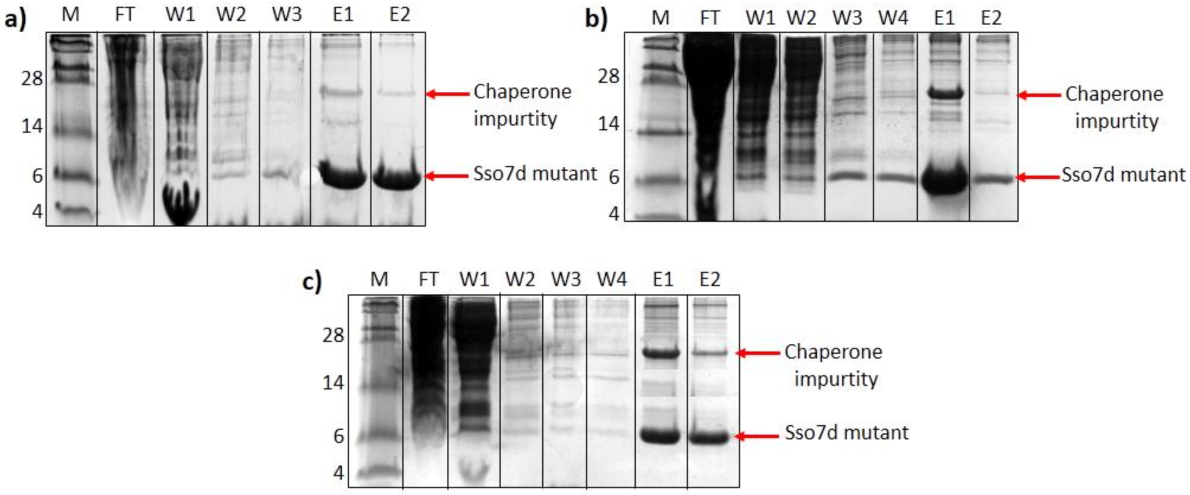


Figure 64*. 15% tricine gels for the IMAC purification of Sso7d-based mutants: a) M11.1.3(I23C), b) M11.1.3(W25C), c) M11.1.3(G26C). M: marker (See-blue Plus2 prestained), FT: flow-through W: washes, E: elution.*

As it can be seen in **Figure 64**, all the expressed Sso7d-based cysteine mutants co-eluted with a chaperone impurity. To remove it we first tried to selectively precipitate the chaperone by salting out with ammonium sulfate. Salting out is often used for protein fractionation to separate proteins that precipitate at different salt concentrations, making it a good initial purification step for small soluble proteins. Often, proteins that contain positively and negatively charged regions self-aggregate under very low salt conditions. When salt is present, however, the salt ions neutralize charges on the protein surface, preventing aggregation. At high salt concentrations, in contrast, proteins are less soluble and precipitate. This effect (salting out) is a result of the competition between the added salt ions and the protein molecules for the solvent (water) molecules.[269] Ammonium sulfate, $(NH_4)_2SO_4$, is often used for salting out because of its high solubility, which allows for solutions of very high ionic strength, low price, and availability of pure material. Additionally, NH_4^+ and SO_4^{2-} are within the most stabilizing ions in the Hofmeister series. We also decided to digest and sequence the chaperone impurity, obtaining the following sequence:

MKVAKDLVVSLAYQVRTEDGVLVDESPVSAPLDYLHGHGSLISGLETALEGHEVGDKFDVAVGANDAYG
QYDENLVQRVPKDVFMGVDELQVGMRFLAETDQGPVPVEITAVEDDHVVVDGNHMLAGQNLKFNVE
VVAIREATEEELA**H**G**H**V**H**GA**H**D**HHH**D**H**D**H**DGCCGG**H**G**H**D**H**G**H**E**H**GGEGCCGGKGNGGCGC**H**

We noticed that the chaperone had a histidine rich segment (highlighted in bold), which explains why this protein was binding the Ni-NTA resin and co-eluting with the cysteine mutants at high imidazole concentrations. We also observed that the chaperone had an isoelectric point (pI = 5.04) which significantly differs from the one calculated for the Sso7d cysteine mutants (pI = 8.22). For this reason, we decided to also try cation exchange chromatography working with 50 mM HEPES buffer at pH 7, so the Sso7d-based mutants and the chaperone will be positively and negatively charged, respectively. We also added 6 M urea to the buffers in order to avoid unspecific binding between the Sso7d-based mutants and the chaperone. To perform cation exchange chromatography we needed to dialyzed the samples against 2 L of 50 mM HEPES, 5 mM NaCl, 6 M urea pH 7.

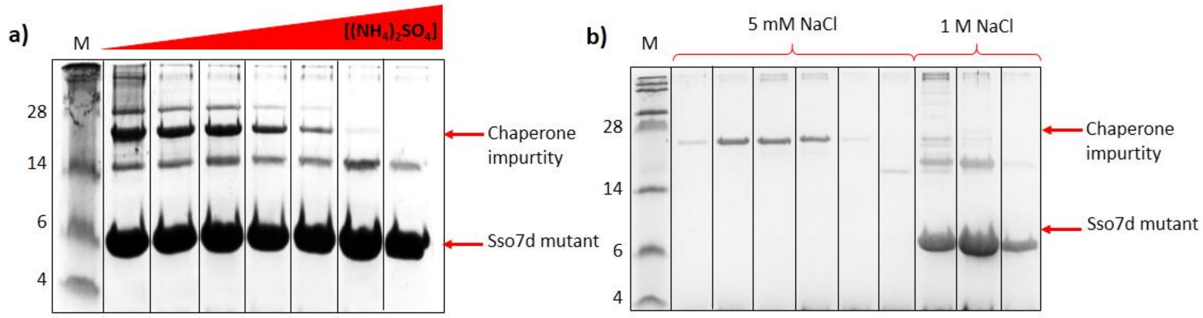


Figure 65. *15% tricine gel showing the different fractions obtained during protein purification: a) by salting out using ammonium sulfate; b) by cation exchange using sodium chloride. M: marker (See-blue Plus2 prestained).*

As it can be observed in **Figure 65a** we were finally able to obtain fractions containing pure protein without the chaperone impurity by salting out precipitation. Nevertheless, with this methodology we were actually losing some protein that seemed to precipitate at low salt concentration. On the contrary, cation exchange chromatography gives very promising results, as we were able to perfectly separate both proteins without losing material (**Figure 65b**). Fractions containing each Sso7d-based cysteine mutant were pooled together and dialyzed against 2 L of 50 mM Na_2HPO_4, 300 mM NaCl, 0.5 M urea, pH 8.0. Protein concentrations were determined by UV absorbance at 280 nm (ε = 20910 M^{-1} cm^{-1}).

Labeling of Sso7d cysteine mutants with 4-DMN derivatives

Next step was the labeling of the three Sso7d-based cysteine mutants with 4-DMN derivatives. We used five thiol-reactive derivatives of 4-DMN with different linker lengths (**Figure 66**).

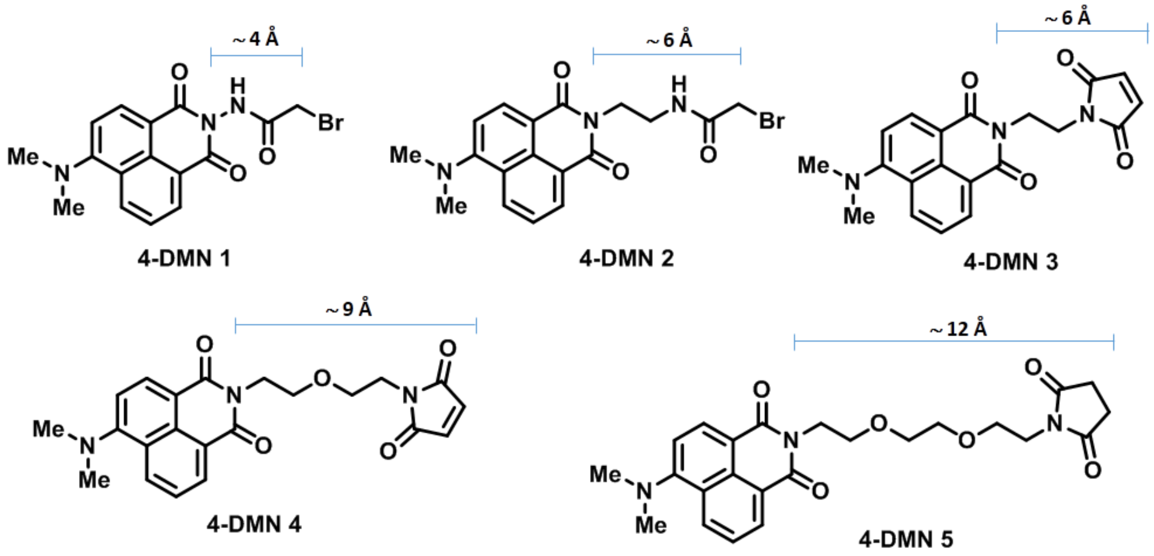


Figure 66. *4-DMN based cysteine modifying agents used for labeling Sso7d mutants*

Protein labeling was done with samples containing 50 to 100 µM Sso7d cysteine mutant in 50 mM Na_2HPO_4, 300 mM NaCl, 0.5 M urea, pH 8.0 buffer, adding 100 mM of TCEP for thiol reduction and 10 eq. of 4-DMN based cysteine modifying agent. Labeling reactions were performed overnight at room temperature with gentle tumbling and protected from light. After overnight tumbling at room temperature, the reactions were all purified via size exclusion chromatography with BioRad Bio-Gel P2 stationary phase. Fractions from the column were analyzed by tricine (15%) gel both by UV (360 nm) and Coomassie stain. Results are shown in **Figure 67**, **Figure 68** and **Figure 69**.

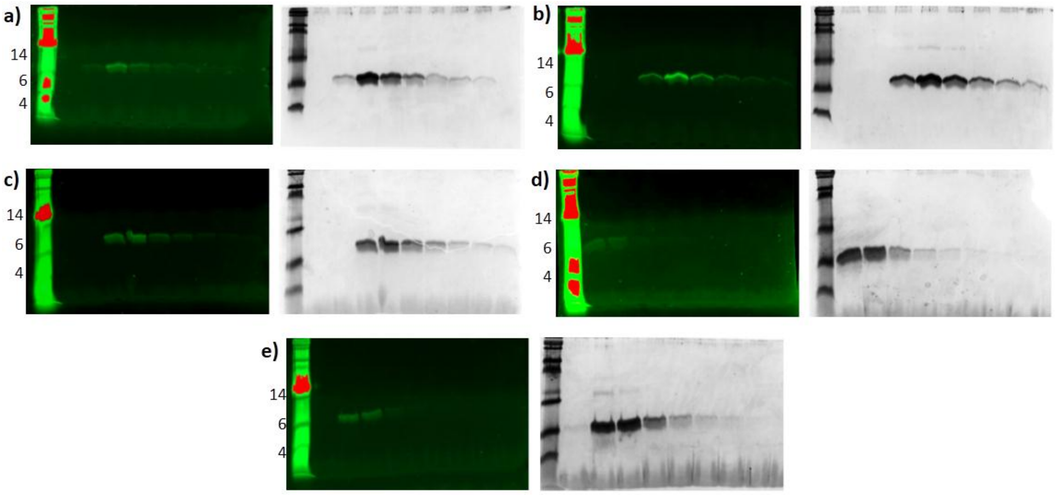


Figure 67. *Tricine (15%) gels showing the results for Sso7d mutant I23C labeled with: a) 4-DMN 1, b) 4-DMN 2, c) 4-DMN 3, d) 4-DMN 4 and e) 4-DMN 5. Proteins were loaded on the gel after excess dye removal and visualized by UV (360 nm) (left gel in each panel) and Coomassie blue staining (right gel in each panel).*

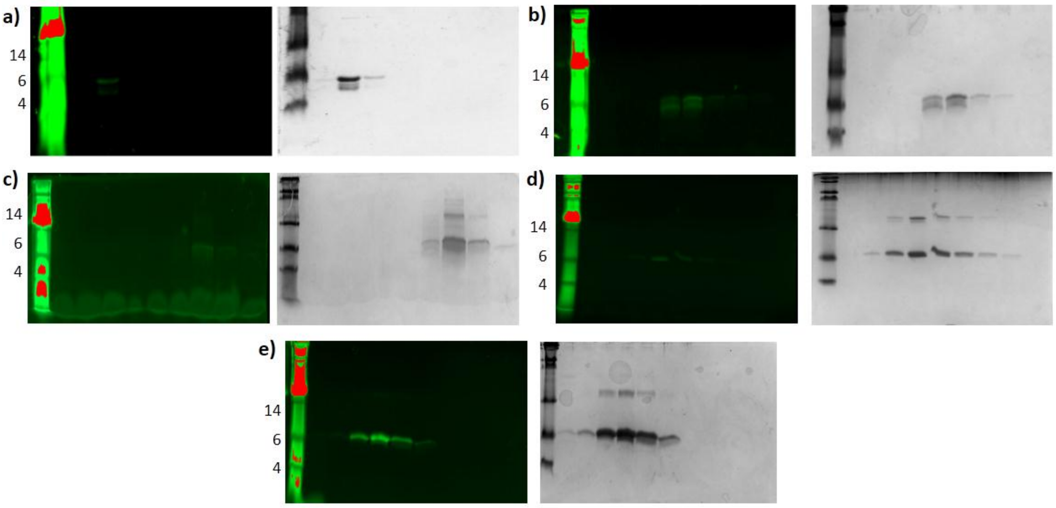


Figure 68*. Tricine (15%) gels showing the results for Sso7d mutant W25C labeled with: a) 4-DMN 1, b) 4-DMN 2, c) 4-DMN 3, d) 4-DMN 4 and e) 4-DMN 5. Proteins were loaded on the gel after excess dye removal and visualized by UV (360 nm) (left gel in each panel) and Coomassie blue staining (right gel in each panel).*

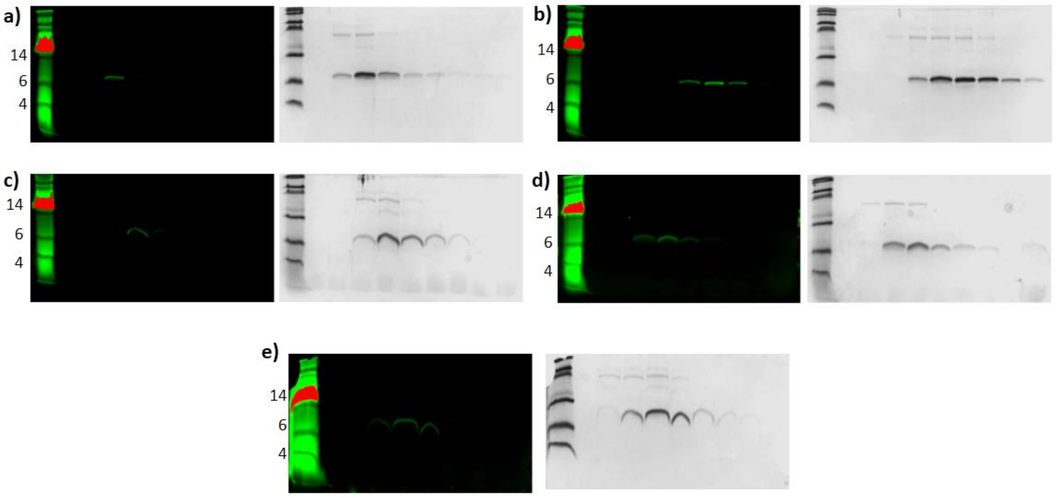


Figure 69. *Tricine (15%) gels showing the results for Sso7d mutant G26C labeled with: a) 4-DMN 1, b) 4-DMN 2, c) 4-DMN 3, d) 4-DMN 4 and e) 4-DMN 5. Proteins were loaded on the gel after excess dye removal and visualized by UV (360 nm) (left gel in each panel) and Coomassie blue staining (right gel in each panel).*

When we did fluorescence assays with the labeled Sso7d cysteine mutants we were not able to obtain reproducible results. We hypothesized that the lack of reproducibility may be caused by the presence of free dye in the samples, due to the incomplete removal of the excess of non-reacted dye by size exclusion chromatography. We decided to try alternative methodologies for the removal of non-reacted dye after the labeling reactions. First we tried desalting columns (Sephadex G25 superfine) using 50 mM HEPES, 150 mM NaCl buffer at pH 7.4.

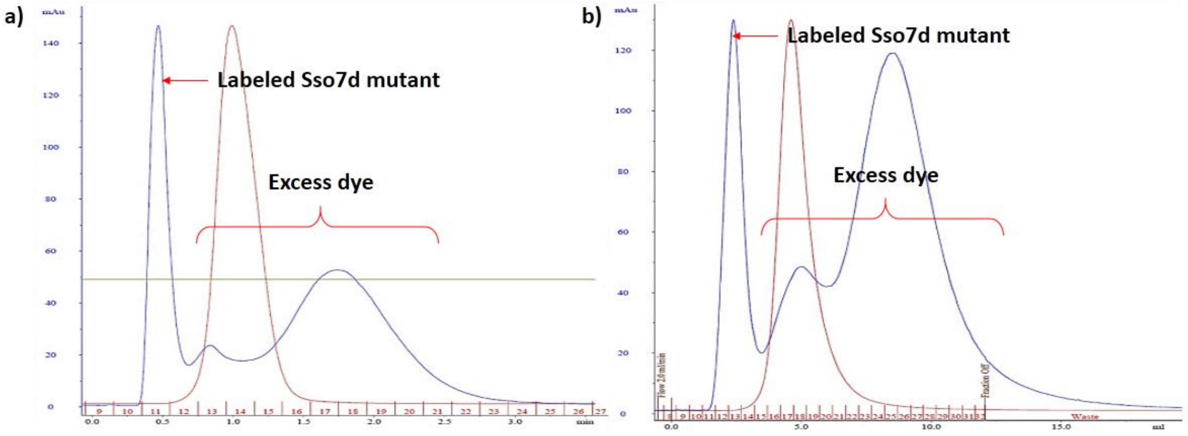


Figure 70. *Desalting columns purification after protein labeling, working at: 5 mL/min (a) and 2 mL/min (b) flow rate.*

As it can be noticed in **Figure 70**, complete separation of excess dye from the protein of interest was not achieved by desalting columns. We decided to try a cation exchange purification step (HiTrap SP FF) using 50 mM HEPES, 5 mM NaCl buffer at pH 7.0. Working at these conditions, the Sso7d cysteine mutant should be positively charged, so it will bind to the column and the dye will go through it. The protein can be eluted at high salt concentration and then dialyzed to remove the salt excess.

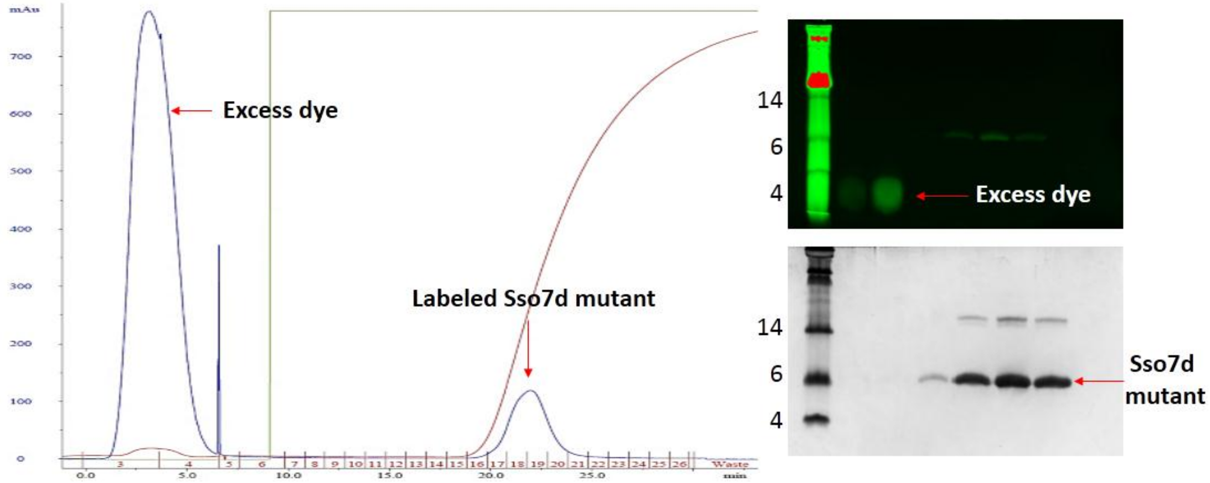
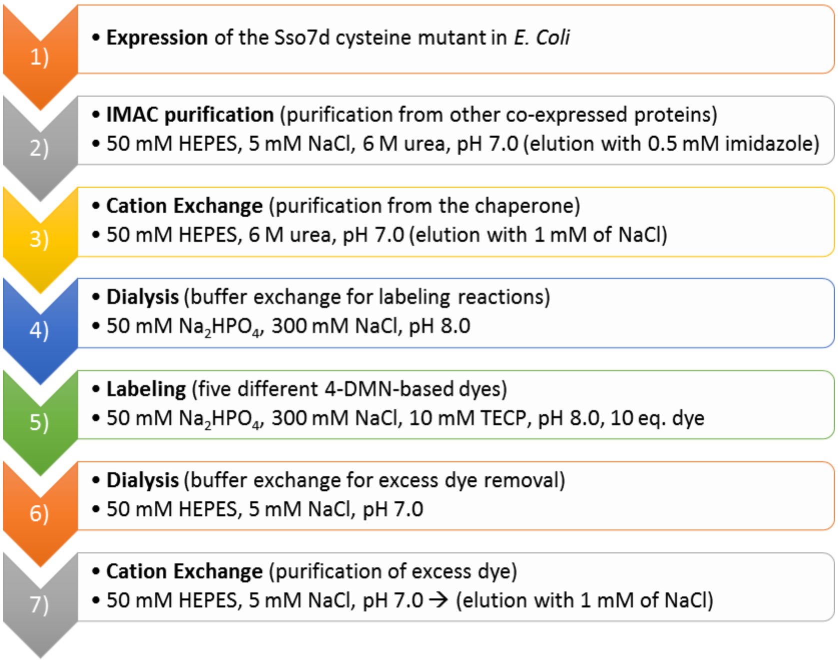


Figure 71. *Cation exchange crhromatogram (left) after protein labeling, and electrophoretic analysis of the fractions obtained (right). The labeling mixture was purified at 0.5 mL/min flow rate, loaded on tricine (15%) gel and analyzed by UV (360 nm) and Coomassie blue staining.*

Cation exchange chromatography seemed to give the best results as we were able to perfectly separate the excess dye from the protein of interest (**Figure 71**). These results were also confirmed by gel electrophoresis where we were able to observe the excess dye perfectly separated from the labeled Sso7d protein. At that point we ended up with a complete protocol for the expression, purification and labeling of the Sso7d mutants that could be applied to different clones of this protein scaffold (**Figure 72**).



1)
- **Expression** of the Sso7d cysteine mutant in *E. Coli*

2)
- **IMAC purification** (purification from other co-expressed proteins)
- 50 mM HEPES, 5 mM NaCl, 6 M urea, pH 7.0 (elution with 0.5 mM imidazole)

3)
- **Cation Exchange** (purification from the chaperone)
- 50 mM HEPES, 6 M urea, pH 7.0 (elution with 1 mM of NaCl)

4)
- **Dialysis** (buffer exchange for labeling reactions)
- 50 mM Na_2HPO_4, 300 mM NaCl, pH 8.0

5)
- **Labeling** (five different 4-DMN-based dyes)
- 50 mM Na_2HPO_4, 300 mM NaCl, 10 mM TECP, pH 8.0, 10 eq. dye

6)
- **Dialysis** (buffer exchange for excess dye removal)
- 50 mM HEPES, 5 mM NaCl, pH 7.0

7)
- **Cation Exchange** (purification of excess dye)
- 50 mM HEPES, 5 mM NaCl, pH 7.0 → (elution with 1 mM of NaCl)

Figure 72. *Schematic representation of the optimized protocol for the expression, purification and labeling of Sso7d cysteine mutants.*

Florescence assays with 4-DMN labeled Sso7d-based cysteine mutants

Spectroscopic assays were performed by measuring the fluorescence intensity increase achieved by the labeled Sso7d mutants upon MSA binding. We used the same concentrations previously reported by Dr. Seymour De Picciotto (300 nM of labeled Sso7d protein and 10 µM of MSA). Fluorescence intensity increase was expressed as intensity fold and it was calculated by dividing F/F0 where F is the maximum fluorescence intensity upon MSA binding and F0 is the fluorescence intensity of the free labeled protein measured at the wavelength corresponding to F.

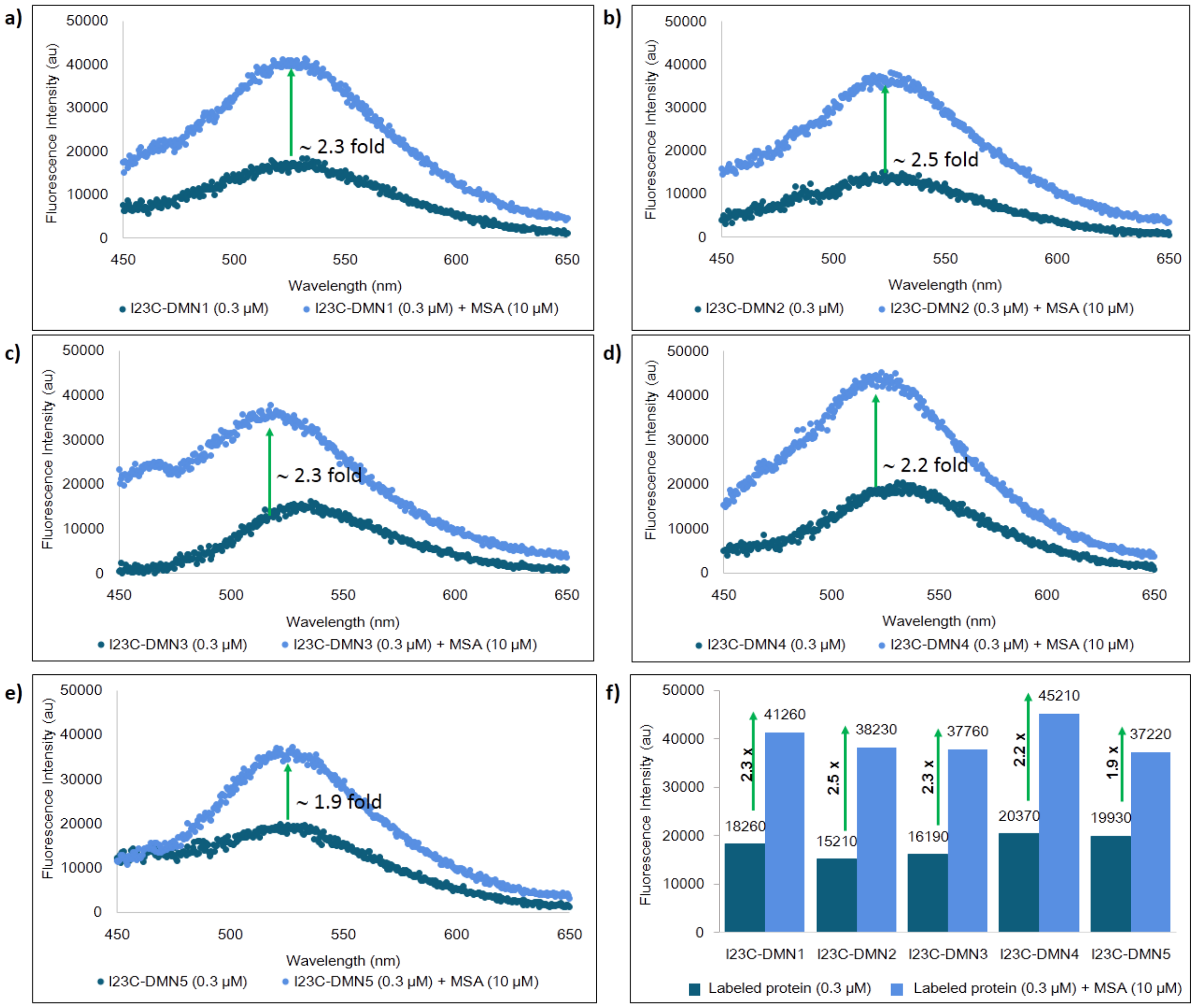


Figure 73. *Graphs for the fluorescence assays performed with 300 nM labeled protein adding 10 µM MSA. I23C mutant labeled with: a) 4-DMN 1, b) 4-DMN 2, c) 4-DMN 3, d) 4-DMN 4, f) 4-DMN 5. f) Plot comparing the response of the five dyes..*

We selected the compound with the highest increase in fluorescence intensity of this series to perform a titration assay. A sample containing 300 nM I23C-DMN2 mutant was titrated with increasing concentrations of MSA (0.1, 0.3, 1.0, 3.0, 10.0 µM).

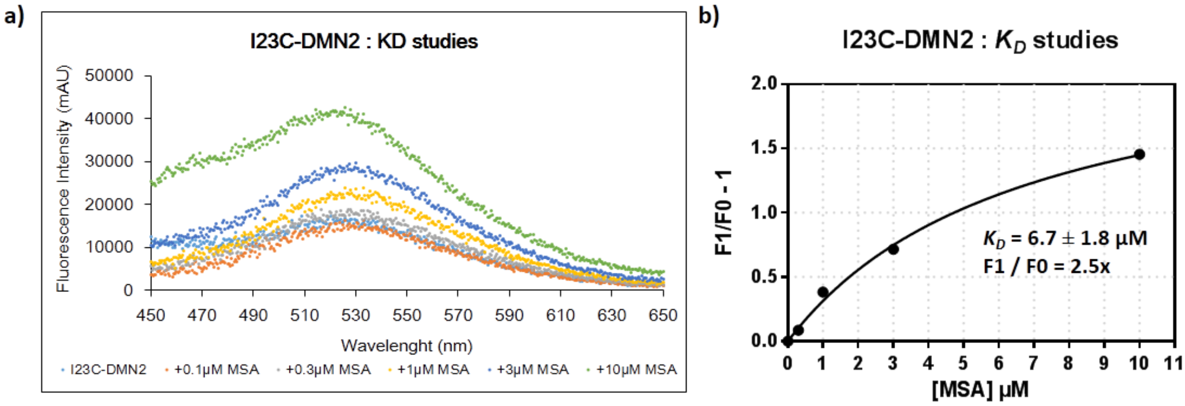


Figure 74. *a) Titration assay of the I23C-DMN2 protein with increasing amounts of MSA. b) Fit of the F1/F0 ratio to the "one site-specific binding" equation of GraphPad Prism software. The obtained K_D value is shown.*

After fitting the data using the "one site-specific binding" equation in the GraphPad Prism software, we obtained a K_D value that was pretty similar to the ones obtained by Dr. De Picciotto for the M11.1.3 clone labeled with NBD in different positions (K_D = 1.9 and 1.4 µM for the W25C and G26C mutants respectively).

Similarly we performed an equivalent evaluation of the performance of the five DMN-based dyes for the other two mutants (W25C and G26C) (**Figure 75** and **Figure 76**).

The results obtained for the three M11.1.3 mutants (I23C, W25C and G26C) labeled with 4-DMN derivatives (**Figure 73**, **Figure 75** and **Figure 76**), were in agreement with those obtained by Dr. De Picciotto, where he found that W25C was the one giving the highest fluorescence intensity increase upon MSA binding, followed by G26C and I23C. The three compounds with which we performed titration assays showed K_D values around 1 µM (**Figure 74**, **Figure 75** and **Figure 76**), which is the value that was previously determined in Prof. Imperiali's lab for the M11.1.3 binding MSA. This indicates that labeling of the M11.1.3 clone in the three studied positions, did not affect binding between this protein and its partner (MSA).

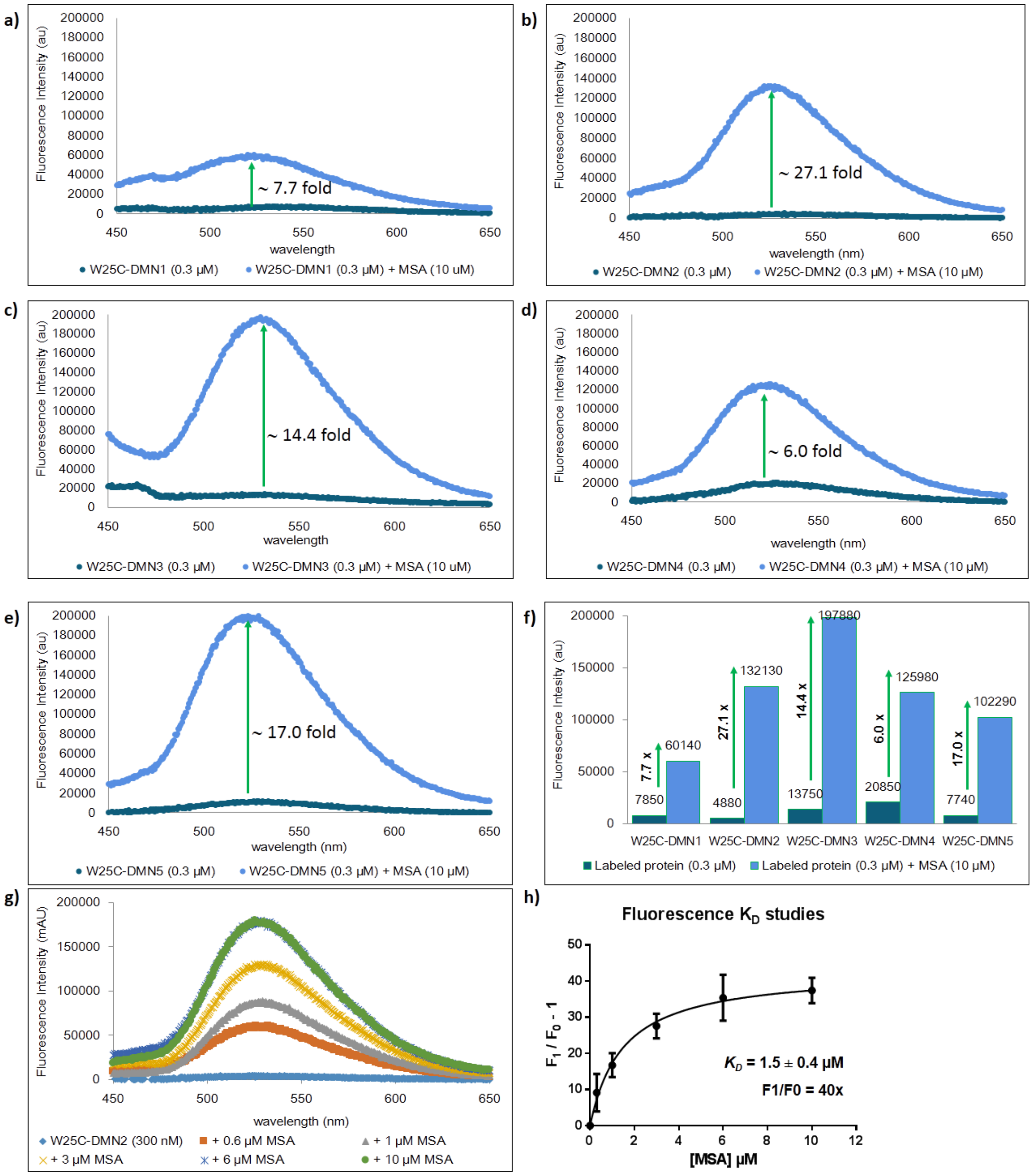


Figure 75. *Graphs for the fluorescence assays performed with 300 nM labeled protein adding 10 µM MSA. W25C mutant labeled with: a) 4-DMN 1, b) 4-DMN 2, c) 4-DMN 3, d) 4-DMN 4, f) 4-DMN 5. f) Plot comparing the results of the five dyes. g) Titration assay of the W25C-DMN2 protein with increasing amounts of MSA. h) Fit of the binding curve to the "one site-specific binding" equation of GraphPad Prism software.*

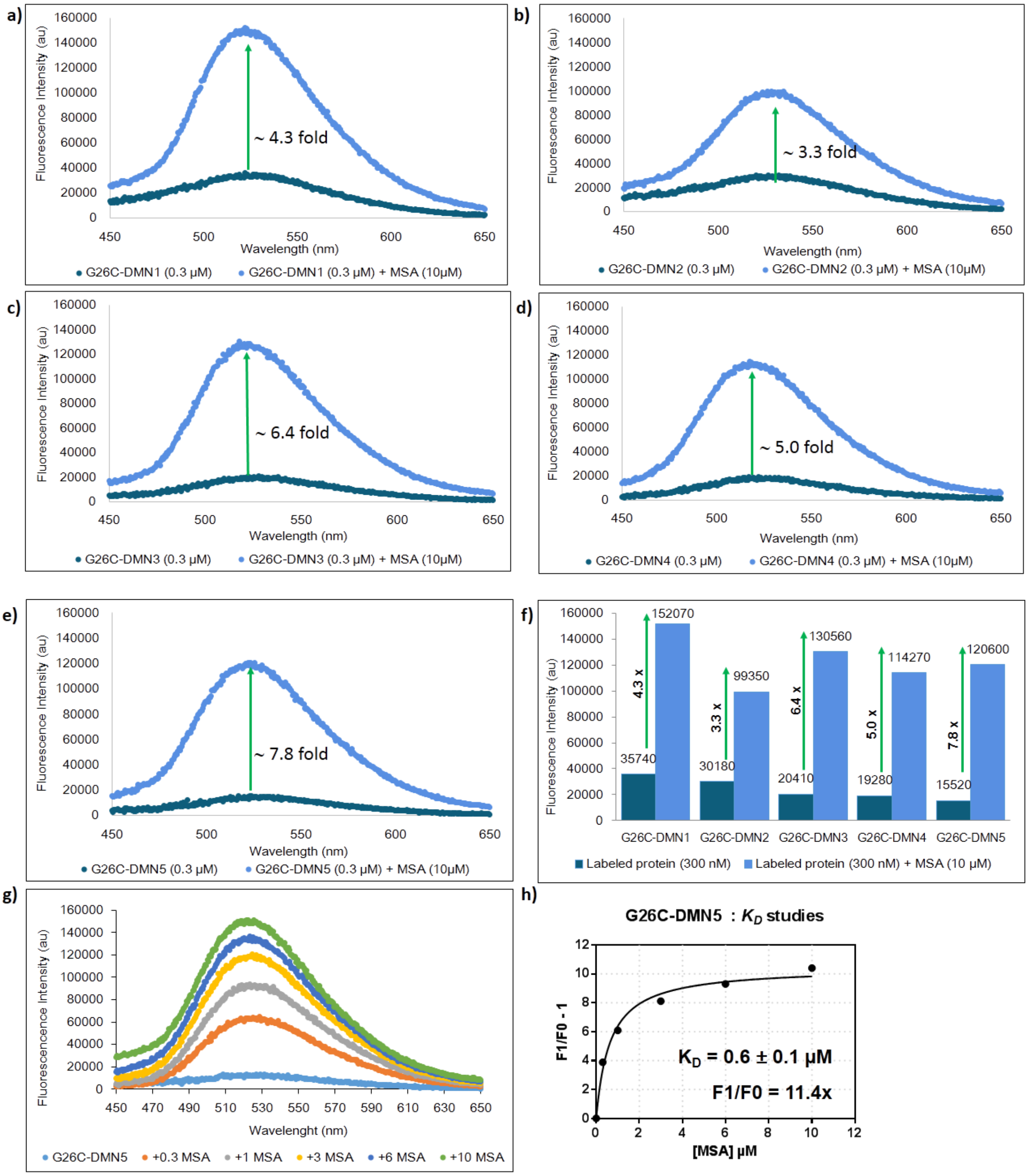


Figure 76. *Graphs for the fluorescence assays performed with 300 nM labeled protein adding 10 µM MSA. G26C mutant labeled with: a) 4-DMN 1, b) 4-DMN 2, c) 4-DMN 3, d) 4-DMN 4, f) 4-DMN 5. f) Plot comparing the results of the five dyes. g) Titration assay of the G26C-DMN5 protein with increasing amounts of MSA. h) Fit of the data to the "one site-specific binding" equation of GraphPad Prism software.*

In order to test the reproducibility of these assays we compared two independent experiments performed in two different days, by two different researchers. As it can be noticed in **Figure 77** we were able to obtain reproducible fluorescence intensity increase values working with the previously described protocol for the expression, purification and labeling of Sso7d mutants.

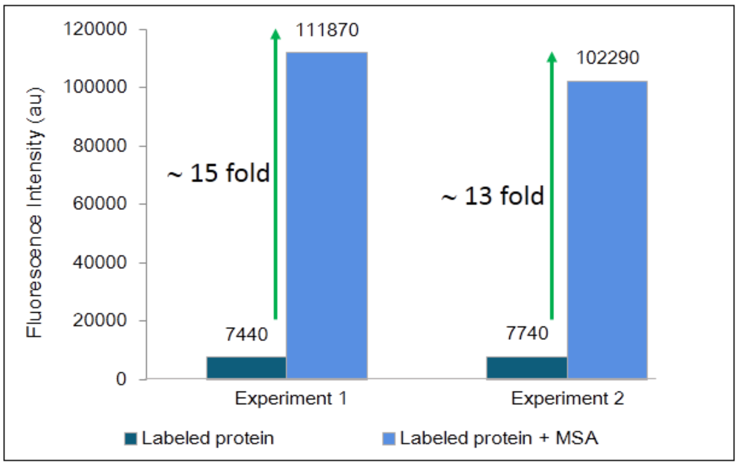


Figure 77. *Reproducibility assay performed to compare the results of two independent experiments with 300 nM W25C-DMN5 adding 10 µM of MSA.*

We can conclude that most of the studied mutants labeled with 4-DMN derivatives seemed to be shielding sensors, as they present high background fluorescence, due to interaction with the labeled protein itself; and low fluorescence intensity increase upon binding with the partner protein. Only W25C mutant labeled with 4-DMN 2, 4-DMN 3 and 4-DMN 5 seemed to present the typical properties of specific binding sensors (low fluorescence background and high fluorescence intensity increase upon binding with its partner). Interestingly W25C labeled with 4-DMN 2 seemed to present no interaction with the labeled protein itself as it can be noticed by its very low fluorescence background (the lowest from all the studied peptides). This may mean that, in this compound, the fluorescent probe may find a cavity in the MSA protein. This will explain the high fluorescence intensity increase observed for this compound.

Comparison between M.11.1.3.(W25C)-4-DMN2 and M11.1.3(W25C)-NBD variants

After seeing the results obtained with the Sso7d mutants when labeled with 4-DMN derivatives, we decided to compare our best compound (M11.1.3(W25C)-4-DMN2) with the same mutant labeled with NBD (M11.1.3(W25C)-NBD). For this purpose we used fresh labeled samples, as we observed that samples kept at -20^{o}C for more than a week were degraded. We labeled again W25C mutant with 4-DMN2 and a cysteine reactive analog of NBD known as IANBD amide (*N*,*N*'-Dimethyl-*N*-(Iodoacetyl)-*N*'-(7-Nitrobenz-2-Oxa-1,3-Diazol-4-yl)Ethylenediamine) (**Figure 79d**). Labeling results are shown in **Figure 78.**

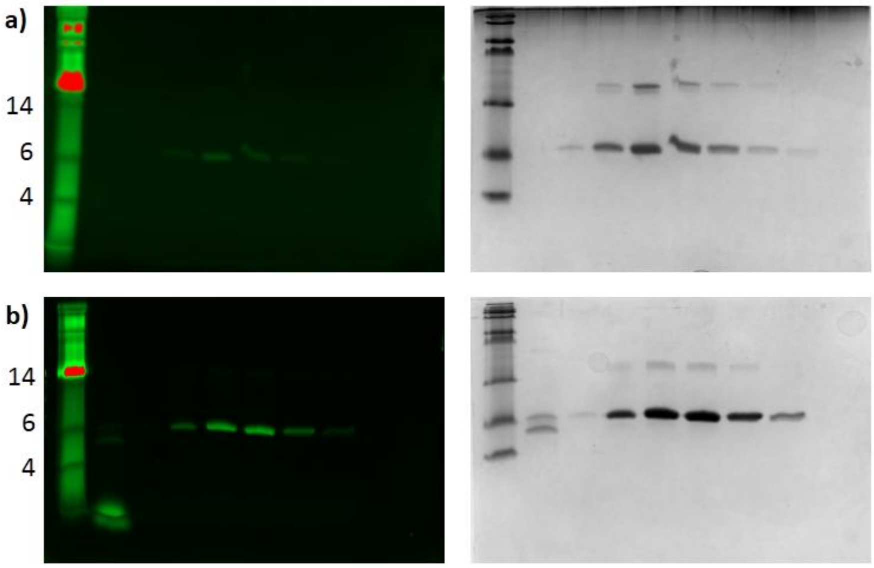


Figure 78. *Tricine (15%) gels for W25C labeled with a) IANBD and b) 4-DMN 2. Both gels were analyzed by UV (360 nm) and Coomassie staining.*

We performed fluorescence experiments with these labeled proteins at 300 nM concentration, adding 10 µM of MSA and measuring the fluorescence intensity increase upon binding as it was previously described in this chapter. M11.1.3(W25C)-4-DMN2 showed similar results to the ones obtained for M11.1.3(W25C)-NBD (F/F0 = 40 and 50 respectively), suggesting that both variants performed similar (**Figure 79**). Furthermore we noticed that M11.1.3(W25C)-4-DMN2 presented a lower fluorescence background when compared with M11.1.3(W25C)-NBD, suggesting less pre-activation due to interaction with the labeled protein itself (**Figure 79c**). Both fluorescent reagents are chemically and structurally similar (**Figure 79d**), and it was not unexpected that they showed a similar performance when attached to M11.1.3(W25C) mutant.

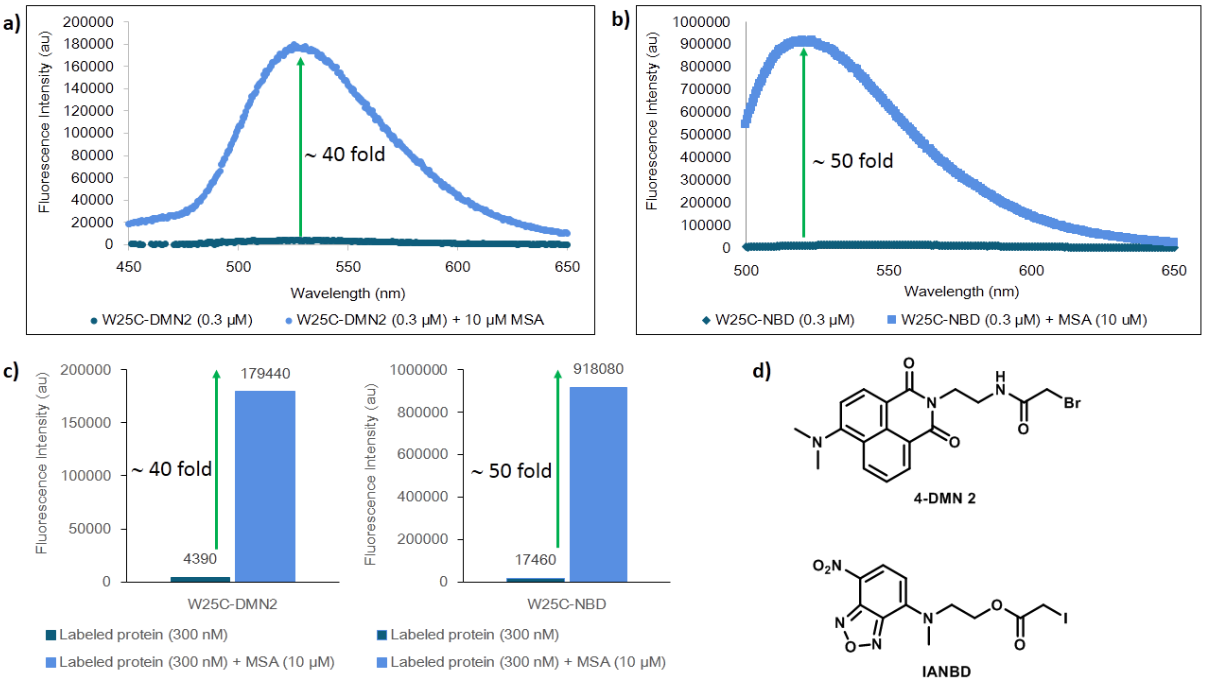


Figure 79. *Graphs for the fluorescence assays performed with 300 nM labeled protein adding 10 µM MSA. W25C mutant labeled with a) 4-DMN 2 and b) NBD. c) Comparison of the fluorescence intensity increase observed upon binding for W25C-DM2 (left) and W25C-NBD (right). d) Structure of the two cysteine reactive agents used for protein labeling.*

In a recent work, De Picciotto et al.[260] described that NBD seemed to find a pocket when binding to MSA which may explain the high increase in fluorescence intensity observed for this compound. In both variants (M11.1.3(W25C)-4-DMN2 and M11.1.3(W25C)-NBD) we could observed that the maximum emission wavelength was significantly blue-shifted upon MSA binding ($|\Delta\lambda|$ = 24 and 10 nm for M11.1.3(W25C)-NBD and M11.1.3(W25C)-4-DMN2 respectively); indicating decrease in the polarity of the environment surrounding the fluorescent dye. This may confirm the hypothesis of NBD and 4-DMN2 interaction with a hydrophobic pocket placed on MSA. The presence of a binding pocket is not an absolute requirement for useful sensor activity, but is an interesting condition as it increases the fluorescence intensity of the biosensor in a very significant way.

Engineered Sso7d-based human epidermal growth factor (hEGF) binders by yeast surface display

After having developed a robust protocol for the expression and purification of Sso7d-based solvatochromic biosensors to monitor protein-protein interactions, we were ready to apply this strategy to a relevant target. For this purpose human epidermal growth factor (hEGF) was selected. The aim of this part of the project was to obtain Sso7d-based hEGF binders using yeast surface display so that we could do selective mutagenesis to obtain cysteine mutants and convert them in solvatochromic biosensors to detect in-vivo hEGF levels.

We worked in collaboration with Prof. Wittrup's lab in order to obtain Sso7d-based hEGF binders. For this purpose we used two Sso7d libraries (library 11 and library 18) previously designed in the Wittrup lab. The libraries were generated by the randomized mutation of nine Sso7d residues (K21, W23, V25, M28, S30, T32, R42, A44, and S46). These residues are located on a planar surface of the protein projecting their side chains to the solvent (**Figure 80**). Library 11 was generated by the randomized substitution of the indicated Sso7d residues with eleven amino acids representing the most common residues found in protein-protein interactions and antibody CDR3 loops (alanine, aspartic acid, glycine, histidine, isoleucine, lysine, asparagine, arginine, serine, tryptophan and tyrosine). In the Sso7d library 18 eighteen possible amino acids (excluding cysteine and proline from the twenty proteinogenic amino acids) were randomly introduced in the nine indicated Sso7d variant positions.

Sequence#	Details	Sequence
1-19	Invariant	ATVKFKYKGEEKQVDISKI
20-26	Strand 1	KKVWRVG
27-33	Strand 2	KMISFTY
34-40	Invariant	DEGGGKT
41-47	Strand 3	GRGAVSE
48-61	Invariant	KDAPKELLQMLEKQ

Figure 80. *Left) Cartoon representation of the Sso7d scaffold. Depicted in red, magenta and blue are the 9 variable positions of the Sso7d libraries for YSD. Right) Table showing the Sso7d scaffold sequence. Depicted in red, magenta and blue are the 9 variable positions in strand 1, 2 and 3 respectively.*

We worked with the *Saccharomyces Cerevisiae* Aga2 protein system,[270] in which protein scaffolds are expressed as fusions to the Aga2 protein, that forms two disulfide bonds with the Aga1 protein, which is anchored to the cell wall via β-glucan linkage. The gene encoding the Aga1 protein is stably integrated into the yeast chromosome, while the gene encoding the protein scaffold-Aga2 fusion is cloned into a circular yeast display plasmid vector (pCT-CON) that is maintained episomally in the yeast using a nutritional marker for selective growth.

YSD requires two main components: the yeast, EBY100; and the plasmid, pCT-CON. EBY100 is deficient in the machinery to synthesize the amino acid tryptophan and contains the Aga1 gene in the yeast genome. The pCT-CON plasmid encodes for (a) the gene TRP1, which is important for tryptophan synthesis; (b) the Aga2p protein fused to the protein of interest; and (c) ampicillin resistance, for plasmid production in *E.coli.* Expression of both Aga1 and Aga2 proteins is under control of the galactose-inducible promoter. When yeast are properly transformed with the pCT-CON plasmid, they can grow in selective media deficient in tryptophan, whereas untransformed EBY100 will not propagate. Switching the yeast from glucose-rich media to galactose-rich media will induce proper display of the protein of interest (**Figure 81b**).

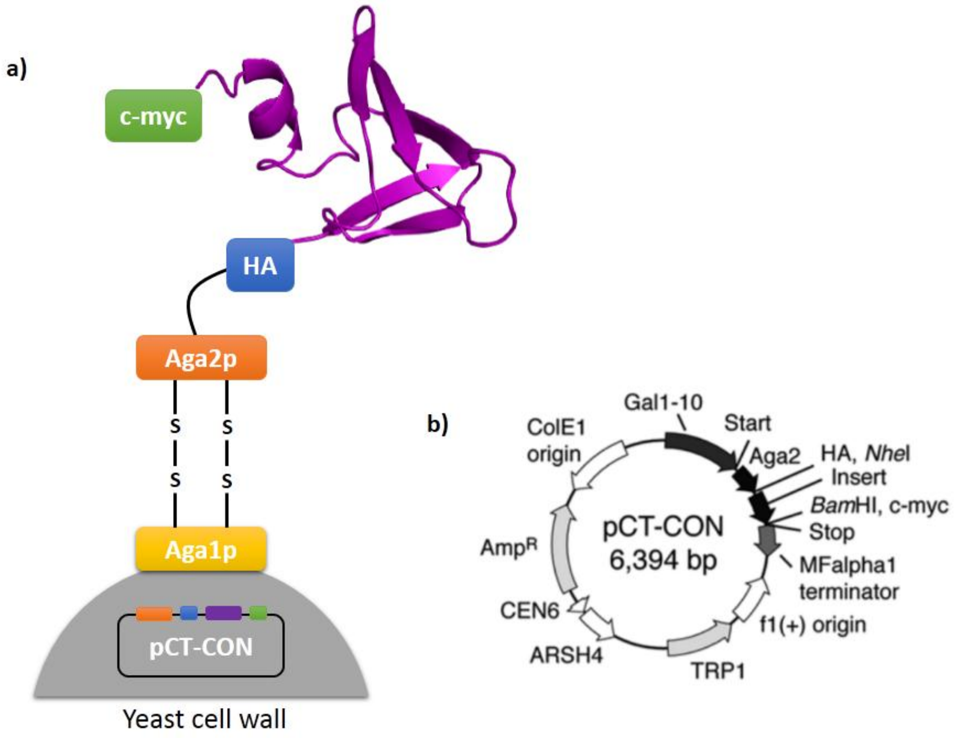


Figure 81. *Schematic representation of YSD. a) Schematic of the Sso7d scaffold displayed on the surface of yeast. b) Vector map of pCT-CON. The "insert" label indicates the position where the Sso7d gene is inserted.*

In our case, the Sso7d scaffolds were displayed on the yeast surface as C-terminal fusion to the Aga2 protein and flanked by two detection tags: hemagglutinin (HA) epitope tag (YPYDVPDYA) at the N-terminus, which is used as an expression marker to monitor plasmid incorporation; and c-Myc epitope (EQKLISEEDL) at the C-terminus, which is used to monitor full length expression of the gene on the yeast surface (**Figure 81a**). Using a two-color labeling scheme, with one fluorophore for the expression (anti-HA or c-myc tag) and another for target binder (anti-biotin), protein scaffolds can be engineered for higher affinity and stability binders.

Selection of Sso7d-based hEGF clones was performed by fluorescence activating cell sorting (FACS). However, FACS is not well suited for isolating de novo binding clones from naïve libraries due to limitations in the size of the population that can be analyzed, the minimum binding affinity of clones that can be reliably captured, the amount of target antigen required, and the likelihood of capturing artifactual binders to the reagents.[267] FACS instruments analyze and sort on the order of 10^8 cells per hour. To ensure that almost all clones are analyzed at least once, a 10-fold excess of the library diversity should be analyzed, requiring a day of sort time for a library of 10^8 cells, and setting this size as a reasonable upper limit for the diversity that can be easily analyzed with this degree of diligence. Unfortunately, a library of 10^{10} cells may be necessary to obtain binders with affinities in the nanomolar range.[271]

We needed a step before FACS in order to reduce the library diversity. In our case, initial library screening was first performed by magnetic beads selection using hEGF-coated Biotin Binder Dynabeads. The multivalency of the Dynabeads (~5×10^6 biotin-binding sites per bead) and the yeast (~10^5 Sso7d per cell) allows for the isolation of weak affinity Sso7d-based binders through avidity interaction. After reducing the diversity of the libraries by magnetic beads selection, FACS can be perfectly used to select high affinity EGF binders.

The whole process that was performed to obtain the Sso7d-based hEGF binders is depicted in **Figure 82**. Briefly, different rounds of positive and negative magnetic beads selection were used to select for weak binders and to get rid of magnetic bead binders. Yeast cells isolated after the magnetic bead-based processes were further screened using FACS to isolate clones that exhibit higher binding affinity towards the target. Finally, single clone analysis and characterization was achieved.

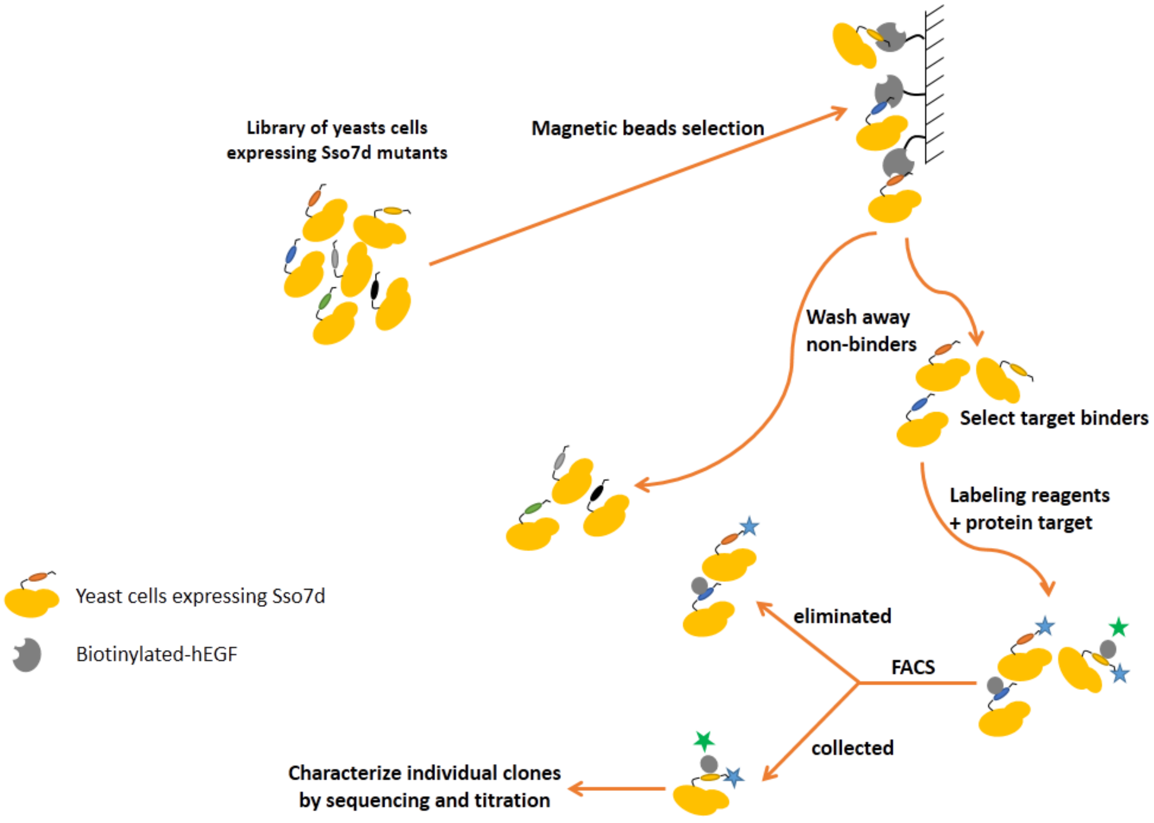


Figure 82. *Schematic representation for selection of protein binders from a naïve yeast display library using magnetic bead screening followed by flow cytometry sorting.*

hEGF selective biotinylation

Biotinylated hEGF was required for both immobilization on magnetic beads and also for future labeling purposes when performing FACS selection. We decided to obtain the biotinylated hEGF by sortase mediated ligation (**Figure 83**).

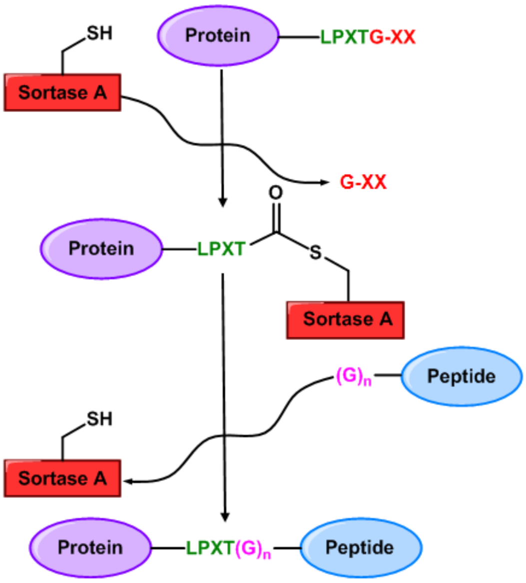


Figure 83. *Schematic representation of the sortase mediated ligation used to produce biotinylated hEGF*

For this purpose Dr. Silvano Sanchini expressed and purified hEGF-LPRTGGG and he also synthesized and purified the peptide linker containing the biotin moiety. We used a flexible serine-glycine linker to place the protein of interest far enough from the biotin moiety so that it should not interfere with protein binding (**Figure 84**).

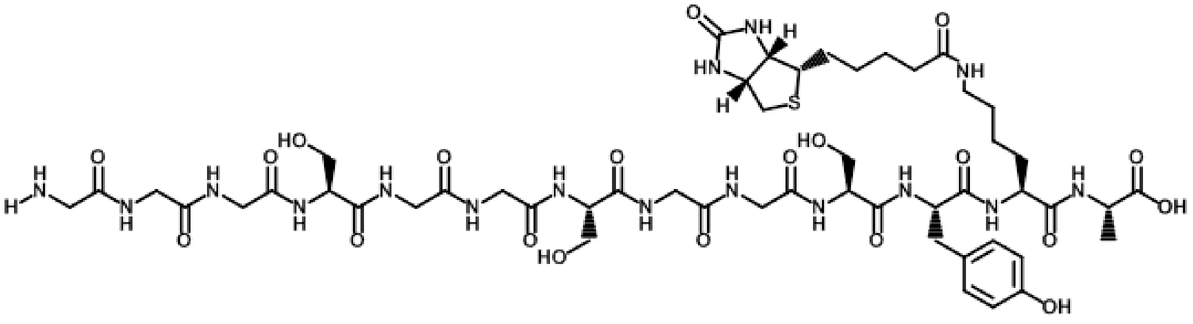


***Figure 84**. Flexible serine-glycine peptide linker used for the sortase mediated ligation.*
Peptide sequence: GGGSGGSGGSYK(biotin)A.

Sortase mediated ligation was performed in 50 mM HEPES, 150 mM NaCl, 10 mM $CaCl_2$, 1 mM TCEP, pH 7.5 using 30 µM hEGF-LPRTGGG, 5 eq. of biotinylated peptide and 0.3 eq. of Sortase A. Reaction was followed by HPLC-MS. After completion, as the SrtA was expressed with a His-tag, the final product was purified by IMAC purification. Protein biotinylation was confirmed by HPLC-MS. Size exclusion chromatography was performed to remove the excess peptide, and the final protein concentration was measured by UV absorption at 280 nM (ε = 19420 $M^{-1}cm^{-1}$).

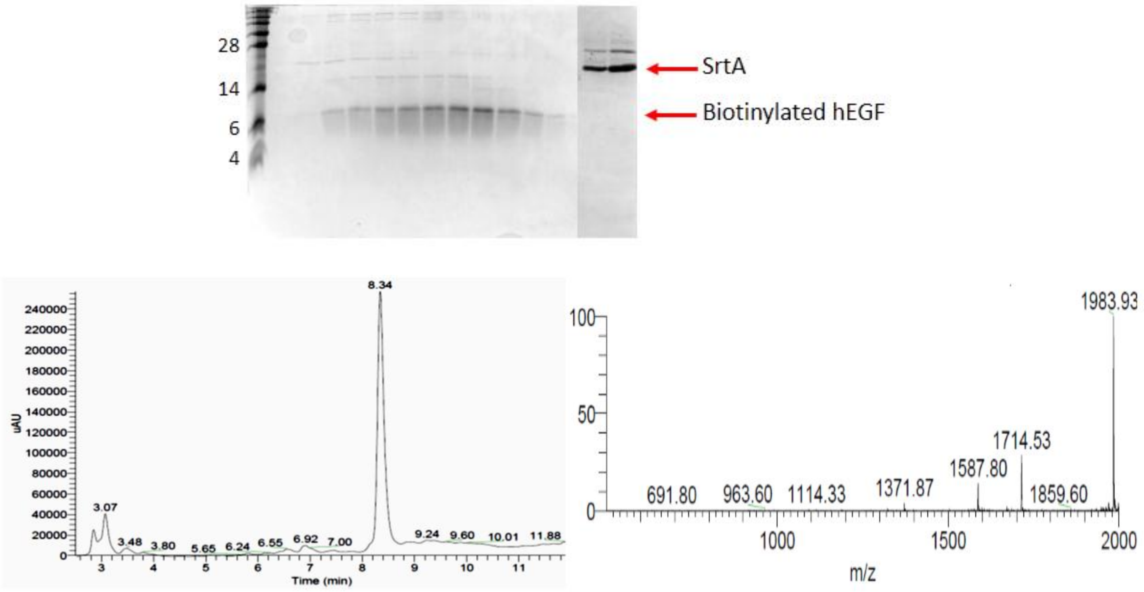


***Figure 85.** Top) Tricine (15%) gel showing the results of the IMAC purification after sortase mediated ligation. Proteins were visualized by Coomassie staining. Bottom) HPLC chromatogram and TIC obtained for the purified biotinylated protein. MW expected: 7932.5, MW calculated: 7932.8.*

Naïve library growth

We used glycerol stocks of both Sso7d libraries (Sso7d-11 and Sso7d-18) to culture the naïve libraries in SD-CAA media. Protein expression was induced by changing the cultivation dextrose-containing SD-CAA media for SG-CAA media (similar to SD-CAA media, except that glucose is used instead of dextrose).

Full length protein expression after naïve library growth was assessed by flow cytometry. Each library was as shown in **Table 21**. As we were using the same secondary antibody for both anti-HA and anti-c-myc detection, samples were labeled separately. Also an unlabeled sample, where only the secondary antibody was added, was used as a negative control. Samples were sent to the Witrup's lab for full length protein analysis.

Table 21. *Labeling strategy to assess full length protein expression*

	Sample 1 (100 µL)	**Sample 2 (100 µL)**	**Sample 3 (100 µL)**
[Cells]	2x10^6 cells/mL	2x10^6 cells/mL	2x10^6 cells/mL
Chicken anti-HA IgY	1:200 (0.5 µL)	-----	-----
Chicken anti-c-myc IgY	-----	1:200 (0.5 µL)	-----
Alexa Fluor 488-goat anti-chicken IgG	1:200 (0.5 µL)	1:200 (0.5 µL)	1:200 (0.5 µL)

In the flow cytometry plots showed in **Figure 86** each dot corresponds to a single yeast cell. The y axis indicates fluorescence intensity of Alexa Fluor 647, while x axis accounts for fluorescence intensity of Alexa Fluor 488. As in our experiments, we used the Alexa Fluor 488 goat anti-chicken to label the HA and c-myc tags, higher values along the x axis indicated higher display of these tags on the cell surface. Alexa Fluor 647 streptavidin was used to label the biotinylated hEGF, so higher values along the y axis indicated higher number of Sso7d proteins binding to hEGF. Dots that appeared in the Q4 quadrant corresponded to unlabeled cells. Those cells that stained positive in both the anti-HA and anti-c-myc plots are the ones displaying full length Sso7d protein (dots observed in the Q3 quadrant). Different intensity values on the x axis were attributed to diverse number of protein copies expressed on one yeast cell wall. In truncations or frame-shift mutants, the c-myc tag detection would be lost and, in the case of plasmid loss, the HA tag would not be detected. As it can be noticed, after naïve library growth, around 50-60% of the yeast cells expressed full length protein.

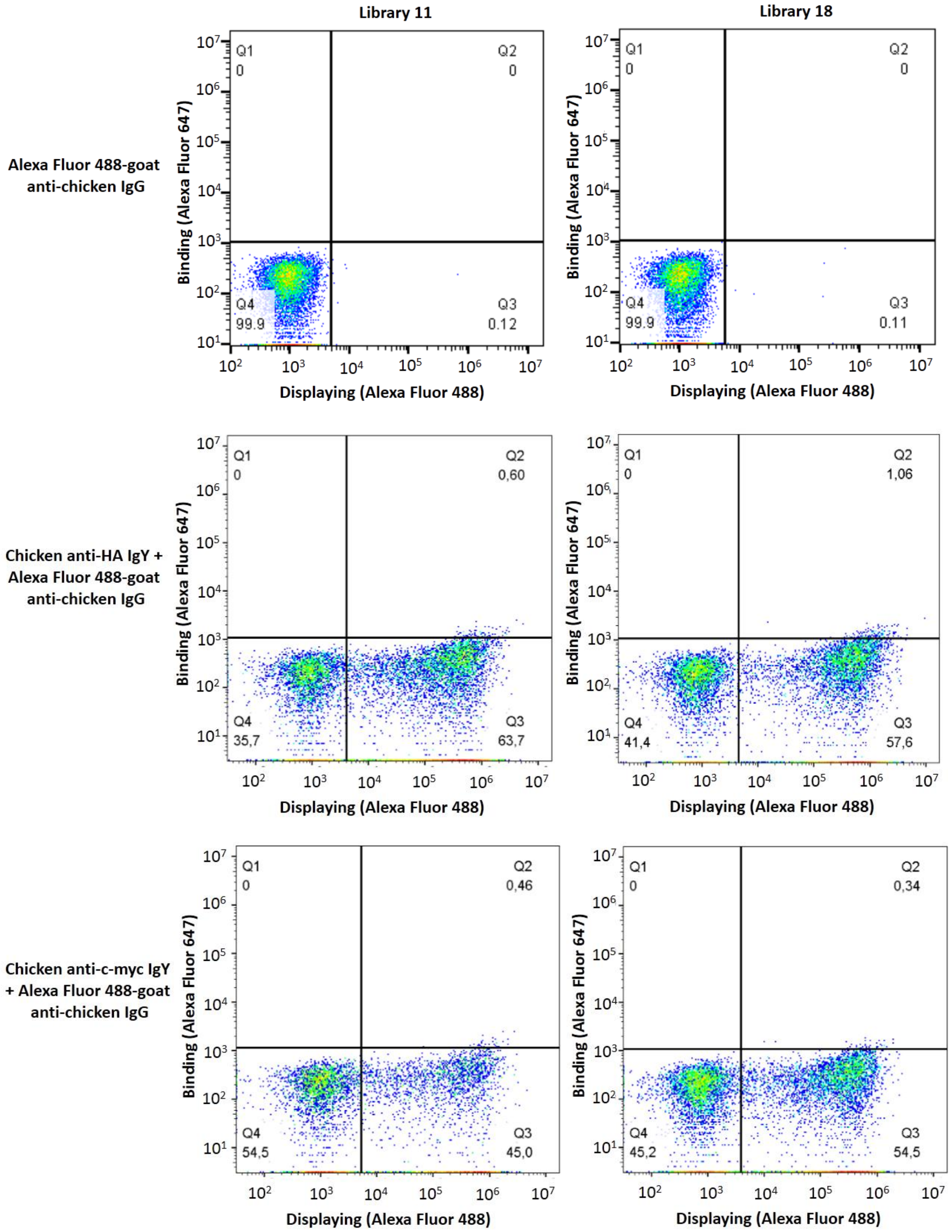


Figure 86. *Flow cytometry plots showing the results for full length protein expression.*

First round of magnetic beads selection

First round of magnetic beads selection was done with hEGF-coated beads, prepared with the biotinylated hEGF previously obtained. For this purpose, yeast cells expressing Sso7d mutants were combined with hEGF-coated magnetic beads and selection was performed. After magnetic beads selection, yeasts-beads complex was inoculated in SD-CAA media and grown again. 1 µL of each yeast media was diluted with SD-CAA media to a final volume of 200 µL, and 50 µL of this dilution were plated on SD-CAA plates and incubated at 30ºC for at least 72h.

After 72 hours colonies grown on the SD-CAA plates were counted to have an idea of the maximum diversity that we could have in both libraries. Using **Equation 3** we could estimate the number of viable cells in each culture.

$$\frac{CFUs}{mL} = \frac{number\ of\ colonies}{(dilution\ factor)\cdot(volume\ plated\ in\ mL)} \quad \textbf{Equation 3}$$

After the first round of magnetic beads selection we counted 107 and 96 colonies for library 11 and 18 respectively; meaning that we had 3.0 x 10^7 and 2.7 x 10^7 viable cells respectively. If we assume that each colony is a different clone, this should be the maximum diversity of our library.

After growing overnight, cells were pelleted and resuspended in fresh SD-CAA media. Magnetic beads were removed and remaining yeast cells were inoculated again in SD-CAA media, grown and induced.

Full length protein expression after first round of magnetic beads selection was assessed again by flow cytometry, as it was previously described in this chapter. We could observe that, after first round of magnetic beads selection, around 50-60% of yeast cells expressed full length protein on cell walls (**Figure 87**).

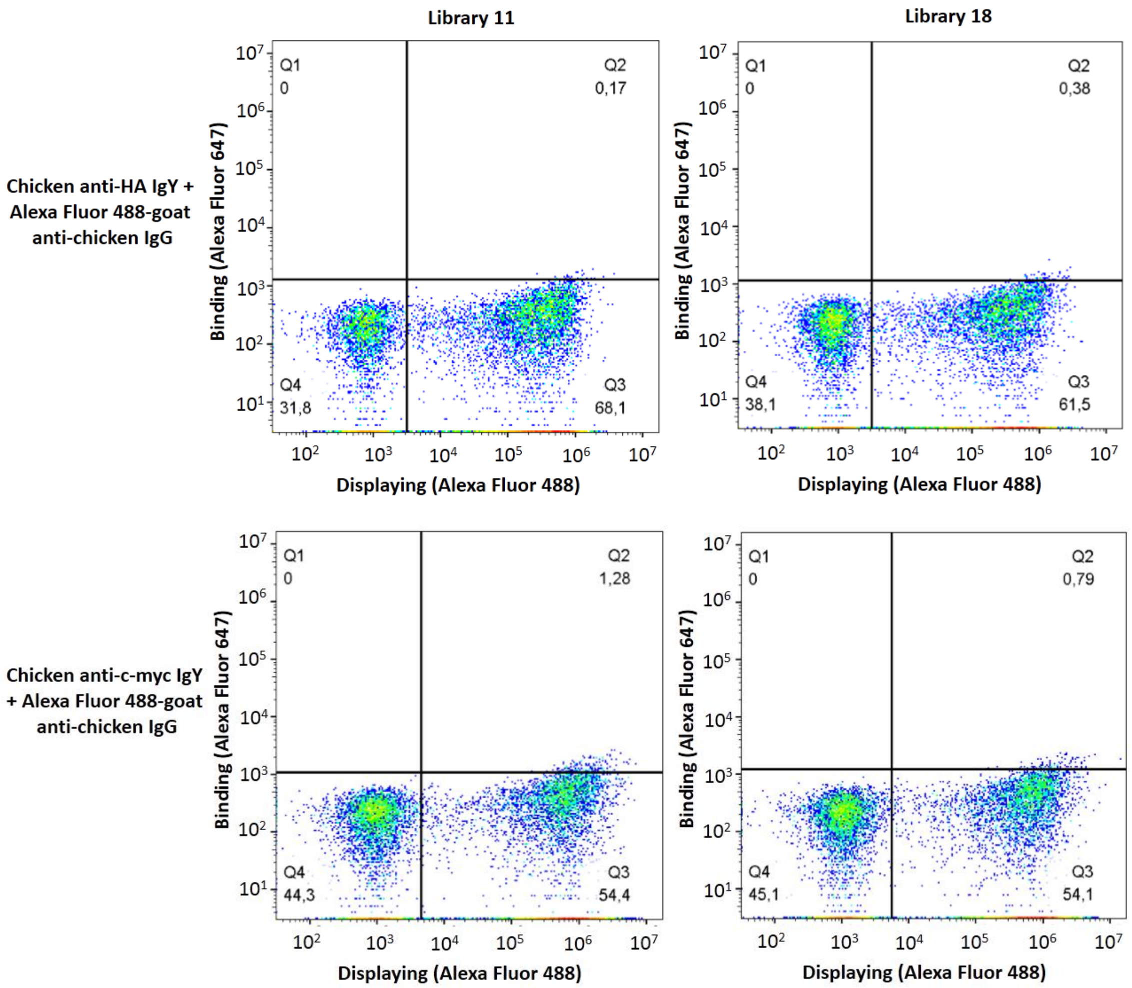


Figure 87. *Flow cytometry plots showing the results for full length protein expression*

Second round of magnetic beads selection

As the maximum diversity present in the cultures after magnetic beads selection was around 3 x 10^7 cells, we used 1x10^9 cells to be sure that we are not omitting any of the expressed clones. After subsequent washes, cells were resuspended with ice-cold PBSA to a final volume of 1 mL and combined with uncoated magnetic beads. Selection was performed to eliminate magnetic beads binders and false positives. This time we performed three rounds of negative selection and one round of positive selection using hEGF-coated magnetic beads. Once the four rounds of magnetic beads selection were completed, yeasts-beads complex was inoculated into SD-CAA media. 100 µL of each yeast culture was diluted with SD-CAA media to a final volume of 1 mL and 20 µL of this mixture were plated on SD-CAA plates and incubate at 30ºC for at least 72h. Finally, cells were grown overnight and then, magnetic beads were removed once again and yeast cells were resuspended in SD-CAA media, grown overnight, passaged and induced as it was previously done.

After 72 hours colonies grown on the SD-CAA plates were counted to have an idea of the maximum diversity that we could have in both libraries after the second round of magnetic beads selection. Using **Equation 3** we estimate the number of viable cells in each culture. This time we only counted one colony for each library which means that the dilution made for plating the yeast culture was not correct. However we estimated 1.5 x 10^4 viable cells in both libraries. As the diversity was significantly reduced using magnetic beads sorting, we were then able to use FACS for the selection of higher affinity hEGF binders.

Previous to FACS sorting we needed to assess once again that yeast cells were expressing full length protein after the second round of magnetic beads selection. So, once induction of protein expression was done, cells were pelleted and resuspended in ice-cold PBSA and labeling was done as previously described.

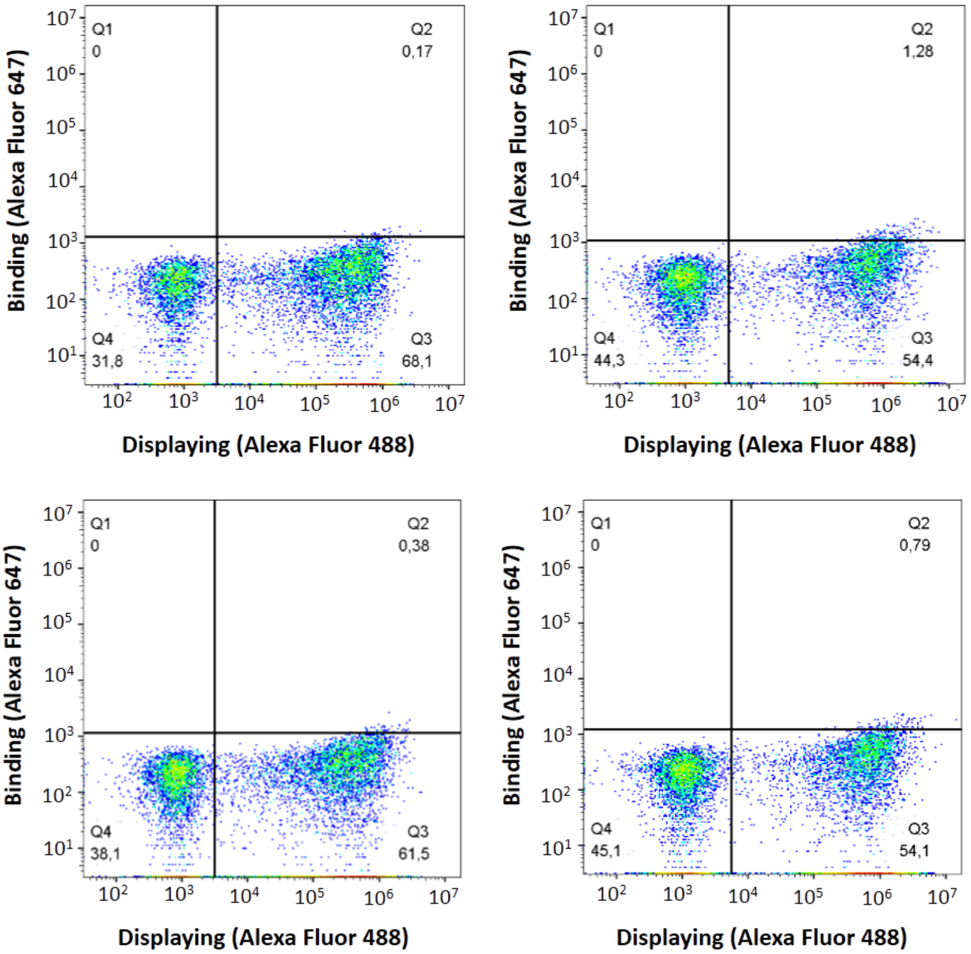


Figure 88*. Flow cytometry plots showing the results for full length protein expression in: top) library 11, and bottom) library 18.*

Once again, full length protein expression was assessed in almost 60% of yeast cells. So we were ready for FACS sorting.

First round of FACS sorting

First, we needed to incubate yeast cells with biotinylated hEGF. For this purpose we used two different strategies as it is shown in **Table 22.**

Table 22. *Labeling strategies used fot FACS sorting.*

	Sample 1 (200 µL)	Sample 2 (200 µL)
[Cells]	2 x 10^7 cells/mL	2 x 10^7 cells/mL
Chicken anti-HA IgY	1:200 (1 µL)	1:200 (1 µL)
Biotinylated hEGF	1 µM no-preloaded	250 nM preloaded
Alexa Fluor 488-goat anti-chicken IgG	1:200 (1 µL)	1:200 (1 µL)
Streptavidin-Alexa Fluor 647	1:200 (1 µL)	-----

For the hEGF-preloaded strategy (sample 2) we first incubated streptavidin-Alexa Fluor 647 with biotinylated-hEGF on ice for 30 min, shielded from light, at a streptavidin/hEGF molar ratio of 1:4 in as a small volume as possible. This sample was then mixed with yeast cells and PBSA was added to a final volume of 200 µL. Labeling with chicken anti-HA IgY was performed as always and the secondary labeling was done just adding Alexa Fluor 488-goat anti-chicken IgG. After labeling, FACS sorting was performed in order to enrich cultures with high affinity hEGF binders. FACS selection was done in the Koch Institute Flow Cytometry Core at MIT. First, analysis of labeled cells was performed. Plotting SSC-A vs FSC-A we first selected for living cells as they should present similar morphology (size and complexity). Dead cells may present lower FSC and higher SSC. Then, plotting SSC-W vs SSC-H one could select for single cells and eliminate any debris, clumps or doblets (**Figure 89**).

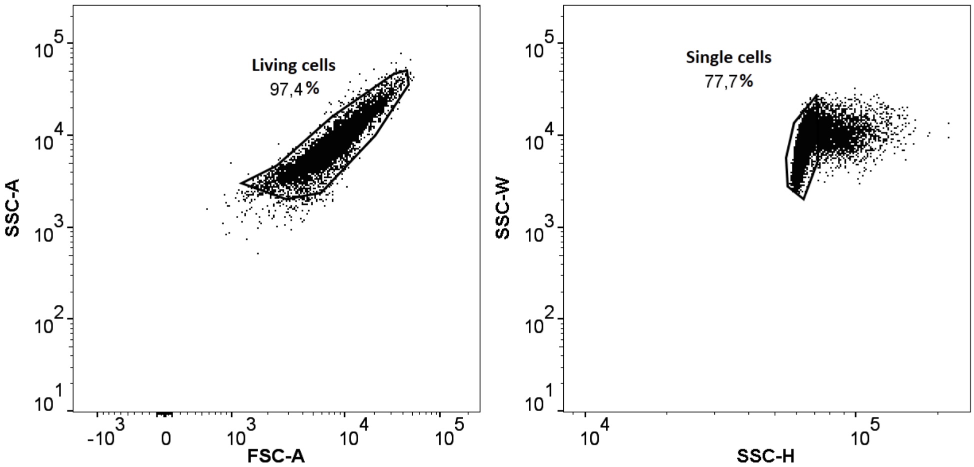


Figure 89. *Analysis of yeast cells prior to FACS sorting. Examples of selection for: living cells (left) and single cells (right).*

As we were using Alexa Fluor 647 (far-red fluorescent dye) to detect hEGF binding, and Alexa Fluor 488 (green fluorescent dye) to detect full length protein expression; we plotted APC-A (far-red fluorescence detector) vs FITC-A (green fluorescence detector) to select for Sso7d-based hEGF binders.

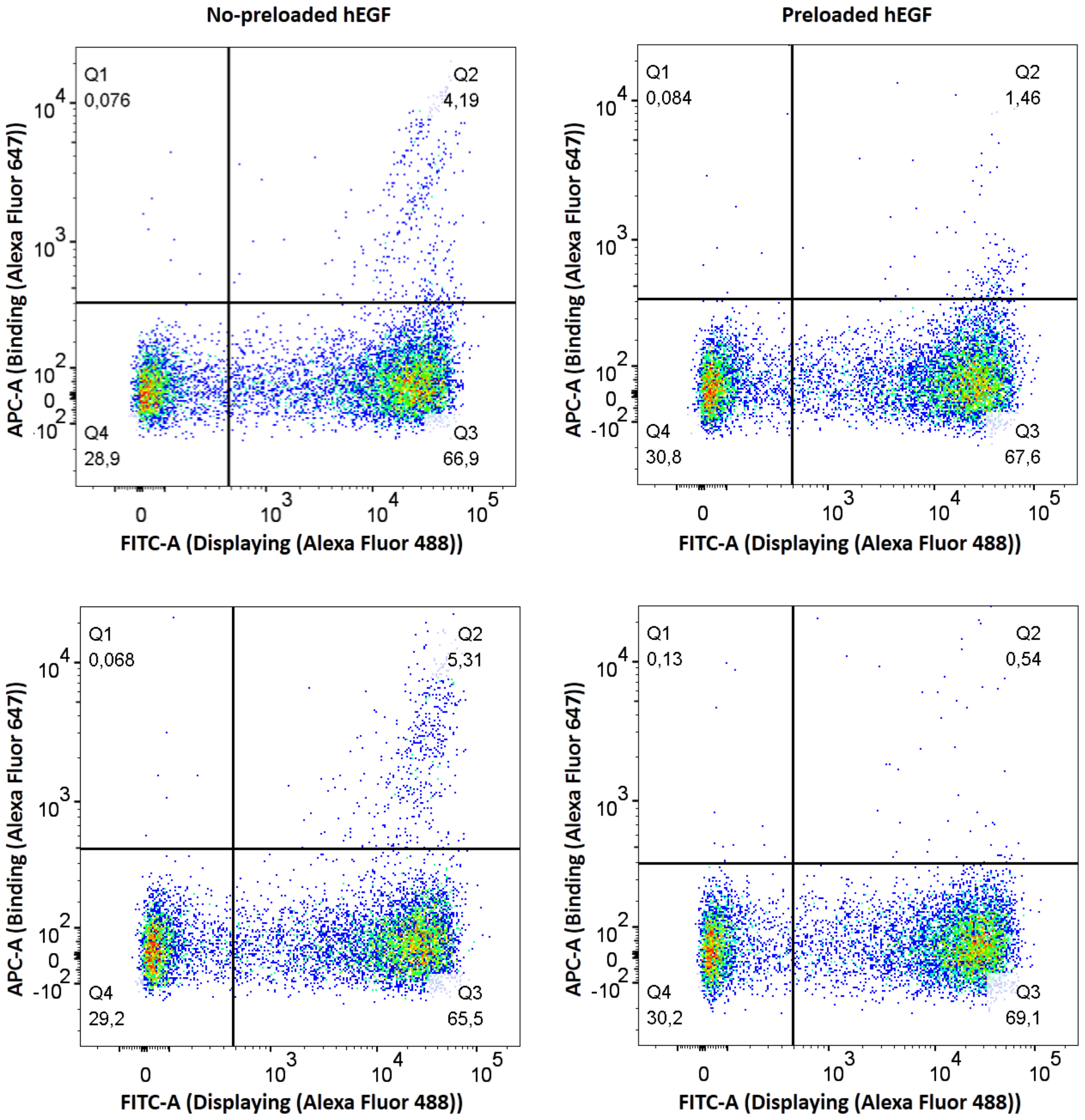


Figure 90. *Analysis of yeast cells searching for hEGF binders in: library 11 (top) and library 18 (bottom).*

As it can be noticed in **Figure 90**, some yeast cells showed positive staining for both Alexa Fluor 647 and 488 (cells that appeared in Q2), indicating full length protein expression and hEGF binding. As it could be observed the non-preloading methodology seemed to work better with this system, as we could observe more binders in the samples that were treated using this process, compared with those that were combined to preloaded hEGF. In initial FACS sorts it is recommended to select 1-3% of displaying cells that show detectable binding, as we wanted to make sure that we were not omitting even low affinity binders we sorted between 3-6% of displaying cells showing hEGF-binding (**Figure 91**). Collected cells were inoculated into 2 mL SD-CAA media and grown for 48h at 30ºC with shaking (250 rpm). When cells were completely grown ($OD_{600} \sim 9$) we passaged and induced them as previously described.

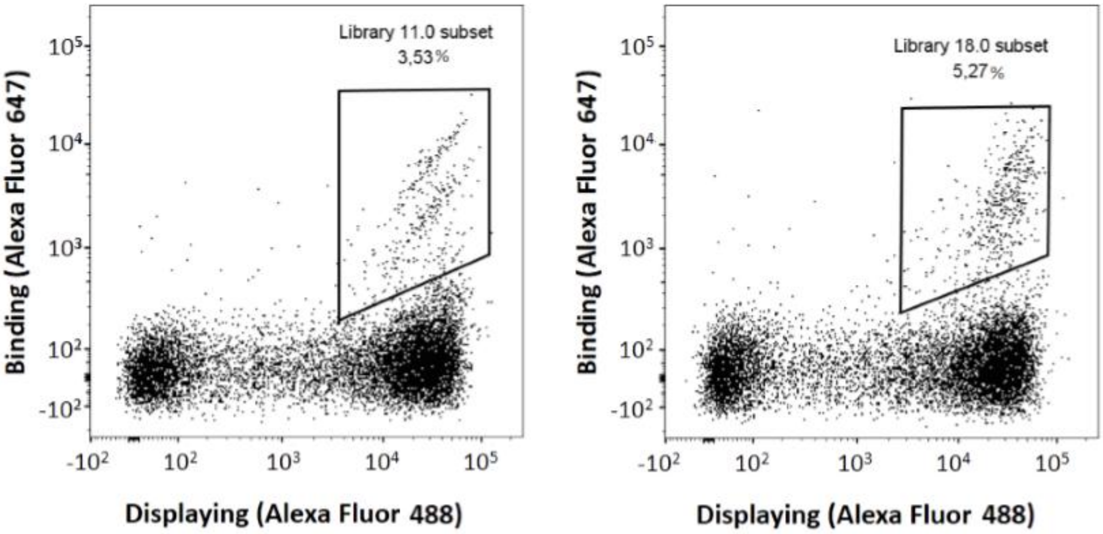


Figure 91. *Flow cytometry plots showing the first sorting of improved clones for: left) library 11, and right) library 18.*

Second round of FACS sorting

Newly induced cells were labeled as it was described in **Table 22** and FACS sorting was performed again. The results obtained for this second round of FACS selection are shown in **Figure 92**.

As it can be noticed, after the first round of FACS sorting, we were able to enrich for improved hEGF binders in both library 11 and library 18. In the results obtained for library 11 we could clearly see two different diagonals that may contain diverse hEGF binders so we selected them separately to enrich for both of them separately. In the results obtained for library 18 we could guess a diffused diagonal and some dispersed binders so we decided to collect them all together to enrich the culture media with all of them. In this second round of FACS we noticed that the process in which we used preloaded hEGF, seemed to be displaying higher affinity binders, especially for library 18. We also selected those binders in order to grow them separately (**Figure 93**).

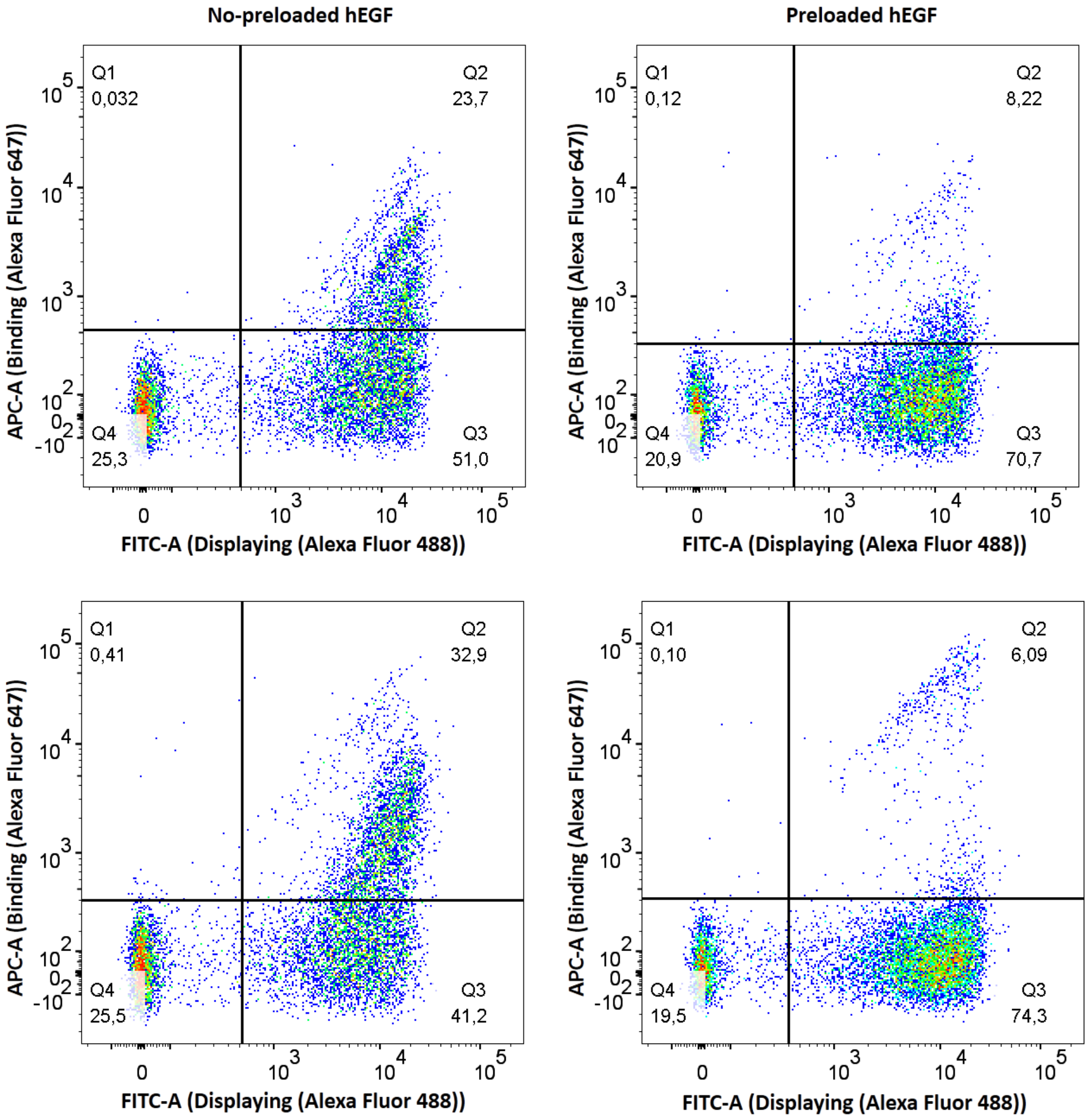


Figure 92. *Analysis of yeast cells searching for hEGF binders in: library 11 (top) and library 18 (bottom).*

Collected cells were inoculated again in SD-CAA media and grown for 48h at 30ºC. As I was running out of time because my stay at MIT was arriving to its end, this time we decided to induce protein expression in all samples except library 11 subset 2, which was used for plasmid extraction and single clone characterization to characterize at least some of the obtained binders.

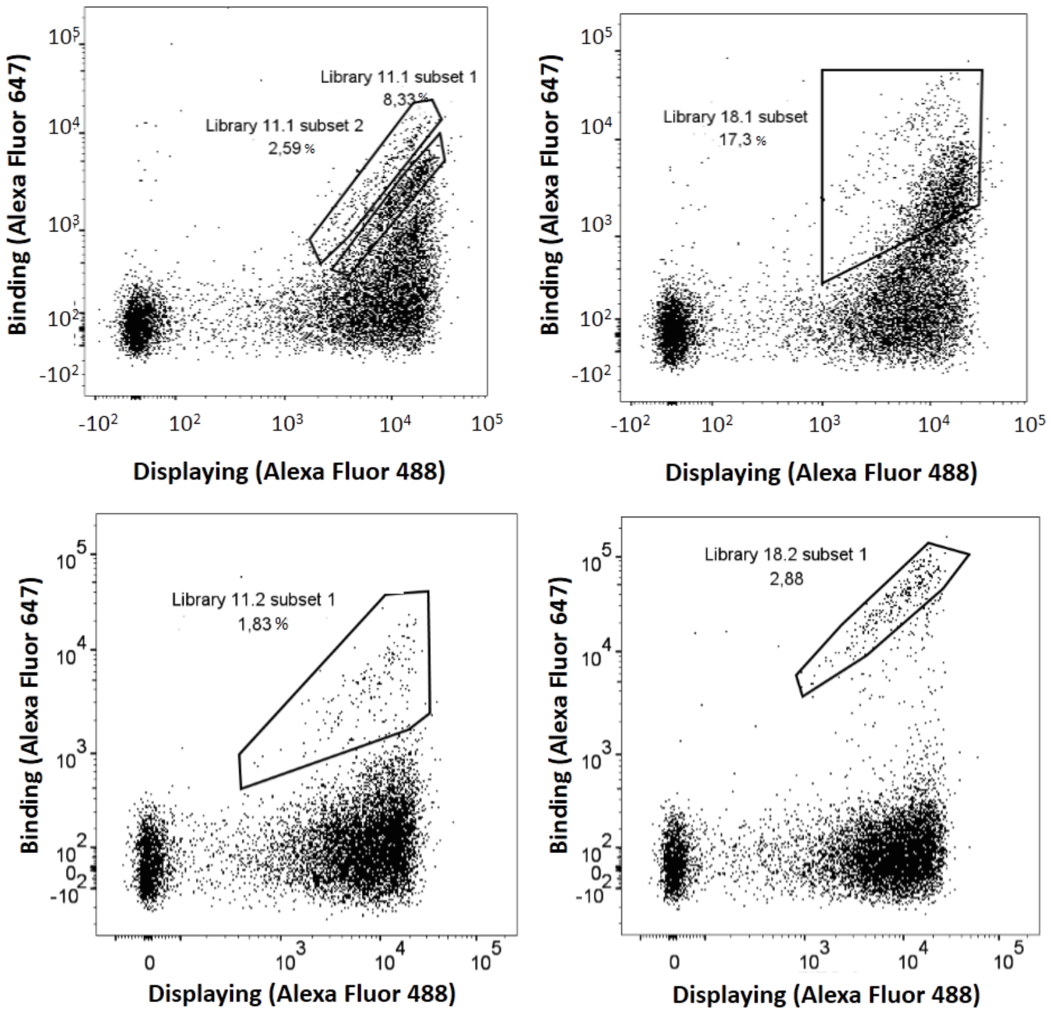
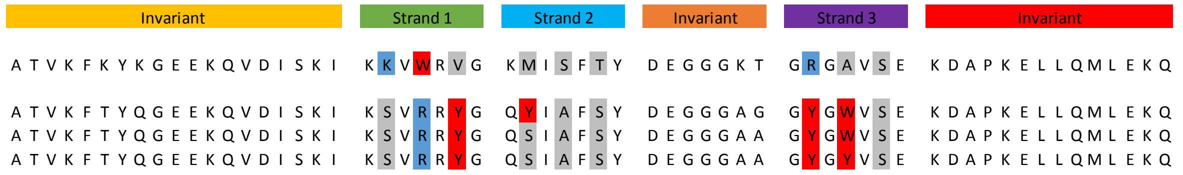


Figure 93. *Flow cytometry plots showing the second sorting of improved clones for: left) library 11, and right) library 18.*

Plasmid extraction and identification of individual cells

For this purpose 1 x 10^8 cells were pelleted and plasmid DNA was extracted using Zymoprep Yeast Plasmid Miniprep II. To facilitate DNA sequencing, the plasmids extracted from yeasts were amplified in *E. coli* XL1-Blue supercompetent cells and isolated using Wizard® *Plus* SV Minipreps DNA Purification Kit. Finally, the amino acid sequence of the selected hEGF binders was determined by DNA sequencing of the purified plasmids (**Figure 94**).



Invariant	Strand 1	Strand 2	Invariant	Strand 3	Invariant
A T V K F K Y K G E E K Q V D I S K I	K K V W R V G	K M I S F T Y	D E G G G K T	G R G A V S E	K D A P K E L L Q M L E K Q
A T V K F T Y Q G E E K Q V D I S K I	K S V R R Y G	Q Y I A F S Y	D E G G G A G	G Y G W V S E	K D A P K E L L Q M L E K Q
A T V K F T Y Q G E E K Q V D I S K I	K S V R R Y G	Q S I A F S Y	D E G G G A A	G Y G W V S E	K D A P K E L L Q M L E K Q
A T V K F T Y Q G E E K Q V D I S K I	K S V R R Y G	Q S I A F S Y	D E G G G A A	G Y G Y V S E	K D A P K E L L Q M L E K Q

Figure 94. *Comparison between wild type Sso7d sequence and those obtained for the first selected Sso7d-baed hEGF binders. Variant positions are highlighted: blue) charged residues, grey) non-charged residues, red) aromatic residues*

At that point of the project my stay at the MIT finished so I was not able to continue the next steps of the project. So far, Dr. Sanchini has been able to express one of the obtained Sso7d mutants in *E. coli* to perform interferometry assays in order to determine the affinity of this hEGF binder. The assayed compound showed a hEGF-affinity of ~3 µM. He also conducted docking simulations to select the best positions of one of these binders to be mutated to cysteine for dye introduction. W44 seemed to be the best position to introduce a fluorescent dye (**Figure 95**).

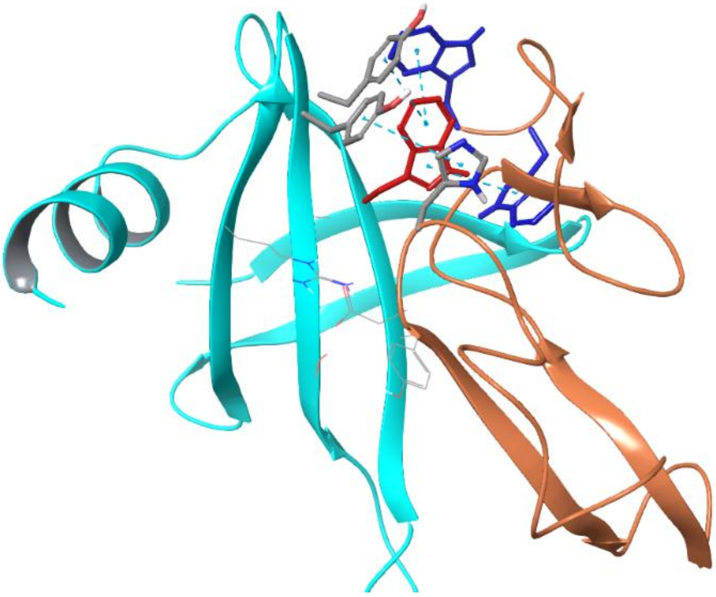


Figure 95. *Cartoon representation showing the interaction between one of the Sso7d mutants (depicted in blue) and hEGF (depicted in brown). Sticks illustrate a dense π-stacking network around W44 (red) on Sso7d variant involving W49 and W50 (blue) on hEGF and Y25 and Y28 (grey) on Sso7d itself.*

A third round of FACS was planned to be done with newly induced cells coming from the second round of FACS sorting to select for highest affinity binders. After that, error prone PCR will be used to introduce additional diversity or to combine potentially favorable mutations, followed by display of the new library and subsequent rounds of cell sorting by FACS (**Figure 96**). Typically, multiple rounds of mutagenesis and/or library sorting are applied to isolate high-affinity variants with equilibrium dissociation constants in the low nanomolar to picomolar range, which is the aim of this project.

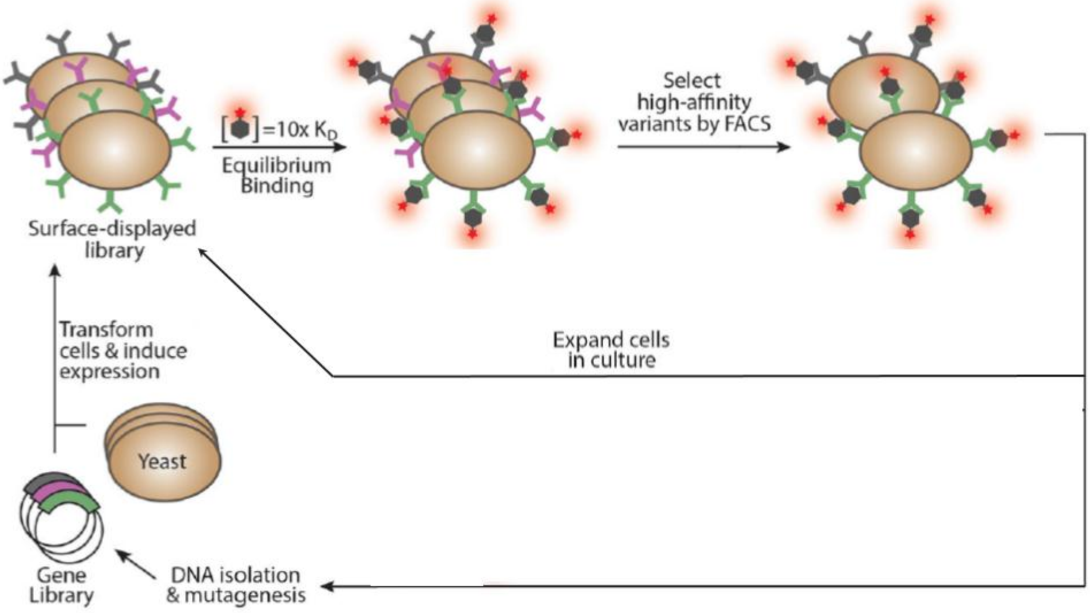


Figure 96. *Isolating high-affinity protein variants from a yeast-displayed library by FACS. Following transformation of yeast cells with a gene library and induction of surface expression the displayed library was labeled prior to screening. Sorted pools of yeast clones can be expanded in culture for either analysis or a subsequent round of sorting, or DNA from these clones can be isolated, subjected to mutagenesis, and used to transform a new batch of yeast for further directed protein evolution. Adapted from Cherf et al.*[272]

Once a hEGF binder with the desired affinity will be obtained, docking simulations will be performed to select the best place to introduce the fluorescent dye. Different 4-DMN derivatives will be tested and fluorescent studies will be conducted to select the best hEGF biosensor.

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 acquired to Schralau; while acetone, acetonitrile and toluene were obtained from SDS. All amino acids used for peptide synthesis were purchased to Iris Biotech and HOAt was acquired from Applied GL Biochem. 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.

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 Shiff-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^{o}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 n 50 mL of ethanol. This solution had to be protected from the light.

Chlroranil 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 µL of acetone. Subsequently, 20 µL of a stock solution of chloranil is added. After a short mixing the mixture is left at room temperature for 5 min and the beads inspected. Blue to green beads means free primary amines (positive test), whereas colorless 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 cyclic peptides were synthesized on 2-chlorotrityl chloride resin with a substitution of 0.6 mmol/g.

Treatment	Reagents
Wash	DCM (5 x 30s)
Wash	DMF (5 x 30s)
Wash	DCM (5 x 30s)
Coupling	0.7 eq Fmoc-Aa-OH, 10 eq DIEA, DCM (1h)
Capping	MeOH (0.8 mL/g resin), 15 min
Wash	DCM (5 x 30s)
Wash	DMF (5 x 30s)
Wash	DCM (5 x 30s)

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). To remove the Fmoc group from Fmoc-L-Pro-OH, an additional treatment with DBU, toluene, piperidine and DMF (5:5:20:70) was performed (1 x 1 min and 2 x 5 min). After deprotection, the resin was washed with DMF (5 x 30s), DCM (5 x 30s) and again with DMF (5 x 30s).

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 the following equation:

$$X = \frac{A \cdot V}{\varepsilon \cdot m \cdot b}$$

where X is the loading of the resin, A is Fmoc absorbance at 301 nm, V is the volume of solvent, ε is the molar extinction coefficient of Fmoc at 301 nm (7800 $M^{-1}cm^{-1}$), m is the mass of the resin in g and b is the optical path in cm.

Chain elongation

Method 1 was used as a default for all Fmoc-Aa-OH couplings. Method 2 was used for the third amino acid coupling to avoid diketopiperazine formation. When the first coupling was not complete, a second coupling was performed. When the recoupling was unsuccessful, method 3 was used.

Method 1 - HATU (3 eq.)

Treatment	Reagents
Wash	DMF (5 x 30s)
Coupling	3 eq Fmoc-Aa-OH, 3 eq HATU, 6 eq DIEA, DMF (35 min)
Wash	DMF (5 x 30s)
Wash	DCM (5 x 30s)
Wash	DMF (5 x 30s)

Method 2 - HATU (6 eq + 3 eq)

Treatment	Reagents
Wash	DMF (5 x 30s)
Coupling	6 eq Fmoc-Aa-OH, 6 eq HATU, 12 eq DIEA, DMF (20 min)
Recoupling	3 eq Fmoc-Aa-OH, 3 eq HATU, 6 eq DIEA, DMF (30 min)
Wash	DMF (5 x 30s)
Wash	DCM (5 x 30s)
Wash	DMF (5 x 30s)

Method 3 - PyAOP

Treatment	Reagents
Wash	DMF (5 x 30s)
Coupling	4 eq Fmoc-Aa-OH, 4 eq PyAOP, 12 eq DIEA, DMF (1h)
Wash	DMF (5 x 30s)
Wash	DCM (5 x 30s)
Wash	DMF (5 x 30s)

Cleavage without deprotection of side chains

Protected linear peptides were cleaved using 2% TFA in DCM (5 x 30s) to keep the protecting groups for the subsequent head to tail cyclization.

Head to tail cyclization

Linear protected peptides were dissolved at 5 mM concentration in DCM containing 2% DMF. Subsequently 2 eq of PyAOP and 6 eq of DIEA were added and cyclization was let to occur at room temperature for 16h. Reaction was followed by HPLC. After complete cyclization, the solvent was evaporated and the cyclic peptide was deprotected.

Side chains deprotection

Protecting groups' removal was performed using TFA, H_2O and TIS (92.5:5:2.5), letting the reaction to occur during 1.5h.

Work-up

After side chains deprotection, the solvent was evaporated applying a current of N_2. The residue was washed 3 times by suspension in diethyl ether and subsequent centrifugation. After that, the cleaved peptides were dissolved in H_2O/MeCN (1:1) and lyophilized.

Purification

Crude peptides were dissolved in MeCN and purified by ISCO CombiFlash®. Fractions corresponding to the peak of interest were analyzed by UPLC, pooled together and lyophilized.

Peptides were purified with a CombiFlash® automated flash chromatography (Teledyne ISCO). The column used was a RediSep Rf Gold® Reversed-phase C18, solvents MeCN (0.1 % TFA) and H_2O (0.1 % TFA), flow rate: 35 mL/min. Purity was checked by reverse-phase UPLC.

Peptide characterization

Identification and purity assessment

UPLC chromatograms were obtained on an Acquity high class (PDA eλ detector, sample manager FNT and Quaternary solvent manager) using an Acquity BEH C18 column (50 x 2 mm x 1.7 µm). The flow rate was 0.61 mL/min using MeCN (0.036 % TFA) and H_2O (0.045 % TFA). 2 min linear gradients were used in all cases.

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). The flow rate was 1 mL/min using MeCN (0.036 % TFA) and H_2O (0.045 % TFA). 8 min linear gradients were used in all cases.

The determination of the molecular weight of all synthesized peptides were performed using an Acquity high class (PDA eλ detector, sample manager FNT and Quaternary solvent manager) coupled with SQ detector 2. The column used was an Acquity BEH C18 (50 x 2 mm x 1.7 µm). The flow rate was 0.6 mL/min using MeCN (0.07 % formic acid) and H_2O (0.1 % formic acid). 2-min linear gradients were used in all cases.

Sometimes MALDI-TOF MS was used for molecular weight determination 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 MeCN/H_2O 1:1 (v/v) containing 0.1% TFA.

Quantification by amino acid analysis

The content and ratio of amino acids present in peptide samples were determined by ion exchange chromatographic analysis after acid hydrolysis. The hydrolysis was performed with 6M HCl at 110ºC for 16h. 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 spectropolarimeterdic, equipped with a CDF 426S/426L peltier. Spectra were obtained between 190 and 260 nm, with a time response of 2s, a scanning speed of 20 nm/min and a step resolution of 0.2 nm. Molar ellipticiy values were calculated from experimental ellipticity (in mdeg) using the following equation:

$$\theta = \frac{\theta_{exp}}{b \cdot C \cdot n}$$

where θ is molar ellipticity in mdeg·cm^2·mol^{-1}, θ_{exp} is measured ellipticity in mdeg, b is the optical path in cm, C is peptide concentration in M and n is the number of residues in the peptide.

Nuclear magnetic resonance

NMR experiments were carried out on a Bruker Avance III 600 MHz spectrometer equipped with a TCI cryoprobe. Samples for the characterization of hexacyclic peptides were prepared by dissolving peptides in 25 mM sodium phosphate, 50 mM NaCl, pH 7.4 buffer containing 10% D_2O at 3-4 mM. Chemical shifts were referenced to internal sodium-3-(trimethylsilyl)propanesulfonate (DSS).

Suppression of the water signal was achieved by excitation sculpting.[273] Residue specific assignments were obtained from 2D total correlated spectroscopy (TOCSY)[274] and correlation spectroscopy (COSY)[275] experiments, while 2D nuclear Overhauser effect spectroscopy (NOESY)[276] permitted sequence specific assignments. ^{13}C resonances were assigned from 2D ^{1}H-^{13}C HSQC spectra. All experiments were performed at 25ºC except NOESY spectra that were acquired at 5ºC. Amide proton temperature coefficients were determined from a series of one dimensional spectra acquired between 5ºC and 35ºC. The TOCSY and NOESY mixing times were 70 and 450 ms, respectively.

Apamin analogues were dissolved in H_2O:D_2O (90:10) at 4 mM concentration and a drop of chlorhidric acid was added to adjust the pH at 2-3. Additional samples of these peptides were also dissolved in 10 mM phosphate buffer at pH 7.4 to perform the NMR experiments at physiological pH. Assignments were obtained from 2D homonuclear (TOCSY, COSY and NOESY) and heteronuclear (^{1}H-^{13}C HSQC) experiments. TOCSY and NOESY mixing times were 70 and 250 ms, respectively.

Simulated Annealing

Structures for MiniAp-1 were generated by the standard simulated annealing protocol implemented in the CNS software.[277] Only the distance restraints from inter-residue NOEs were used for the calculation. NOEs were classified as strong, medium and weak (upper limits for structure calculation were set as 2.5 Å, 3.5 Å and 4.5 Å, respectively). The necessary pseudoatom corrections were applied for non-stereospecifically assigned protons at prochiral centers and for the methyl group of aliphatic side chains. The Φ and Ψ backbone torsion angle restraints included in the calculation were derived from experimental ^{1}H, ^{13}C and ^{15}N chemical shifts values using the PREDITOR server.[278] 80 structures were generated and 10 were selected based on lowest overall energy and on the basis of no violations of NOE or dihedral angle constraints greater than 0.2 Å and 5º, respectively. PROCHECK[279] was used to generate Ramachandran plot statistics of the final structures.

Replica exchange molecular dynamics (REMD)

REMD simulation began with a cyclic structure of the peptide built with XLEaP program from the AMBER14 molecular mechanics package[204] and corresponding chiralities where assigned to different residues. For all the systems, the Amber ff99SB force field[205] was used, together with the re-optimized omega-bond angle parameters.[206] The initial peptide structure 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 503 K.[280] Generalized Born model[281] 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[282]. Integral time step was set to 1 fs. Temperature was regulated using Berendsen thermostat[283] with a coupling time constant of 1 ps. SHAKE algorithm[284] was used to constrain all the covalent bonds involving hydrogen atoms. Swaps were attempted every 2 ps and MD simulations were extended to 50 ns for each replica. Snapshots were saved every 2 ps. Matplotlib software[285] was used to generate Ramachandran plot of the final cyclic hexapeptide structures.

Protein production

Expression and purification of selectively labeled ^{13}C-methyl methionine VEGF

Protein expression

Competent B834 (DE3) E.coli cells were transformed with p6xHisVEGF$_{11\text{-}109}$ and pMS421 plasmids and plated on solid LB agar media plates. 5 mL of overnight carbenicilin/streptomycin containing LB rich media were inoculated into 1 L of M9 minimal media containing 80 mg/mL methionine-(methyl-^{13}C) and 1 g/L ammonium-^{15}N chloride. Growth was monitored using optical density at 600 nm until a value of 0.7-0.8 was reached and then protein expression was induced with IPTG (final concentration 1 mM) for 4 to 6 hour at 37^{o}C.

Lysis and IMAC purification

Harvested cells were dissolved in 6 M guanidine HCl, 0.1 M NaH_2PO_4, 10 mM Tris, 10 mM 2-mercaptoethanol, 15 mM imidazole, pH 8. Cells were lysated by sonication. The lysate was filtered and then purified by FPLC (ÄKTA explorer) using a 5 mL HisTrap^{TM} HP column (GE Healthcare) at 4^{o}C. Subsequent washes were performed with: a) 8 M urea, 0.1 M NaH_2PO_4, 10 mM Tris, 5 mM DTT, 15 mM imidazole, pH 8, b) 8 M urea, 0.1 M NaH_2PO_4, 10 mM Tris, 5 mM DTT, 15 mM imidazole, pH 6.3, and c) 8 M urea, 0.1 M NaH_2PO_4, 10 mM Tris, 5 mM DTT, 100 mM imidazole, pH 5.9. Protein was eluted with 8 M urea, 0.1 M NaH_2PO_4, 10 mM Tris, 5 mM DTT, 500 mM imidazole, pH 5.9. Fractions were analyzed by SDS-PAGE electrophoresis and pure fractions were pooled.

SDS-PAGE electrophoresis

SDS-PAGE electrophoresis was carried out using BioRad system (Mini-protean cell) 7.5% Tris gel and 25 mM Tris, 192 mM glycine, 0.1% SDS (w/v) as running buffer. Sample buffer was 62.5 mM Tris-HCl, pH 6.8, 2% SDS (w/v), 25% (v/v) glycerol, 0.01% (v/v) bromophenol blue, 5% β-mercaptoethanol (mercaptoethanol was not added for non-reducing SDS-PAGE). Protein molecular weights were approximated by comparison to a protein marker (Perfect Protein Markers 15–150 kDa from Novagen). Gels were visualized by coomassie staining (staining solution: 10% (v/v) AcOH, 0.25 g brilliant blue; discoloration solution: 20% (v/v) MeOH, 3% (v/v) AcOH glacial, in water).

Protein refolding

Protein concentration was estimated by UV absorbance at 280 nm (using $\varepsilon_{280} = 0.47$ mL · mg^{-1} · cm^{-1}) and diluted to 1 mg/mL. The pH of the sample was adjusted to 8.0 and DTT was added to a final concentration of 20 mM. Complete reduction was allowed to proceed for 3h at room temperature with gentle stirring in the dark. Refolding was then proceed stepwise through dialysis at 4^{o}C, using Spectra Pore dialysis membrane (MWCO 6-8 KDa from Spectrum Labs), starting with 20 mM Tris HCl, 8 M urea, 25 mM cysteine, pH 8.4. Refolding yield was assessed by non-reducing SDS-PAGE electrophoresis.

Enzymatic digestion

Histag proteolytic cleavage was performed using a 100:1 VEGF/Genease I (from New England Biolabs) ratio in 20 mM Tris HCl, 200 mM NaCl, 1 mM EDTA, pH 8.4 buffer. Reaction was let to occur overnight at room temperature with gentle stirring. Digested protein was then dialyzed into 20 mM Tris HCl, pH 7.5 at 4^{o}C.

Anion exchange chromatography

Further protein purification was performed by anion exchange chromatography (ÄKTA explorer) using a 5 mL HiTrap^{TM} Q HP column (GE Healthcare) at 4^{o}C. 20 mM Tris HCl, pH 7.5 was used as the binding buffer, while protein was eluted with increasing NaCl concentrations up to 1 M, in 20 mM Tris HCl, pH 7.5. Protein purity was assessed by SDS-PAGE electrophoresis and pure fractions were pooled and concentrated by ultrafiltration prior to S75 size exclusion purification.

Size exclusion chromatography

Final size exclusion purification chromatography by FPLC was performed at 4^{o}C with a HiLoad 16/600 Superdex 75 pg column (GE Healthcare), using 25 mM phosphate buffer, 50 mM NaCl, pH 7.4 buffer. Pure protein fractions were checked by non-reducing SDS-PAGE electrophoresis, pooled together and concentrated using Amicon filters MWCO 10 KDa (Merck Millipore).

Expression and purification of Sso7d cysteine mutants

Protein expression

While Dr. De Picciotto left glycerol stocks for W25C and G26C M11.1.3 mutants, only a semi-empty microcentrifuge tube containing cDNA for I23C was found. This residual DNA from Dr. Picciotto was used to transform BL21(DE3) competent cells. Transformed cells were plated on LB plates with kanamycin (30 µg/mL) and chloramphenicol (30 µg/mL) and incubated at 37^{o}C overnight. Five colonies were inoculated in 5 mL LB media each with kanamycin (30 µg/mL) and chloramphenicol (30 µg/mL) and incubated at 37^{o}C with shaking (250 rpm) overnight. Plasmid DNAs were isolated using Wizard® *Plus* SV Minipreps DNA Purification Kit and send for sequencing to check correct protein expression. 10 mL of overnight culture that was not used for plasmid extraction, was diluted with 10 mL of LB, 20% (v/v) glycerol to prepare glycerol stocks.

Glycerol stocks of the Sso7d cysteine mutants were inoculated in LB media containing 30 µg/mL kanamycin and grown at 37^{o}C overnight with shaking (250 rpm). 5 mL of pre-culture media were inoculated into 1 L of TB with 30 µg/mL kanamycin and grown at 37^{o}C with shaking (250 rpm) until OD_{600} = 0.8. Induction was performed adding IPTG (final concentration 1 mM) and incubating at 16^{o}C with shaking (250 rpm) overnight.

Lysis and IMAC purification

Cultures were pelleted at 3500 rpm for 30 min at 4^{o}C and cells were resuspended in 50 mM sodium phosphate, 300 mM NaCl, 1% (v/v) Triton-X-100, 3% (v/v) glycerol, 2 mM mercaptoethanol, pH 8.0, and pelleted again at 3500 rpm for 30 min at 4ºC. Lysis was done by resuspending the cells in 50 mM sodium phosphate, 300 mM NaCl, 1% Triton-X-100, 3% (v/v) glycerol, 2 mM mercaptoethanol, pH 8.0, and sonicating (3 cycles of 45 s) at 50% amplitude. Finally cells were centrifuged at 41KG for 45 min at 4^{o}C and the supernatant was incubated with Ni-NTA resin (Qiagen) previously equilibrated with 50 mM Na_2HPO_4, 300 mM NaCl, 2 mM mercaptoethanol, 6 M urea, 15 mM imidazole, pH 8.0 buffer. Protein was washed with 50 mM Na_2HPO_4, 300 mM NaCl, 2 mM BME, 6 M urea, pH 8.0 buffer with increasing amounts of imidazole (15, 25 and 40 mM). Protein was finally eluted with the same buffer containing 500 mM imidazole. Protein purity was assessed by Tris-Tricine electrophoresis.

Tris-Tricine electrophoresis (tricine gels)

Tris-tricine electrophoresis was carried out using BioRad system (Mini-protean cell) 16.5 % Tris gel and 100 mM Tris, 100 mM Tricine, 0.1% SDS as running buffer. Sample buffer was 200 mM Tris-HCl, pH 6.8, 2% (w/v) SDS, 40% glycerol, 0.04% Coomassie Brilliant Blue G-250 (ThermoFisher), 2% (v/v) mercaptoethanol (added freshly). Protein molecular weights were approximated by comparison to a protein marker (SeeBlue® Plus 2 pre-stained protein standard from ThermoFisher). Gels were visualized by Coomassie staining (Imperial^{TM} Protein Stain from ThermoFisher).

Salting out precipitation

Small amounts of ammonium sulfate were added followed by incubation at 4°C for 10 min and centrifugation at 3750 rpm for 5 min at 4°C. Precipitate was resuspended in the minimal volume of 1 x PBS buffer and the supernatant was subsequently treated with increasing amounts of ammonium sulfate until we reach a saturated concentration. Protein purity was assessed by Tris-Tricine electrophoresis.

Cation exchange chromatography

$(His)_6$-Sso7d mutants were purified by FPLC using a 1 mL HiTrap SP^{TM} FF column (GE Healthcare). 50 mM HEPES, 5 mM NaCl, 6 M urea, pH 7.0 was used as binding buffer. Protein was eluted with the same buffer containing 1 M NaCl. Protein purity was assessed by Tris-Tricine electrophoresis.

For labeled proteins, in order to get rid of excess dye, the same method was used but no urea was added to any of the buffers. Protein purity and labeling was assessed by Tris-Tricine electrophoresis both by UV (360 nm) and Coomassie staining.

Protein labeling with 4-DMN analogs and IANBD

Labeling was done incubating the Sso7d mutants at 50 - 100 µM concentration with 10 equivalents of dye, in 50 mM Na_2HPO_4, 300 mM NaCl, 100 mM TCEP, pH 7.9 buffer. Reaction was performed overnight at room temperature and protected from light. Excess dye was removed by cation exchange chromatography.

Expression and purification of biotynilated-hEGF

Protein expression

hEGF-LPRTGGG glycerol stocks were grown in 5 mL of LB containing kanamycin (20 µg/mL) while shaking (220 rpm) for 12-16 h at 37°C. This culture was used to inoculate 2 x 1 liter of LB containing kanamycin (20 µg/mL). The inoculated cultures were grown while shaking (220 rpm) at 37°C until the optical density reached 0.6-0.8 as measured by UV absorbance at 600 nm. At this time, the temperature was reduced to 16°C and protein overexpression was induced by the addition of IPTG to a final concentration of 1 mM. The cultures were grown for an additional 20-24 h at 16°C and cells were harvested by centrifugation at 4°C for 30 min at 3600 rpm. The pelleted cells were resuspended in lysis buffer (50 mM HEPES, 150 mM NaCl, 5% (v/v) glycerol, pH 7.9), and centrifuged for an additional 30 min (4°C, 3600 rpm). The supernatant was decanted and the resulting cell pellet was stored at -80°C until lysis. For lysis, cell pellets were resuspended at 4°C in 40 mL of chilled lysis buffer containing 1:1000 of a protease inhibitor cocktail (Calbiochem) and lysed by sonication (40% amplitude, 3 x 1 min with one second on/off pulsing and 5 min ice incubations between sonications). The lysate was then centrifuged at 4°C for 65 min at 37000 rpm.

IMAC purification

Lysates were then batch-bound for at least 2 h to ~ 3-5 mL of Ni-NTA resin (Qiagen) equilibrated with Ni-NTA column buffer (50 mM HEPES, 300 mM NaCl, pH 7.9) with 15 mM imidazole. The flow through was collected and the protein fusions were eluted by washing the resin with column buffer containing increasing concentrations of imidazole (15 mM, 40 mM, 250 mM). All fractions were analyzed by tricine gels. Pure protein-containing fractions were combined, concentrated using Amicon 10K MWCO centrifuge filters (Merck Millipore), and imidazole was removed by dialysis.

SUMO cleavage

SUMO was cleaved by addition of 20 nmol of SUMO protease to the dialyzed protein solution with 2 mM final concentration of DTT and incubating at 30°C for 1-2h. SUMO protease and the cleaved SUMO domain were removed via binding to Ni-NTA resin (Qiagen), providing the free sortase-sequence-modified cytokine in the flow through.

Protein refolding

EGF was refolded by dialysis against 20 mM Tris HCl, 150 mM NaCl, 1 mM reduced glutathione, 0.1 mM oxidized glutathione, pH: 7.9 (2 x 4L, 40 h, 4°C). The protein was then concentrated using Amicon 3K MWCO centrifuge filters (Merck Millipore), as described and buffer exchanged into 50 mM HEPES, 150 mM NaCl, 5% (v/v) glycerol, pH 7.9, aliquoted, and stored at -80°C until further use. Concentrations were determined by UV absorbance at 280 nm using the molar extinction coefficient of 18.140 M^{-1} cm^{-1}.

Sortase-mediated ligation

10 eq. of GGG-PEG_{12}-YK(biotin)KT peptide were added to a sample containing ~40 µM hEGF-LPRTGGG and 0.3 eq. SrtA, in 50 mM HEPES, 150 mM NaCl, 10 mM $CaCl_2$, pH 7.5. Reaction was set up at room temperature with tumbling during 30 min. Protein biotinylation was confirmed by HPLC-MS. Purification was done by means of IMAC adding 1 mL of Ni-NTA resin (Qiagen) to the sample and incubating for 15 min at 4°C. Protein containing fractions were pooled together and concentrated using 3K MWCO Amicon centrifuge filters (Merck Millipore). Final purification was done by means of size exclusion chromatography with an AKTA Pure FPLC at 4°C.

Size exclusion chromatography

Size exclusion purification was performed with a Superdex 200 pg column (GE Healthcare), using 25 mM phosphate buffer, 50 mM HEPES, 150 mM NaCl, pH 7.5 buffer. Pure protein fractions were checked by non-reducing SDS-PAGE electrophoresis and pooled together.

Protein-ligand interaction experiments

VEGF sample preparation for NMR

Protein concentration was adjusted to 50 µM VEGF. NaN_3 was added to a final concentration of 0.02% (w/v). All protein samples contained 10% deuterated water (Cortecnet), necessary to lock the magnetic field. Aliquots of 170 µL were prepared and shock frozen to be stored at -20°C. Aliquots for STD experiments were prepared by changing the buffer to 25 mM phosphate buffer, 50 mM NaCl, 0.02% NaN_3, pH 7.4 in 100% D_2O using PD-10 desalting columns (GE Healthcare). Final protein concentration was 1 µM in this case.

Peptide stocks preparation for NMR

Peptide stocks were prepared by dissolving pure peptides in DMSO-d_6 (Cortecnet) at 100 mM concentration, in order to improve peptide solubility. Stocks were aliquoted and kept at -20°C.

Nuclear magnetic resonance (NMR)

Protein-based NMR experiments

HSQC experiments were conducted on Bruker Avance III 600 MHz spectrometer equipped with a TCI cryoprobe. For chemical shift perturbation (CSP) experiments, 2D ^1H-^{13}C HSQC spectra were recorded with 128x512 to 96x256 complex points with 8 transients per increment. 2D ^1H-^{15}N HSQC spectra were acquired with 128x2048 complex points with a total of 8 transients per increment. The x-carrier frequency was referenced to internal DSS-$_{d6}$. Indirect referencing was used in the ^{13}C and ^{15}N dimension by using the conversion factors from Wishart et al.[286] 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 qsine function was used for line broadening.

For the determination of CSP values, chemical shift changes for ^{13}C-methyl-methionine or amide ^1H of VEGF in the free and bound forms were calculated using **Equation 1** and **2**:

$$CSP = \sqrt{(\Delta\delta_H)^2 + \left(\frac{1}{4}\Delta\delta_C\right)^2} \qquad \textbf{Equation 1}$$

$$CSP = \sqrt{(\Delta\delta_H)^2 + \left(\frac{1}{5}\Delta\delta_N\right)^2} \qquad \textbf{Equation 2}$$

where $\Delta\delta_H = \delta_H(\text{free}) - \delta_H(\text{bound})$; $\Delta\delta_C = \delta_C(\text{free}) - \delta_C(\text{bound})$ and $\Delta\delta_N = \delta_N(\text{free}) - \delta_N(\text{bound})$.

Assessment of the affinity of cyclic peptides for VEGF was performed acquiring ^{1}H-^{13}C HSQC spectra of 50 µM VEGF at 298 K, with increasing concentrations of peptide (from 250 µM to 2.5 or 5 mM). Chemical shift changes were plotted against ligand concentration and fitted to the "One site-specific binding" equation in the GraphPad Prism software.

Binding sites obtained from ^{1}H-^{15}N HSQC spectra were set up on the basis of significant CSP. This was defined as normalized CSP ≥ (mean value of normalized chemical shifts + SD). Significant chemical shift changes were plotted on the protein surface using PyMOL software. NMR structure of the complex VEGF-v107 (PDB file: 1KAT) was used.

Ligand-based NMR experiments

Saturation transfer difference (STD) experiments were acquired at 5°C for samples containing 10 µM VEGF. Water suppression was achieved by excitation sculpting. For selective excitation of the protein, a train of Gaussian pulses with a length of 50 ms each was used for a total saturation time of 2s. To remove residual protein signals the length of the spin-lock filter was set to 20 ms. Selective irradiation of protein signals was applied at 0 ppm (on-resonance), and irradiation was applied at 60 ppm for acquisition of the off-resonance spectrum.

Virtual screening (docking)

Protein-ligand docking studies were carried out using the docking module implemented in MOE 2015 (Molecular Operating Environment)[196]. The X-ray structure of VEGF was retrieved from Protein Data Bank (PDB id: 1FLT), a protein system co-crystallized with the VEGF receptor, whose residues were removed prior to conduct docking analysis. The preparatory docking steps of the apo VEGF included addition of missing atoms, removal of duplicated atoms (such in the case of double occupancies), and assignment of the most adequate protonation state of histidine residues at physiological pH (7.4). Crystallographic water molecules closer than 3.5 Å from any protein atom were retained. The apo VEGF binding site was assigned with the SiteFinder module implemented in MOE. In this particular case, the binding site amino acids were identified if their coordinates were at less than 3.5 Å of any VEGF receptor atom. In total, a docking binding site surface larger than 1500 Å^{2} was generated. To determine the binding mode of each cyclic peptide structures, a docking multilevel strategy, was carried out. Formerly, cyclic peptide structure were subjected to conformational analysis using Conformational Search module (MMFF94x force field) in MOE in an energy window of 15 kcal/mol, rendering a total of 1000 conformers. The 100 lowest energy conformations were placed at any region of the binding site using the Triangle Matcher and scored by London dG methods respectively. The 15 best poses of each compound were further refined using the GBVI/WSA dG method consisting in 500 minimization steps to identify potential induced fit effects promoted by the ligand when bound. The best scored conformation after refinement protocol are expressed as kcal/mol.

Fluorescence polarization assays

Direct binding assay

Wells of a black Corning 384-well polystyrene plate contained 67 nM fluorescent peptide tracer (Flu-X-**C**DIHVMWEWE**C**FERL-NH_2, where X = 2-(2-amino-ethoxy)-ethoxy}-acetic acid) and increasing concentrations (from 0.63 pM to 3 µM) of VEGF protein in FP Buffer (50 mM NaCl, 16.2 mM Na_2HPO_4, 3.8 mM KH_2PO_4, 0.15 mM NaN_3, 0.15 mM EDTA, 0.5 mg/mL Pluronic; pH 7.4).[287] Plates were read after a 2 h incubation at room temperature using a PerkinElmer EnVision multi-label plate reader (Wellesley, MA) with polarized filters and optical modules for fluorescein ($\lambda_{excitation}$: 480 nm; $\lambda_{emission}$: 535 nm) or BODIPY^{TMR} ($\lambda_{excitation}$: 531 nm; $\lambda_{emission}$: 595 nm). mP values were calculated from raw parallel and perpendicular fluorescence intensities.[288] GraphPad Prism software was used to plot mP vs. VEGF concentration and the curve was fit to a “single-site binding model” to extract a binding dissociation constant (K_D value) for the tracer.[289] Experiments were performed in duplicate.

Competition binding assay

Wells of a 384 plate contained 67 nM fluorescein tracer peptide, 1 µM VEGF protein, and 2 µL tested inhibitor dissolved in DMSO (final concentration from 390 nM to 2.5 mM) in a final volume of 50 µL in FP buffer. Plates were read after a 4 h incubation at room temperature, the time necessary for complete equilibration. Experiments were performed in duplicate. The equilibrium dissociation constant (K_i)[290] was calculated as calculated by GraphPad Prism using a “single-site binding model” equation.

Microscale thermophoresis assays

For performing the MST experiments a fluorescent label (NT-647 from NanoTemper) was covalently attached to the VEGF protein (NHS coupling) following the labeling protocol from NanoTemper. Concentration of NT-647 labeled VEGF was kept constant (20 nM), while the concentration of the non-labeled cyclic peptide was varied between 2.5 mM – 33 µM. The assay was performed in 50 mM, 25 mM phosphate buffer, 2.5% DMSO, pH 7.4 at 298 K. After a short incubation the samples were loaded into MST NT.115 standard glass capillaries (NanoTemper) and the MST analysis was performed using the Monolith NT.115. Data analysis was performed using the NanoTemper analysis software.

Fluorescence spectroscopy assays

Fluorescence spectroscopy assays, in this work, were performed using a FluoroMax-2 fluorometer with an integration time of 0.1 s, excitation and emission slits set at 5 and 10 nm respectively. Stock solutions of 4-DMN labeled Sso7d mutants were prepared at 600 nM in 1 x PBS at pH 7.4. A stock solution of MSA at 100 µM was also prepared in the same buffer. Using these protein stocks, the labeled Sso7d mutant was diluted to 300 nM and exposed to increasing concentrations of MSA (0, 0.6, 2, 3, 6, 10 µM) until signal saturation was reached. The samples were excited at 420 nm and emission collected from 450 - 650 nm. The fold-increase for each Sso7d mutant was determined by comparing the background fluorescence spectra of the labeled protein with the fluorescence spectra containing the highest concentration of MSA measured. The largest difference in emission is reported as the fold-increase upon MSA binding. All spectra were corrected for background fluorescence by subtracting a blank scan of the buffer system.

Yeast surface display

Reagents set up

SD-CAA media (pH 4.5)

This media was prepared by dissolving 20 g dextrose, 6.7 g yeast nitrogen base (Difco), 5 g casamino acids (Bacto), 14.7 g sodium citrate and 4.29 g citric acid monohydrate deionized water to a volume of 1 L. This media was then sterilized by filtration. PenStrep 1:100 (Invitrogen) was added to prevent bacterial contamination

SG-CAA media (pH 4.5)

This media was prepared by dissolving 20 g glucose, 6.7 g yeast nitrogen base (Difco), 5 g casamino acids (Bacto), 14.7 g sodium citrate and 4.29 g citric acid monohydrate deionized water to a volume of 1 L. This media was then sterilized by filtration. PenStrep 1:100 (Invitrogen) was added to prevent bacterial contamination

PBSA

This buffer was prepared by dissolving 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4 and 1 g bovine serum albumin in 1 liter of deionized water. pH was adjusted to 7.4 and solution was sterilized by filtration. This solution can be stored for up to 6 months at 4^oC.

SD-CAA plates

SD-CAA plates were prepared by dissolving 5.4 g Na_2HPO_4, 8.56 g $NaH_2PO_4 \cdot H_2O$, 182 g sorbitol and 15 g agar in deionized water to a volume of 900 mL and autoclave (*Solution A*). Another solution was prepared by dissolving 20 g dextrose, 6.7 g yeast nitrogen base (Difco), 5 g casamino acids (Bacto) in deionized water to a volume of 100 mL and sterilize by filtration (*Solution B*). Then *Solution A* was cooled with stirring until below 50^oC. Both solutions were mixed and plates were poured.

Naïve library growth

Frozen aliquots of Sso7d libraries glycerol stocks (provided by the Wittrup's lab) with 10 x diversity (3x10^{10} cells) were thawed at room temperature and inoculated into 1 L of SD-CAA media. Cells were grown overnight at 30°C with shaking (250 rpm). Yeast cells typically grow overnight to an OD_{600} = 6-8. Cells were passaged into fresh SD-CAA media by pelleting at least 10 x diversity (3x10^{10} cells) and resuspending them into fresh SD-CAA media. An absorbance of 1 at 600 nm is about 1 x 10^7 cells/mL. Fresh cells were grown overnight at 30°C with shaking (250 rpm). Yeasts were passaged again into fresh SD-CAA media as described and grown at 30°C with shaking (250 rpm) until OD_{600} = 2-5. Then at least 10 x diversity cells were pelleted and resuspended in 1L of SG-CAA media with 2 g/L dextrose. Protein expression was induced by incubating overnight at 20°C with shaking (250 rpm).

First round of magnetic beads selection (positive selection to enrich for target binders):

First, 4x10^7 Dynabeads® Biotin Binder (Invitrogen), for each library, were added to a 2 mL microcentrifuge tube containing 1 mL of ice-cold PBSA. Beads were incubated for 5 min at 4°C with tumbling. Then microcentrifuge tubes containing magnetic beads were placed on the DynaMag-2 magnet (Invitrogen) for 3 min and supernatant was removed being careful not to touch the beads. Magnetic beads were washed with 1 mL of ice-cold PBSA, incubated for 5 min at 4°C with tumbling and finally they were placed in the magnet for 3 min to remove the supernatant. Three washes were performed. After supernatant removal 330 pmols of biotinylated-hEGF were combined with the beads in a final volume of 1 mL PBSA. Then beads were incubated at 4°C with tumbling for 2h. After incubation the microcentrifuge tubes containing hEGF-coated-beads were placed on the magnet for 3 min and supernatant was removed. hEGF-coated-beads were washed again as previously done. Coated-beads were finally resuspended with 550 µL of ice-cold PBSA.

In the meantime, 3x10^{10} cells were pelleted at 3750 rpm for 10 min at 4°C (at least 10-fold oversampling should be pelleted to reduce the possibility of losing individual clones). Supernatant was removed and cells were resuspended in 30 mL ice-cold PBSA. Yeasts were pelleted again at 3750 rpm for 5 min at 4°C. Yeast cells were resuspended to a final volume of 10 mL in PBSA and splited in 2 mL microcentrifuge tubes (1 mL/tube). 5 µL of resuspended cells for each library were combined with 495 µL of ice-cold PBSA and kept at 4ºC for full length protein expression analysis. 50 µL of hEGF-coated magnetic beads were added to each microcentrifuge tube containing yeast cells. Yeast-bead complex was incubated for 2h at 4ºC with tumbling. After 2h incubation samples were placed on the magnet for 3 min and supernatant containing non-binders was removed. Magnetic beads containing target binders were washed with PBSA as previously done. Yeast-bead complex was resuspended in SD-CAA media (1 mL/tube) and inoculated into 60 mL SD-CAA media. 1 µL of this culture was diluted with SD-CAA media to a final volume of 200 µL, and 50 µL were plated on SD-CAA plates and incubate at 30ºC. Finally cells were grown at 30ºC with shaking (250 rpm) for 48h.

Magnetic beads removal

After overnight growing, yeast cell cultures were pelleted at 3750 rpm for 1 min at 4ºC (at least 10 fold library viability should be pelleted). Yeast cells were resuspended in SD-CAA media to a final concentration of 1x10^9 cells/mL and splited in 2 mL microcentrifuge tubes (1 mL/tube). Samples were placed on the magnet for 3 min and then supernatant was transferred to a new microcentrifuge tube. This process was repeated twice. Finally supernatant was inoculated into 70 mL SD-CAA media and cells were grown overnight at 30ºC with shaking (250 rpm). Cells were passaged and induced as previously done.

Full-length protein expression confirmation

Samples kept at 4ºC were used to confirm full length protein expression. Antibodies were purchased from Invitrogen. Labeling was done as follows:

	Sample 1 (100 µL)	**Sample 2 (100 µL)**
[Cells]	2x10^6 cells/mL	2x10^6 cells/mL
Chicken anti-HA IgY	1:200 (0.5 µL)	-----
Chicken anti-c-myc IgY	-----	1:200 (0.5 µL)
Alexa Fluor 488-goat anti-chicken IgG	1:200 (0.5 µL)	1:200 (0.5 µL)

Pelleted yeast cells were combined with primary antibody (chicken anti-HA IgY or chicken anti-c-myc IgY) and ice-cold PBSA was added to a final volume of 100 µL. Samples were incubated for 1.5h at 4ºC with tumbling. Cells were pelleted by centrifugation at 14000 rpm at 4ºC for 30s. Primary antibody stained cells were washed with 1 mL of ice-cold PBSA and then pelleted by centrifugation at 14000 rpm at 4ºC for 30s. Yeast cells were then combined with the secondary antibody (Alexa Fluor 488-goat anti-chicken IgG) in a final volume of 100 µL ice-cold PBSA. Cells were incubated for 30 min at 4ºC with tumbling. Then samples were pelleted by centrifugation at 14000 rpm at 4^{o}C for 30s. Cell pellets were kept on ice until analysis. After labeling we ended up with four samples (two per each library) to be analyzed. Cell pellets were resuspended in 1 mL of ice-cold PBSA and labeled samples were analyzed by a BD Accuri™ C6 Plus instrument (fom BD Bioscience). Data was analyzed with FlowJo software.

Colony forming units counting

After cells were grown in SD-CAA plates, colony forming units (CFUs) were counted by visual inspection. The CFUs/mL were calculated using **Equation 3**.

$$\frac{CFUs}{mL} = \frac{number\ of\ colonies}{(dilution\ factor)(volume\ plated\ in\ mL)} \qquad \textbf{Equation 4}$$

Cells viability was calculated by multiplying the CFUs by the final culture volume.

Second round of magnetic beads selection

Negative selections to eliminate magnetic beads or streptavidin binders

This time, 2.4x10^{7} Dynabeads® Biotin Binder (Invitrogen) were added to a 2 mL microcentrifuge tube containing 1 mL of ice-cold PBSA and they were incubated for 5 min at 4ºC with tumbling. Microcentrifuge tubes containing magnetic beads were placed on the magnet for 3 min and supernatant was removed being careful not to touch the beads. Beads washes were performed as previously done. After supernatant removal beads were resuspended with 330 µL of ice-cold PBSA.

In parallel, 2x10^{9} cells were pelleted at 3750 rpm for 1 min at 4^{o}C. Supernatant was removed and cells were resuspended in 1 mL ice-cold PBSA. Yeasts were pelleted at 14000 rpm for 5 min at 4ºC. Yeast cells were resuspended in 1 mL ice-cold PBSA. Cells were pelleted again at 14000 rpm for 5 min at 4ºC and supernatant was removed. Yeast cells washes were performed as previously done. Yeast cells were resuspended in ice-cold PBSA to a final volume of 1 mL. Again, 5 µL of resuspended cells for each library were combined with 495 µL of ice-cold PBSA and kept at 4ºC for FACS analysis (full-length protein expression).

Yeast cells were combined with 50 µL washed uncoated magnetic beads. Yeast-magnetic beads complex was incubated for 1.5h at 4ºC with tumbling. After the incubation period, yeasts-beads complex were placed on the magnet for 3 min and supernatant was transferred to a new 2 mL micocentrifuge tube. Again, 50 µL of non-coated previously washed magnetic beads were added to the supernatant and incubated for 1.5h at 4ºC with tumbling. Three rounds of negative selection were performed to remove magnetic beads binders.

Positive selection to enrich for target binders

For positive selection, $8x10^6$ beads were washed as previously described. To the washed magnetic beads 66 pmols of biotinylated-hEGF were addend and also PBSA to a final volume of 200 µL. Beads were incubated for 2h at 4ºC with tumbling. After 2h incubation we placed the hEGF-coated beads on the magnet for 3 min and removed the supernatant. Coated beads were washed as previously described and resuspended in 120 µL ice-cold PBSA. Coated beads were kept in ice. 50 µL hEGF-coated magnetic beads were combined with the supernant obtained from the third negative selection. Beads-yeasts complex were incubated for 2h at 4ºC with tumbling. After incubation, beads were placed on the magnet for 3 min and supernatant containing non-binders was removed. Beads were washed with PBSA as previously done. Yeasts-beads complex were resuspended in 1 mL SD-CAA media and inoculated into 30 mL SD-CAA media. 100 µL of yeast culture was diluted with SD-CAA media to a final volume of 1 mL and 20 µL of this mixture were plated on SD-CAA plates and incubate at 30°C. Cells were grown at 30°C with shaking (250 rpm) for 48h. Cells were passaged and grown at 30°C with shaking until OD600= 2-5 and then protein expression was induced by incubating overnight at 20°C with shaking.

Flow cytometry assays

For flow cytometry sorting, cells were doubly labeled with the anti-c-myc tag and with the biotinylated target. For this purpose, two different strategies were used.

	Sample 1 (200 µL)	Sample 2 (200 µL)
[Cells]	2 x 10^7 cells/mL	2 x 10^7 cells/mL
Chicken anti-HA IgY	1:200 (1 µL)	1:200 (1 µL)
Biotinylated hEGF	1 µM no-preloaded	250 nM preloaded
Alexa Fluor 488-goat anti-chicken IgG	1:200 (1 µL)	1:200 (1 µL)
Streptavidin-Alexa Fluor 647	1:200 (1 µL)	-----

Sample 1 (no-preloaded)

Pelleted yeast cells were combined with the primary antibody (chicken anti-c-myc IgY) and the biotinyalted hEGF and ice-cold PBSA was added to a final volume of 200 µL. Samples were incubated for 1.5h at 4°C with tumbling. Cells were pelleted by centrifugation at 14000 rpm at 4°C for 30s. Primary antibody stained cells were washed with 1 mL of ice-cold PBSA and then pelleted by centrifugation at 14000 rpm at 4°C for 30s. Stained cells were then combined with the secondary antibodies (Alexa Fluor 488-goat anti-chicken IgG and Streptavidin-Alexa Fluor 647) in a final volume of 200 µL ice-cold PBSA. Cells were incubated for 30 min at 4°C with tumbling. Then samples were pelleted by centrifugation at 14000 rpm at 4°C for 30s. Cell pellets were kept on ice until analysis. After labeling we ended up with two samples (one per each library) to be analyzed.

Sample 2 (preloaded)

For the hEGF-preloaded strategy we first incubated Streptavidin-Alexa Fluor 647 with biotinylated-hEGF on ice for 30 min, shielded from light, at a streptavidin/hEGF molar ratio of 1:4 in as a small volume as possible. This sample was then mixed with yeast cells and PBSA was added to a final volume of 200 µL. Labeling with chicken anti-HA IgY and washes were performed as previously done. The secondary labeling was done just adding Alexa Fluor 488-goat anti-chicken IgG as previously done. After labeling we ended up with two samples (one per each library) to be analyzed.

Sorting

Cell pellets were resuspended in 1 mL of ice-cold PBSA immediately before analysis. Samples were analyzed using the FACSAria™ cell sorter instrument. Yeast cells that were double positive for the c-myc and target were collected into 5 mL glass tubes and kept on ice. After sorting, collected yeast cells were inoculated into 2 mL of SD-CAA media and grown overnight at 30°C with shaking. Some yeast cells need more than 48h to completely grow (OD_{600}= 6-8). Cells were passaged and induced as previously done.

Data was analyzed using the FlowJo software. Briefly living cells were analyzed by plotting the scattering light (SSC) vs the forward scattering light (FSC) and gating those that presented the correct morphology. Doublets and cluster discrimination was done by plotting FSC-W vs FSC-H, where W is the pulse width and H is the pulse intensity, and gating single cells. Final analysis was done by plotting the binding event (Alexa Fluor-647 emission accounting for hEGF binders) vs the protein displaying event (Alexa Fluor-488 emission accounting for c-myc tag display).

Identification of individual clones

After subsequent FACS sorting rounds a selected population of target binders was collected and grown as previously done. Newly passaged non-induced cells were used for yeast DNA plasmid extraction.

Plasmid DNA form selected yeast cells was extracted using the Zymoprep™ Yeast Plasmid Miniprep II (Zymo Research). Cell density was measured and $1x10^8$ cells were pelleted at 300xg for 1 min. Supernatant was removed and 200 mL of Solution 1 was added to the pellet which was resuspended by gentle pipetting. 3 mL of Zymolyase mix were added and incubated at 37^oC for 30 min. 200 mL of Solution 2 was added with gentle mixing. Then 400 mL of Solution 3 was also added and gently mixed. Debris were pelleted at 12000 g for 8 min. Supernatant was removed and transferred to a new microcentrifuge tube. Remaining debris in the supernatant were pelleted at 12000 g for 5 min and the clear supernatant was transferred to a miniprep column (Qiagen). The column was placed on a microcentrifuge tube and it centrifuged at 12000 g for 1 min and flow through was discarded. Column was washed using 550 mL of miniprep wash buffer (Qiagen buffer PE). The column was centrifuged at 12000 g for another 1 min and flow through was discarded. Column was centrifuged again to remove any excess of wash buffer. Column was placed on a new microcentrifuge tube and DNA was eluted with 40 mL of elution buffer (Qiagen) by centrifuging at 12000 g for 1 min.

1 µL of extracted plasmid DNA was transformed into XL1-Blue supercompetent cells. The entire reaction was plated on LB + ampicillin plates and grown at 37^oC for at least 16 h. Individual colonies were inoculated into 5 mL of LB + ampicillin media and grown overnight at 37^oC with shaking at 250 rpm. Cultures were finally minipreped to isolate plasmid DNA. Plasmid DNA was sent for sequencing.

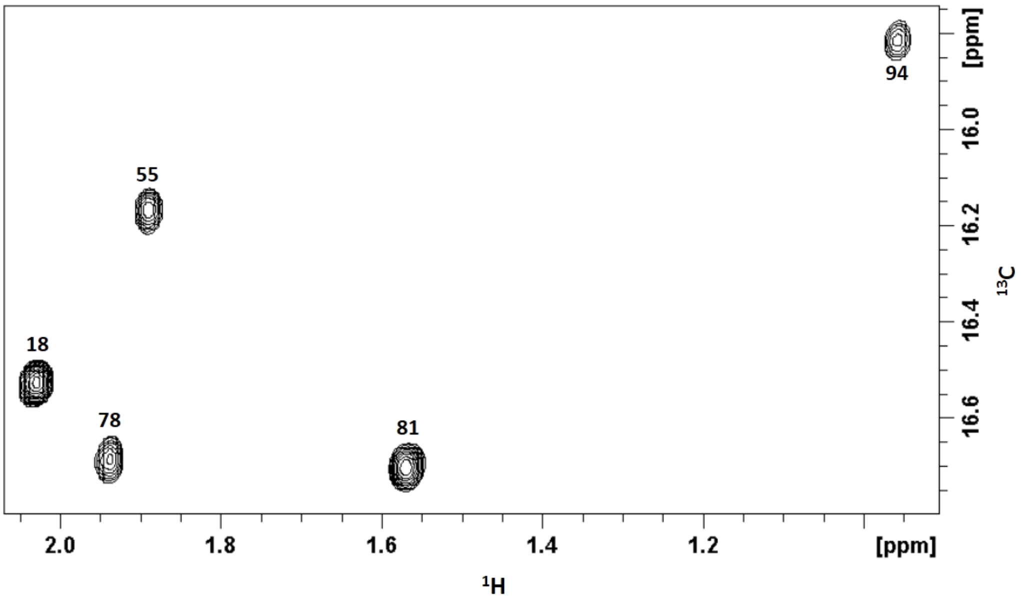
Product characterization

Protein characterization

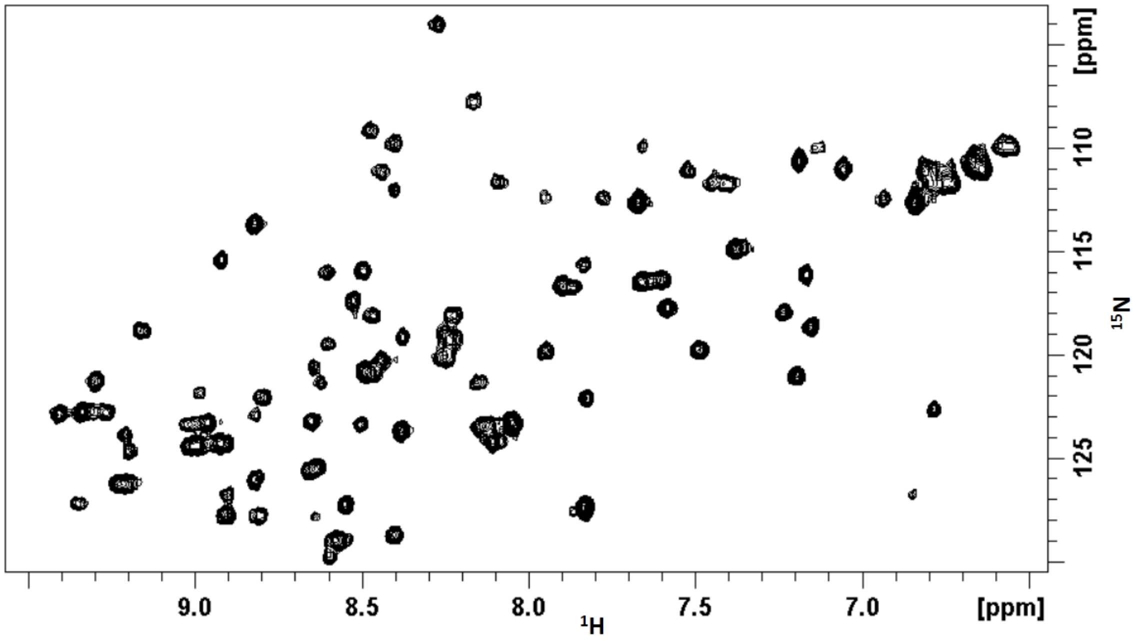
VEGF characterization

HHEVVKFMDVYQRSYCHPIETLVDIFQEYPDEIEYIFKPSCVPLMRCGGCCNDEGLECVPTEESNITMQ IMRIKPHQGQHIGEMSFLQHNKCECRPKKD

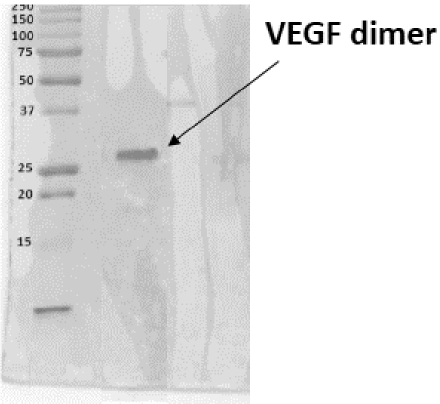
^{1}H-^{13}C-HSQC spectrum of 50 µM VEGF in 25 mM sodium phosphate, 50 mM NaCl buffer, pH 7.4 at 318 K.

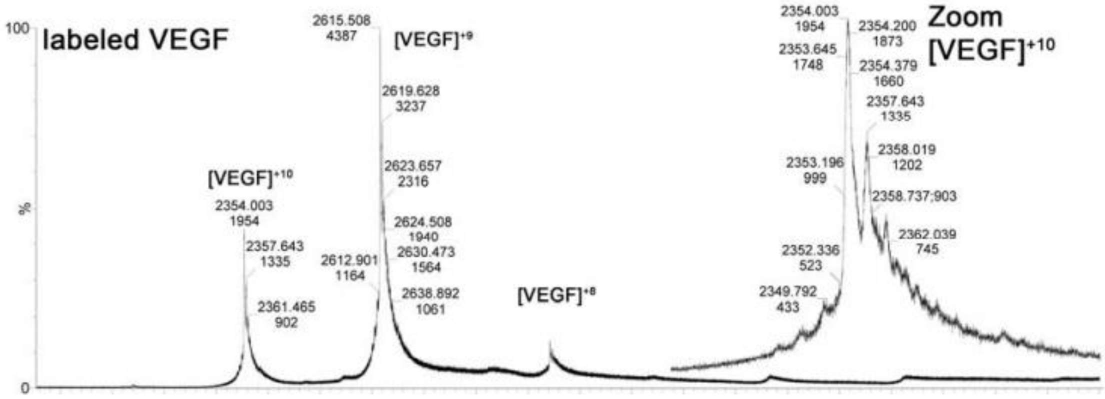


^{1}H-^{15}N-HSQC spectrum of 50 µM VEGF in 25 mM sodium phosphate, 50 mM NaCl buffer, pH 7.4 at 318 K.



Characterization of ^{13}C-methyl-methionine, ^{15}N-uniformly labeled VEGF$_{11\text{-}109}$ by ESI time-of-flight and SDS-PAGE gel electrophoresis

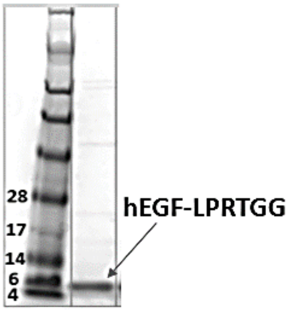


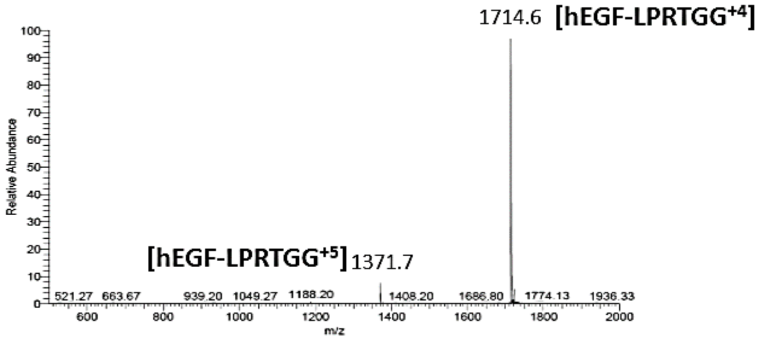


hEGF-LPRTGG characterization

NSDSECPLSHDGYCLHDGVCMYIEALDKYACNCVVGYIGERCQYRDLKWWELRLPRTGGG

Characterization of hEGF-LPRTGG by ESI and SDS-PAGE gel electrophoresis





DNA and protein sequence of Sso7d-based M11.1.3 variants:

DNA sequence of M11.1.3:

ATGGGTCATCACCATCATCATCACGGAGGTGCAACCGTGAAATCCACATACCGAGGCGAAGAA
AAACAGGTGGATATTAGCAAAATCAAGTGGGTG**ATC**CGT**TGGGGC**CAGCATCTTGCTTTTAAA
TATGATGAAGGTGGTGGTGCCGCAGGTTATGGTTGGGTGAGCGAAAAGGATGCACCGAAAGA
ACTGCTGCAGATGCTGGAAAAGCAATAA

Protein Sequence of M11.1.3:

MGHHHHHHGGATVKSTYRGEEKQVDISKIKWV**I**R**WG**QHLAFKYDEGGGAAGYGWVSEKDAPK
ELLQMLEKQ

DNA sequence of M11.1.3(I23C):

ATGGGTCATCACCATCATCATCACGGAGGTGCAACCGTGAAATCCACATACCGAGGCGAAGAA
AAACAGGTGGATATTAGCAAAATCAAGTGGGTG**TGC**CGTTGGGGCCAGCATCTTGCTTTTAAA
TATGATGAAGGTGGTGGTGCCGCAGGTTATGGTTGGGTGAGCGAAAAGGATGCACCGAAAGA
ACTGCTGCAGATGCTGGAAAAGCAATAA

Protein Sequence of M11.1.3(I23C):

MGHHHHHHGGATVKSTYRGEEKQVDISKIKWV**C**RWGQHLAFKYDEGGGAAGYGWVSEKDAPK
ELLQMLEKQ

DNA sequence of M11.1.3(W25C):

ATGGGTCATCACCATCATCATCACGGAGGTGCAACCGTGAAATCCACATACCGAGGCGAAGAA
AAACAGGTGGATATTAGCAAAATCAAGTGGGTGATCCGT**TGC**GGCCAGCATCTTGCTTTTAAA
TATGATGAAGGTGGTGGTGCCGCAGGTTATGGTTGGGTGAGCGAAAAGGATGCACCGAAAGA
ACTGCTGCAGATGCTGGAAAAGCAATAA

Protein Sequence of M11.1.3(W25C):

MGHHHHHHGGATVKSTYRGEEKQVDISKIKWVIR**C**GQHLAFKYDEGGGAAGYGWVSEKDAPKE
LLQMLEKQ

DNA sequence of M11.1.3(G26C):

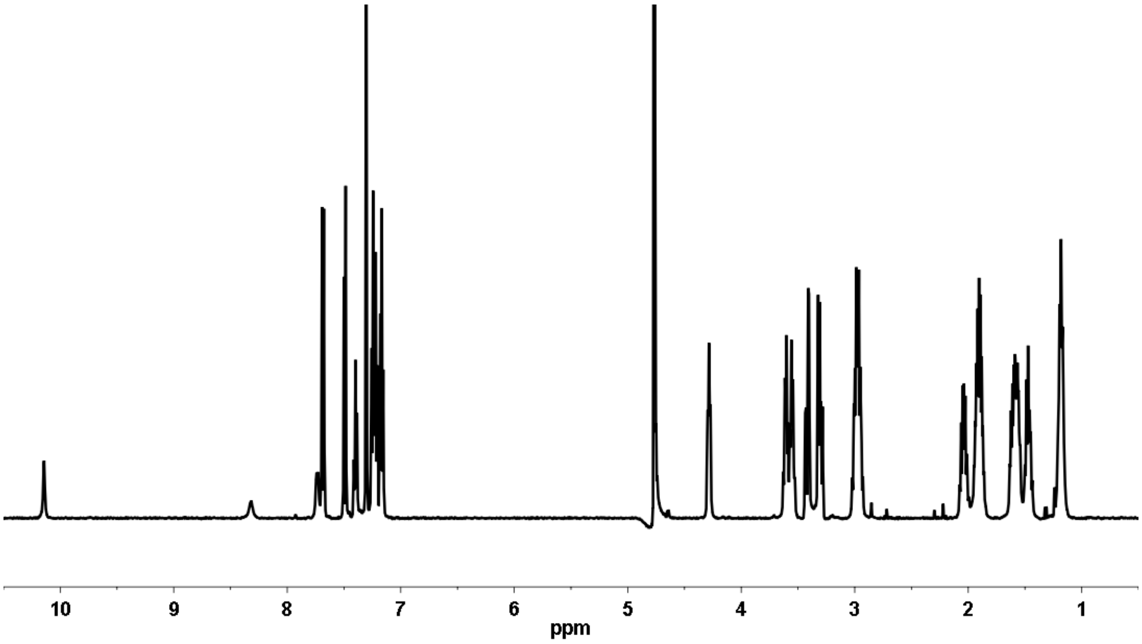
ATGGGTCATCACCATCATCATCACGGAGGTGCAACCGTGAAATCCACATACCGAGGCGAAGAAAAACAGGTGGATATTAGCAAAATCAAGTGGGTGATCCGTTGG**TGC**CAGCATCTTGCTTTTAAATATGATGAAGGTGGTGGTGCCGCAGGTTATGGTTGGGTGAGCGAAAAGGATGCACCGAAAGAACTGCTGCAGATGCTGGAAAAGCAATAA

Protein Sequence of M11.1.3(G26C):

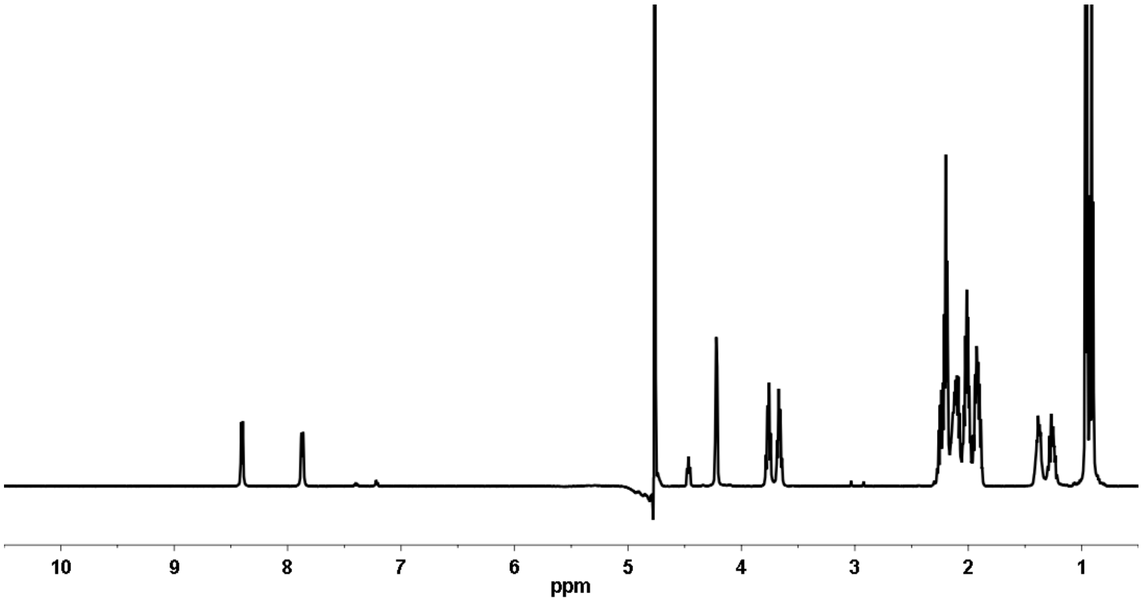
MGHHHHHHGGATVKSTYRGEEKQVDISKIKWVIRW**C**QHLAFKYDEGGGAAGYGWVSEKDAPKELLQMLEKQ

Peptide characterization

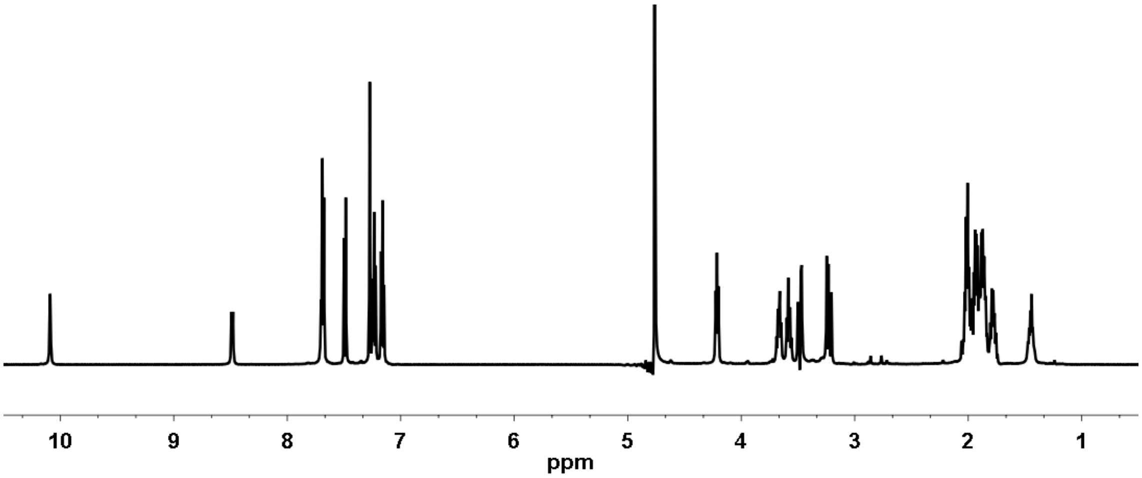
1D ^{1}H-NMR spectrum of peptide 3 c(RpWRpW)



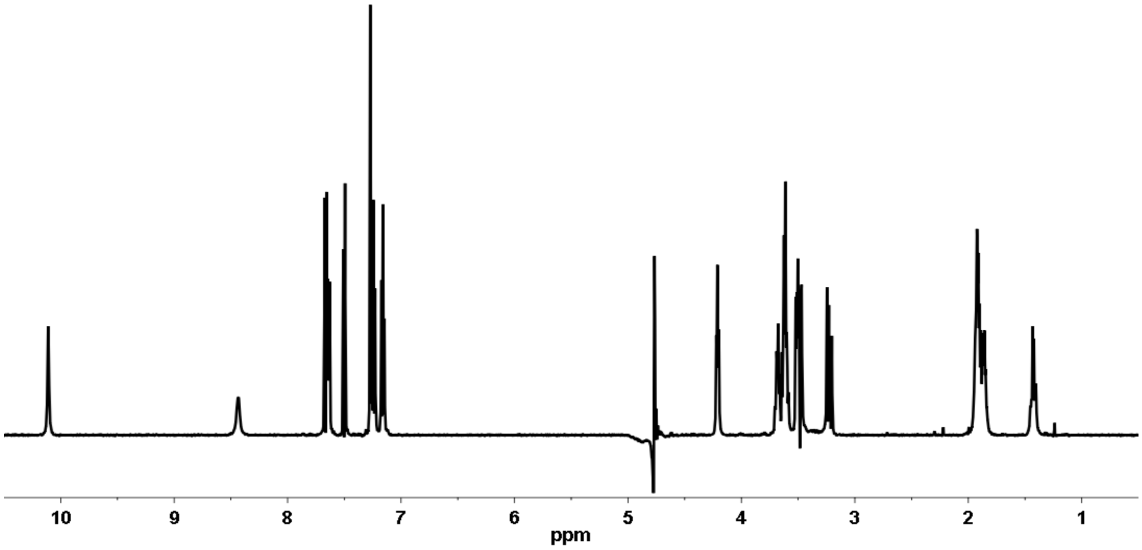
1D ^{1}H-NMR spectrum of peptide 6 c(EpIEpI)



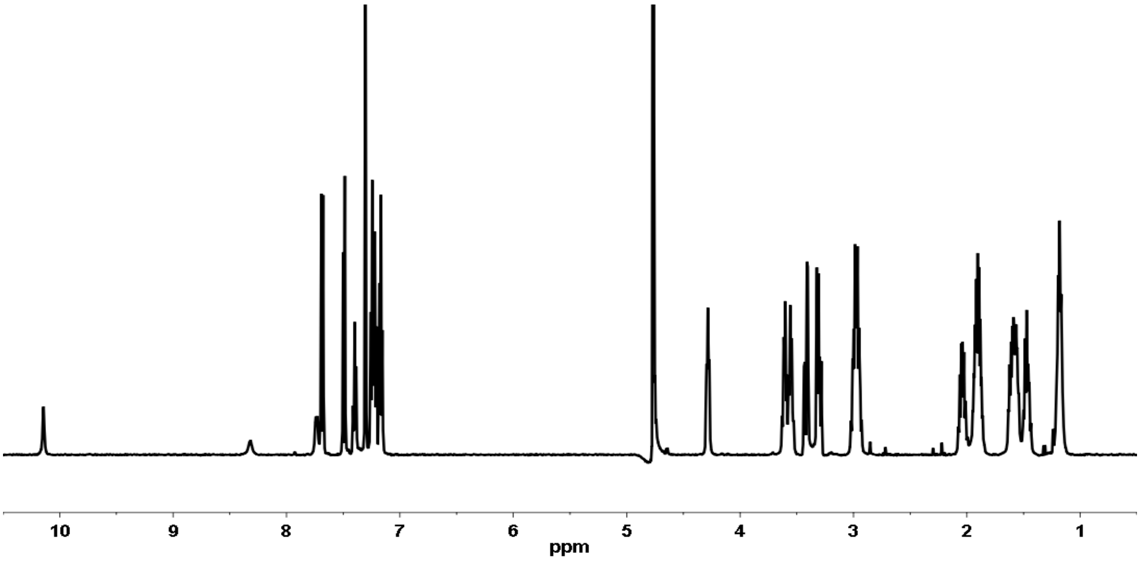
1D ^{1}H-NMR spectrum of peptide 7 c(EpWEpW)



1D ^{1}H-NMR spectrum of peptide 11 c(SpWSpW)



1D ^{1}H-NMR spectrum of peptide 3-P1 c(WpRWpR)



^{1}H-NMR chemical shifts for peptides 3, 6, 7, 11 and 3-P1

Residue	NH					Hα					Hβ				
	3	6	7	11	3-P1	3	6	7	11	3-P1	3	6	7	11	3-P1
Xaa	7.73	7.87	7.70	7.64	7.94	4.65	4.91	4.79	4.92	5.18	1.48	2.01	1.86	3.62	3.31
											1.58	1.91	1.77	3.51	3.26
D-Pro	-	-	-	-	-	4.28	4.47	4.21	4.21	4.27	2.05	2.25	1.93	1.91	2.03
											1.61	1.92	1.44	1.42	1.80
Yaa	8.32	8.40	8.48	8.43	8.39	4.65	4.22	4.63	4.65	4.05	3.42	2.13	3.48	3.49	1.57
											3.30		3.22	3.22	1.40

Residue	Other				
	3	6	7	11	3-P1
Xaa	γCH_2 1.19	γCH_2 2.20	γCH_2 2.01		H_2 7.32
	δCH2 3.00, 2.95				H_4 7.75
					H_5 7.21
					H_6 7.26
					H_7 7.50
					NH 10.15
D-Pro	γCH2 1.91	γCH_2 2.10, 2.00	γCH_2 1.92, 1,90	γCH2 1.91, 1.86	γCH_2 1.95, 1.67
	δCH2 3.61, 3.55	δCH_2 3.76, 3.66	δCH_2 3.66, 3.58	δCH2 3.68, 3.61	δCH_2 3.58, 3.36
Yaa	H_2 7,30	γCH_2 1.38, 1.26	H_2 7,27	H_2 7.27	γCH_2 1.29, 1.20
	H_4 7,68	γCH_3 0.96	H_4 7,69	H_4 7.66	δCH2 3.04, 3.00
	H_5 7,17	δCH_3 0.91	H_5 7,16	H_5 7.16	
	H_6 7,25		H_6 7,24	H_6 7.24	
	H7 7,49		H_7 7,49	H_7 7.49	
	NH 10,15		NH 10,10	NH 10,11	

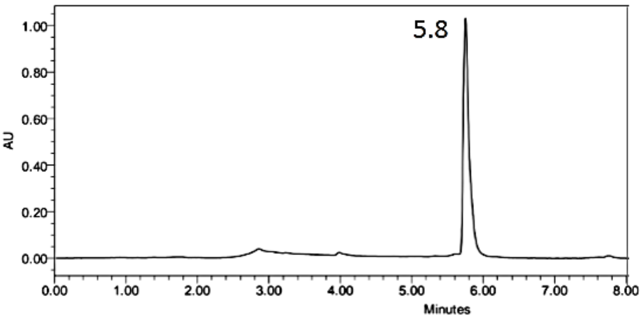
^{13}C-NMR chemical shifts for peptides 3, 6, 7, 11 and 3-P1

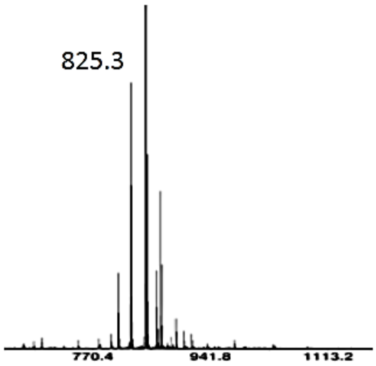
Residue	Cα					Cβ				
	3	6	7	11	3-P1	3	6	7	11	3-P1
Xaa	52.9	52.6	50.7	54.6	54.2	30.8	31.0	29.5	63.8	30.2
D-Pro	63.2	62.9	63.1	63.2	63.1	31.0	31.3	30.8	30.8	31.2
Yaa	57.2	61.6	57.0	57.1	56.4	28.8	38.2	28.8	28.7	29.8

Residue	Other				
	3	**6**	**7**	**11**	**3-P1**
Xaa	γCH_2 26.2	γCH_2 35.6	γCH_2 35.2		H_2 127.2
	δCH_2 43.2				H_4 121.0
					H_5 121.8
					H_6 124.6
					H_7 114.3
D-Pro	γCH_2 27.2	γCH_2 27.1		γCH2 27.2	γCH2 27.1
	δCH_2 50.4	δCH_2 50.2		δCH2 50.4	δCH2 50.3
Yaa	H_2 126.9	γCH_2 26.7	H_2 126.8	H_2 126.9	γCH_2 26.4
	H_4 121.0	γCH_3 17.7	H_4 120.9	H_4 120.9	δCH_2 43.0
	H_5 121.8	δCH_3 13.9	H_5 121.9	H_5 121.8	
	H_6 124.6		H_6 124.5	H_6 124.4	
	H_7 114.3		H_7 114.4	H_7 114.5	

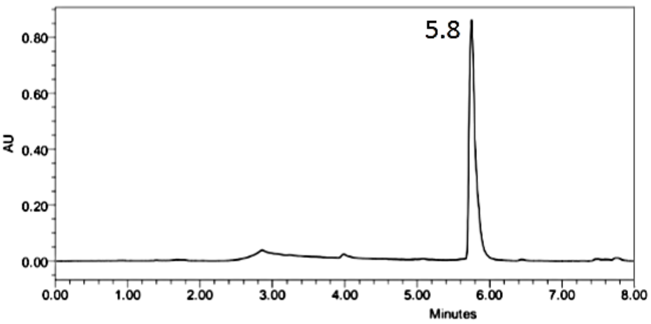
Cyclic peptides characterization (HPLC + MALDI)

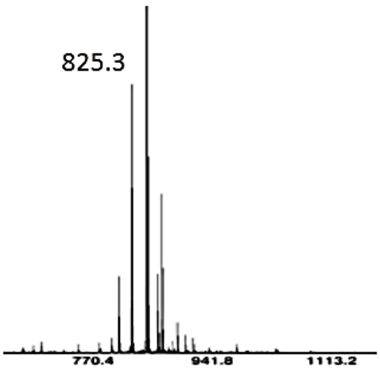
Peptide 7 c(EpWEpE)



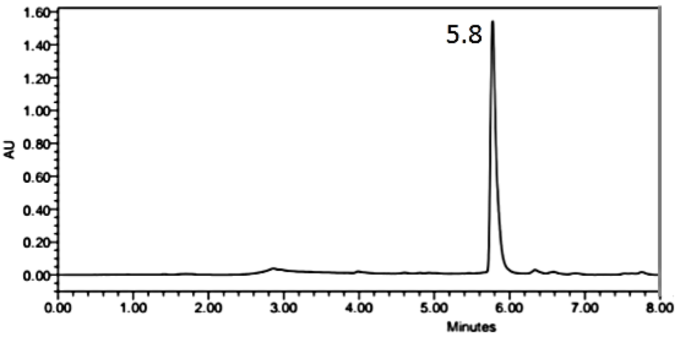


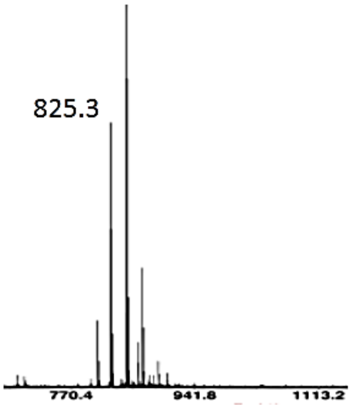
Peptide 7-C01 c(ePwePw)



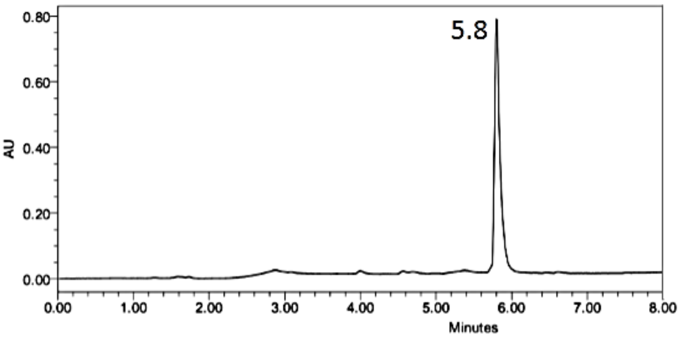


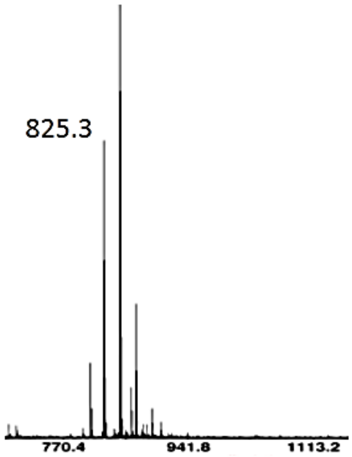
Peptide 7-C02 c(EpwEpw)



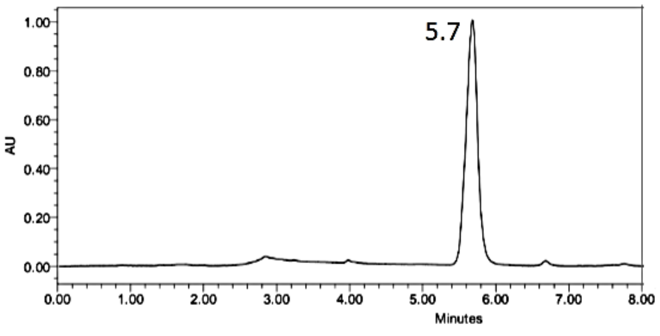


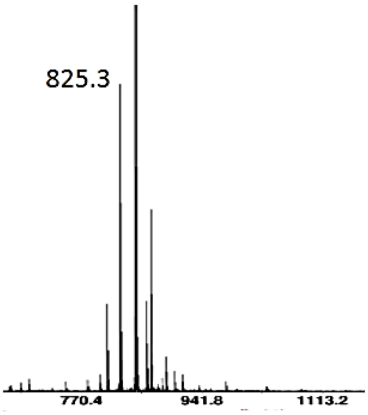
Peptide 7-C03 c(ePWePW)



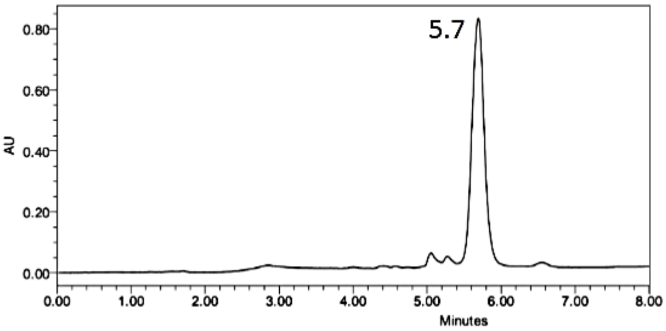


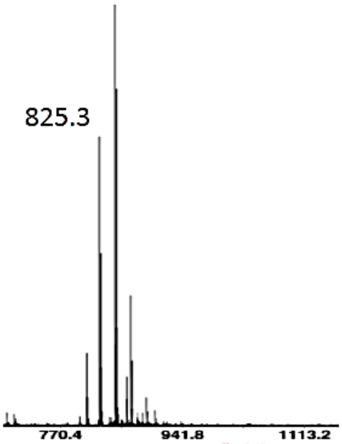
Peptide 7-C04 c(ePWePW)



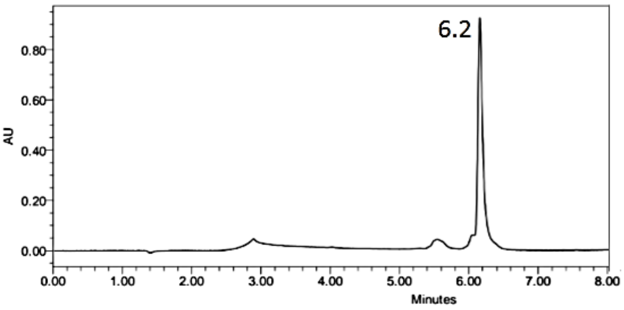


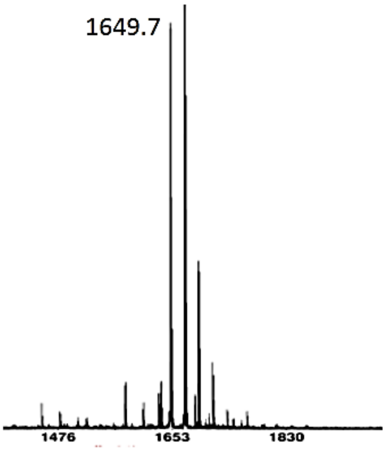
Peptide 7-C05 c(EPwEPw)



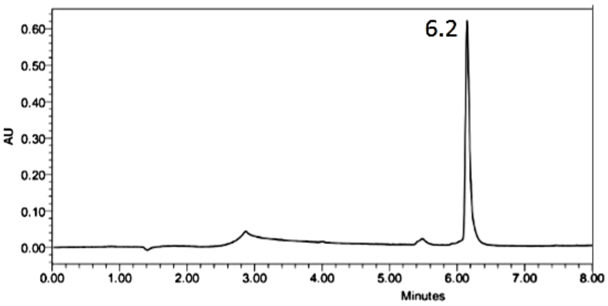


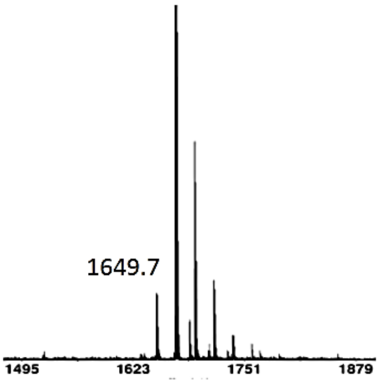
Peptide 7-C06D c(EPWEPW)$_2$



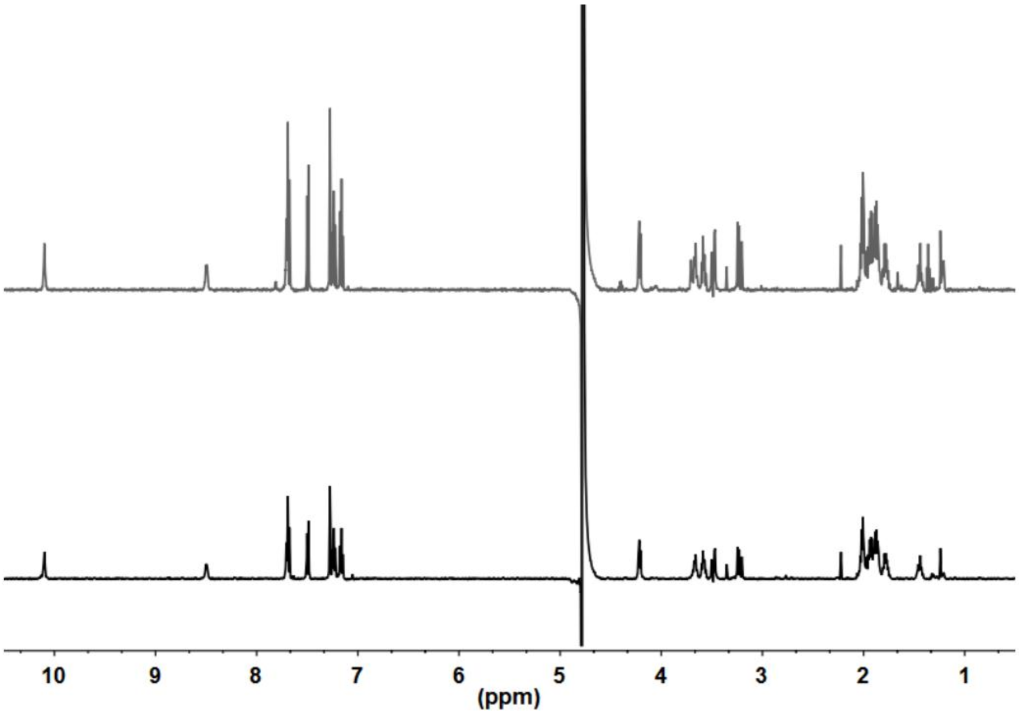


Peptide 7-C07D c(epwepw)$_2$

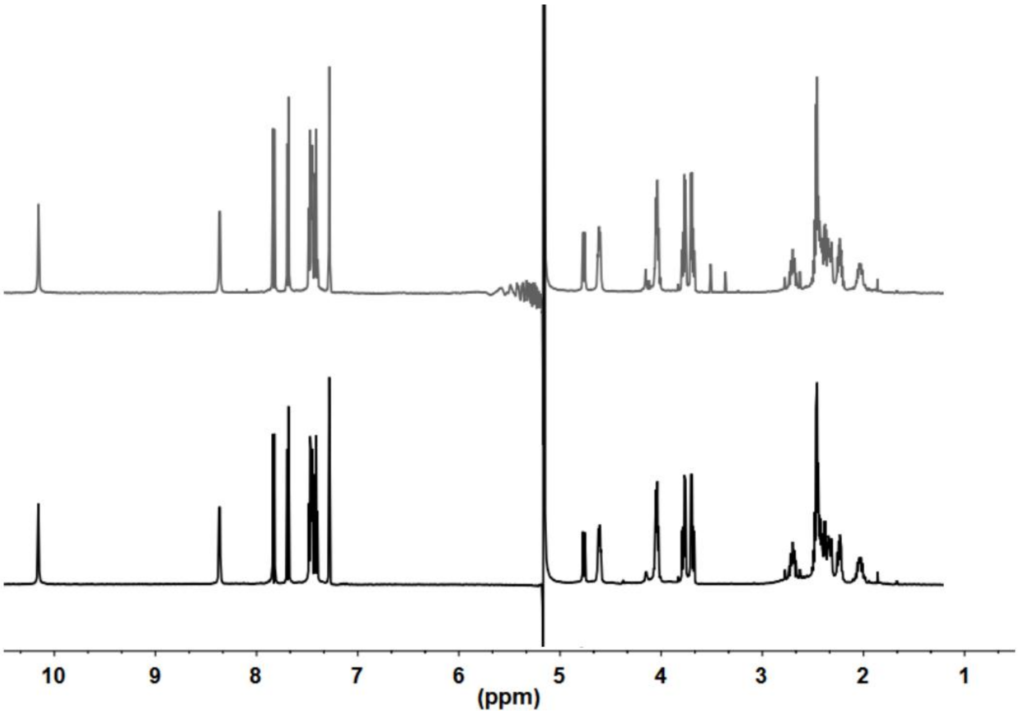




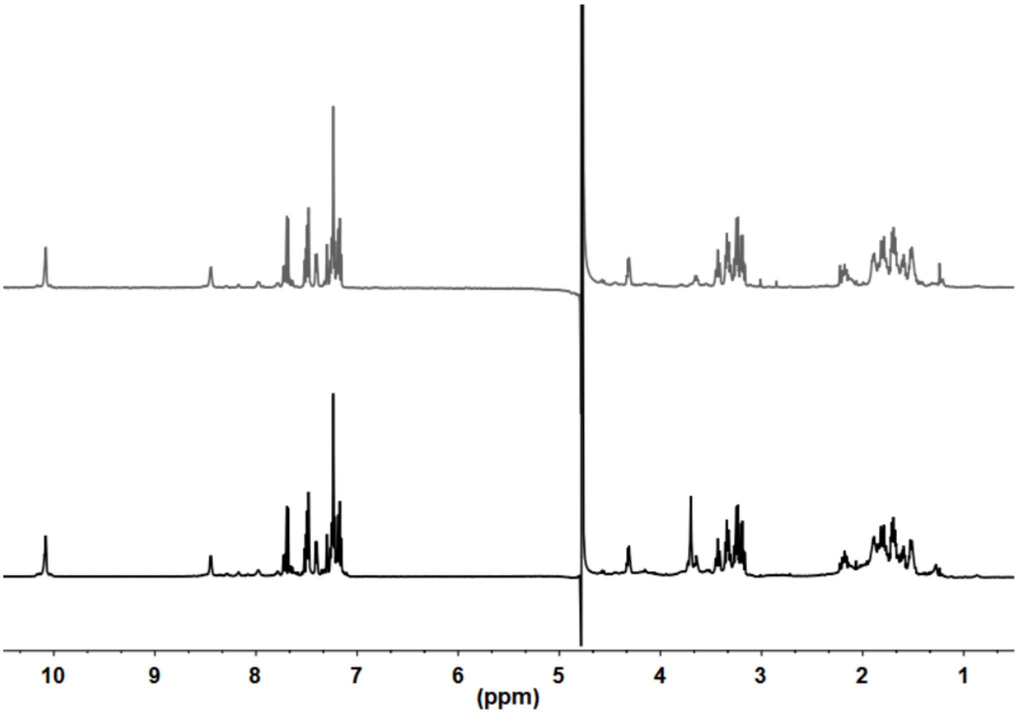
1D ^{1}H-NMR spectra for peptide 7 (grey) and peptide 7-C01 (black)



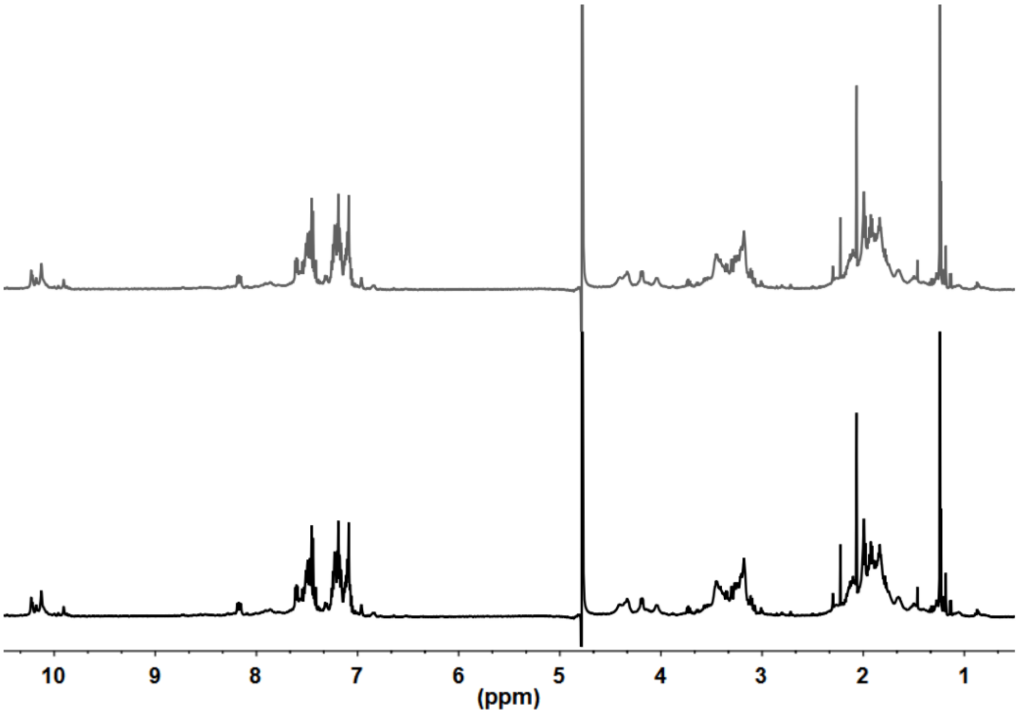
1D ^{1}H-NMR spectra for peptide 7-C02 (grey) and peptide 7-C03 (black)



1D ^{1}H-NMR spectra for peptide 7-C04 (grey) and peptide 7-C05 (black)



1D ^{1}H-NMR spectra for peptide 7-C06D (grey) and peptide 7-C07D (black)



^{1}H-NMR chemical shifts for peptides 7, 7-C02 and 7-C04

Residue	NH				Hα				Hβ			
	7	7-C02	7-C04		7	7-C02	7-C04		7	7-C02	7-C04	
			cc	tt			cc	tt			cc	tt
Xaa	7.70	8.20	7.40	8.46	4.79	4.18	4.31	4.78	1.86	1.79	1.59	1.85
									1.77	1.63	1.51	1.64
D-Pro	-	-	-	-	4.21	4.35	4.56	4.42	1.93	2.14	2.18	2.10
									1.44	1.75	1.70	1.71
Yaa	8.48	7.24	8.45	7.98	4.63	4.73	4.59	4.71	3.48	3.29	3.24	3.33
									3.22	3.21	3.19	3.23

Residue	Other			
	7	7-C02	7-C04	
			cc	tt
Xaa	γCH$_2$ 2,01	γCH$_2$ 1,88	γCH2 1.81, 1.70	γCH2 1.91, 1.82
D-Pro	γCH$_2$ 1.92, 1,90	γCH$_2$ 1.85, 1,42	γCH2 1.78, 1.51	γCH2 1.89
	δCH$_2$ 3,66, 3,58	δCH$_2$ 3,58	δCH2 3.43, 3.34	δCH2 3.64, 3.28
Yaa	H$_2$ 7,27	H$_2$ 7,05	H$_2$ 7.23	H$_2$ 7.29
	H$_4$ 7,69	H$_4$ 7,64	H$_4$ 7.69	H$_4$ 7.71
	H$_5$ 7,16	H$_5$ 7,19	H$_5$ 7.17	H$_5$ 7.19
	H$_6$ 7,24	H$_6$ 7,24	H$_6$ 7.24	H$_6$ 7.25
	H$_7$ 7,49	H7 7,49	H$_7$ 7.49	H$_7$ 7.51
	NH 10,10	NH 10,13	NH 10,08	NH 10,09

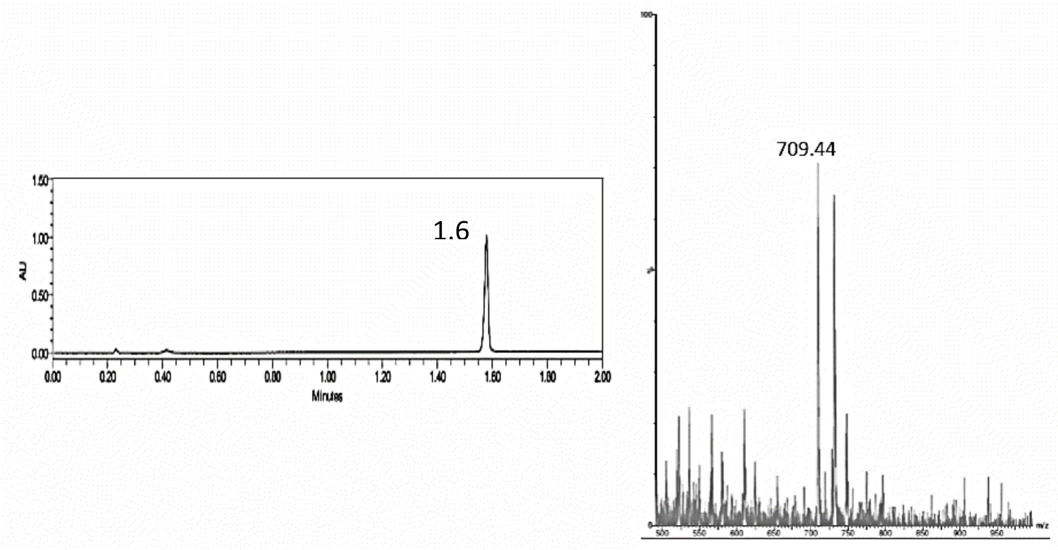
^{13}C-NMR chemical shifts for peptides 7, 7-C02 and 7-C04

Residue	Cα				Cβ			
	7	7-C02	7-C04		7	7-C02	7-C04	
			cc	tt			cc	tt
Xaa	52.9	54.9	53.6	53.0	30.5	28.6	31.2	31.0
D-Pro	63.2	63.7	62.8	64.2	30.7	32.0	34.4	30.8
Yaa	57.1	56.3	56.9	57.4	28.7	30.8	30.0	30.0

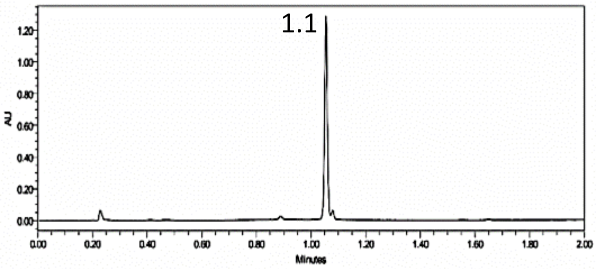
Residue	Other			
	7	7-C02	7-C04	
			cc	tt
Xaa	γCH_2 35.1	γCH_2 35.6	γCH2 34.8	γCH2 35.8
D-Pro	γCH_2 27.1	γCH_2 26.2	γCH2 24.2	γCH2 27.7
	δCH_2 50.0	δCH_2 50.3	δCH2 49.6	δCH2 50.5
Yaa	H_2 126.9	H_2 127.3	H_2 127.3	H_2 127.2
	H_4 121.1	H_4 121.0	H_4 121.1	H_4 121.2
	H_5 122.0	H_5 121.9	H_5 122.0	H_5 121.9
	H_6 124.6	H_6 124.6	H_6 124.6	H_6 124.5
	H_7 114.5	H7 114.6	H_7 114.6	H_7 114.6

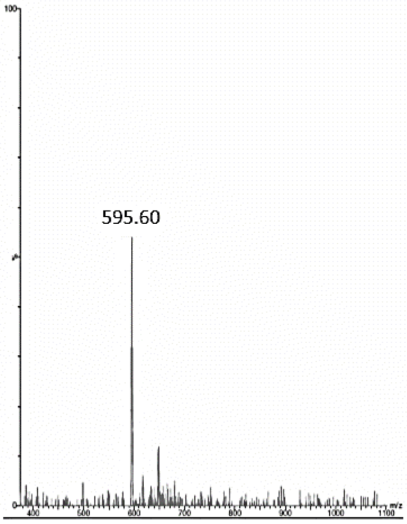
Cyclic peptides characterization (UPLC + MS)

Peptide 7-C04-A1 c(apWapW)

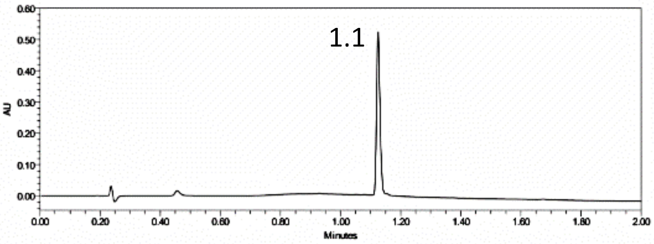


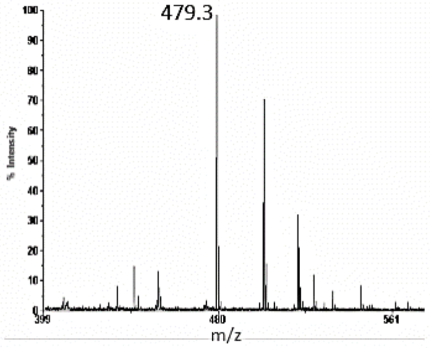
Peptide 7-C04-A2 c(epAepA)



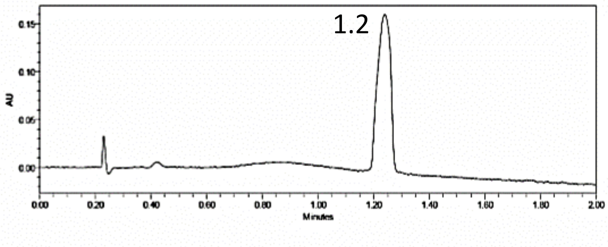


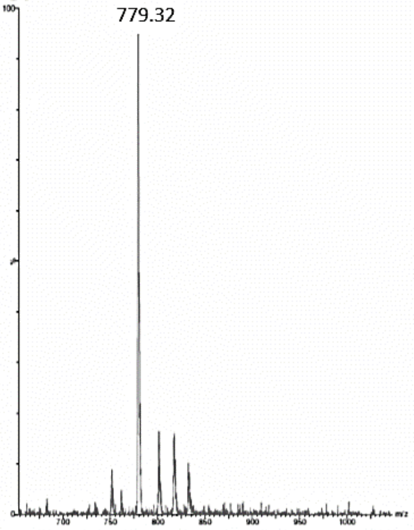
Peptide 7-C04-A3 c(apAapA)



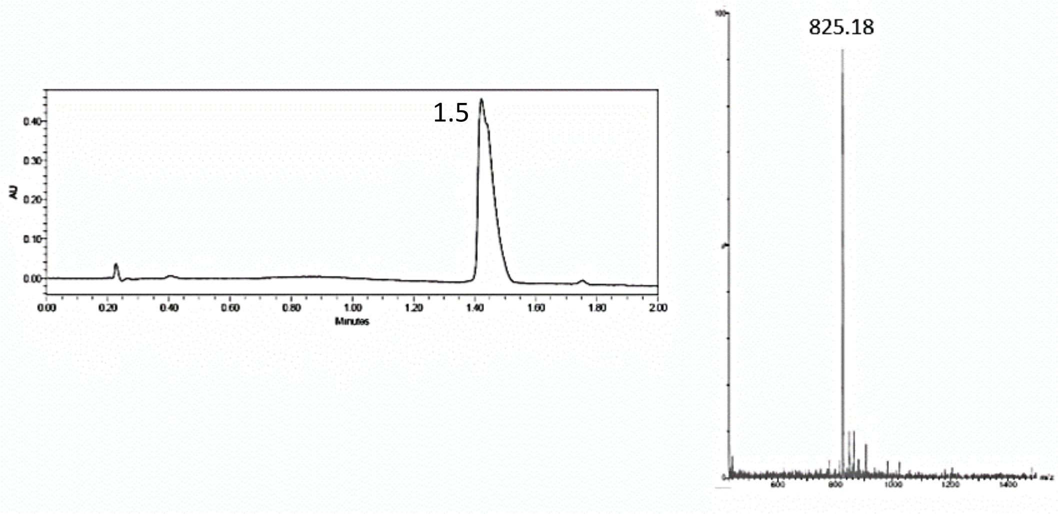


Peptide 7-C04-A4 c(epYepY)

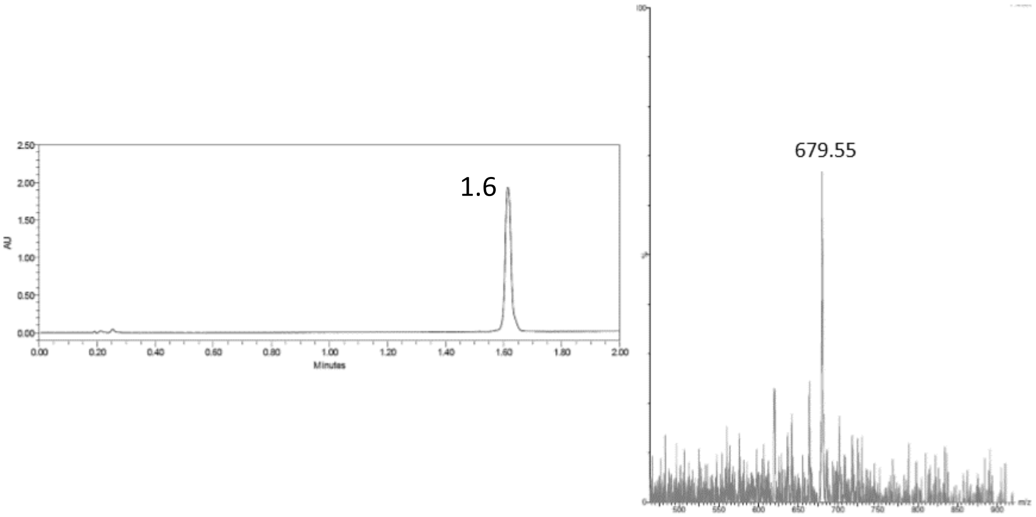




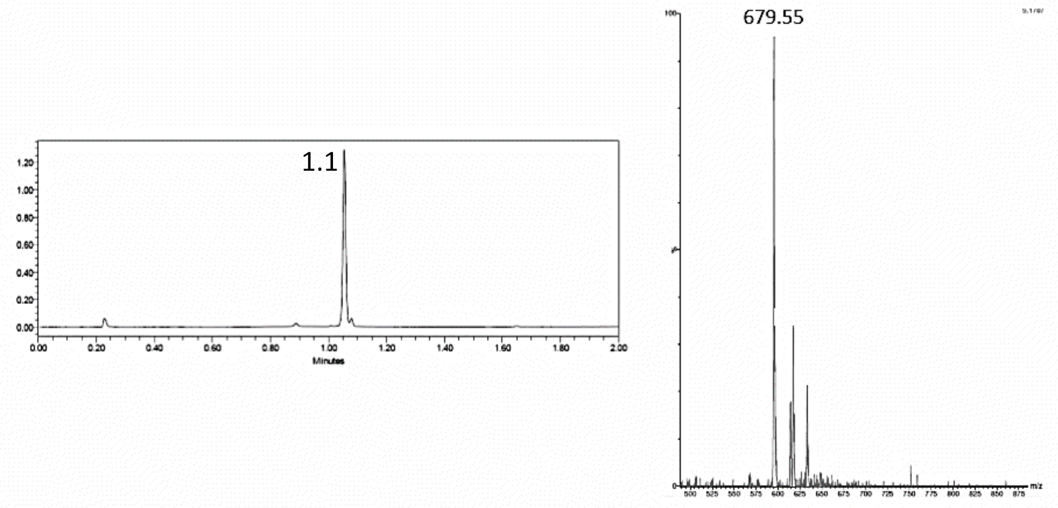
Peptide 7-C04-P1 c(wpEwpE)



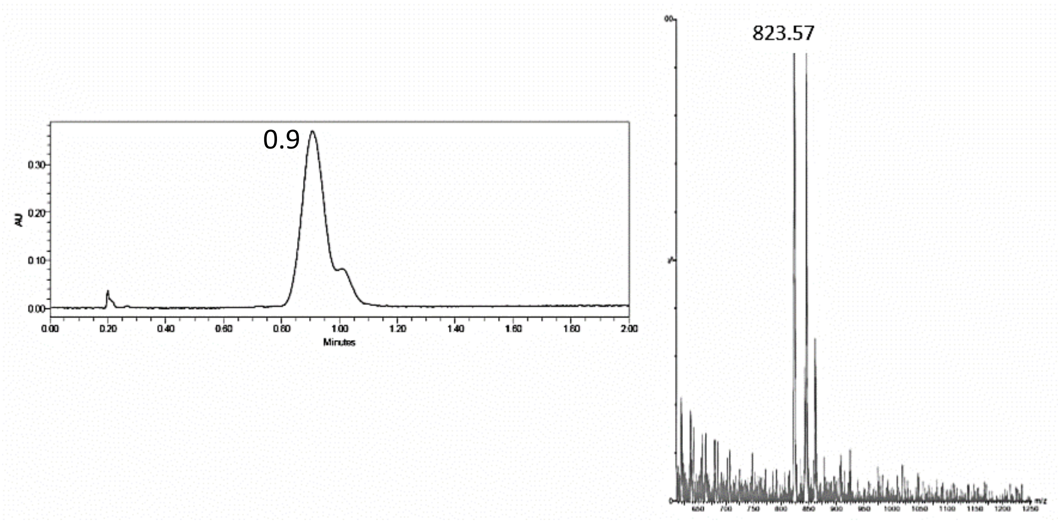
Peptide 7-C04-P1-A1 c(ipEipE)



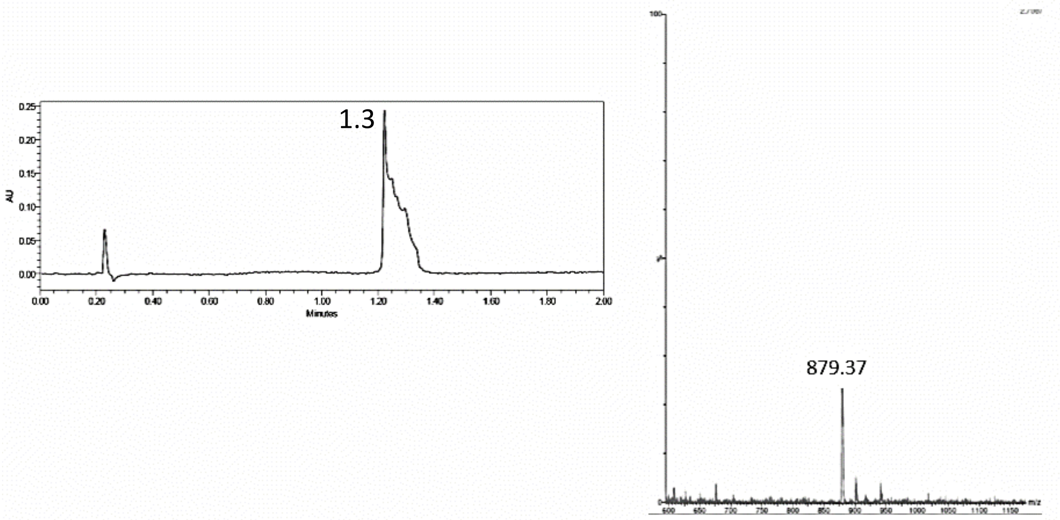
Peptide 7-C04-P1-A2 c(apEapE)



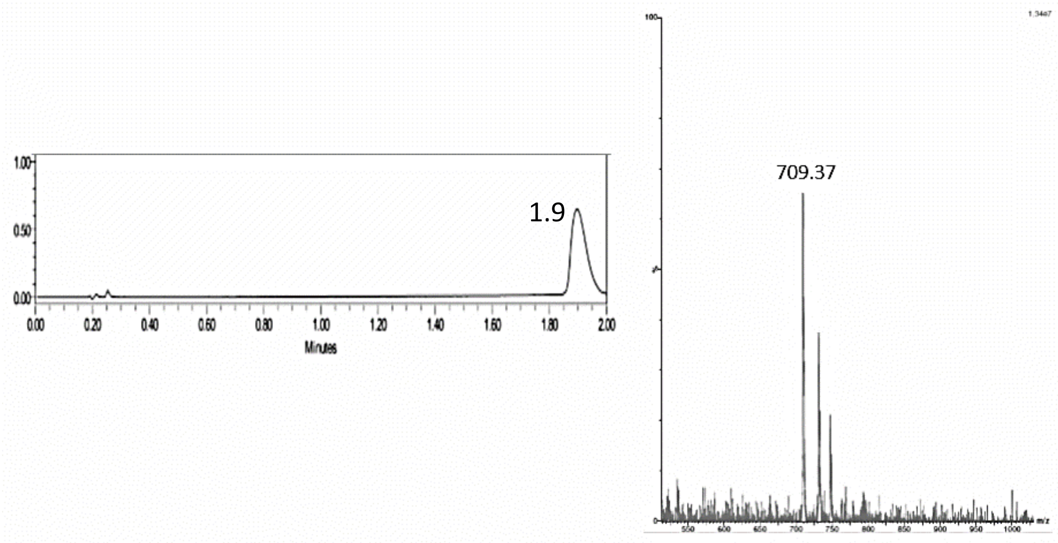
Peptide 7-C04-P1-A3 c(wpQwpQ)



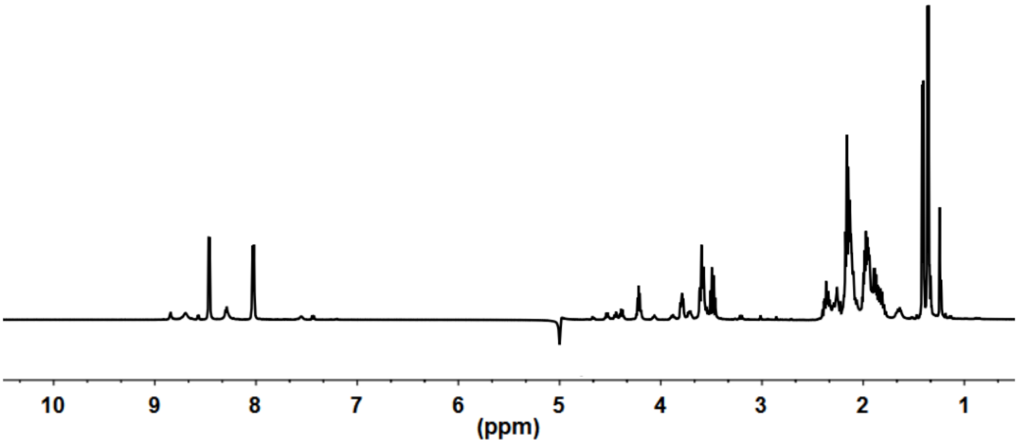
Peptide 7-C04-P1-A4 c(wpRwpR)



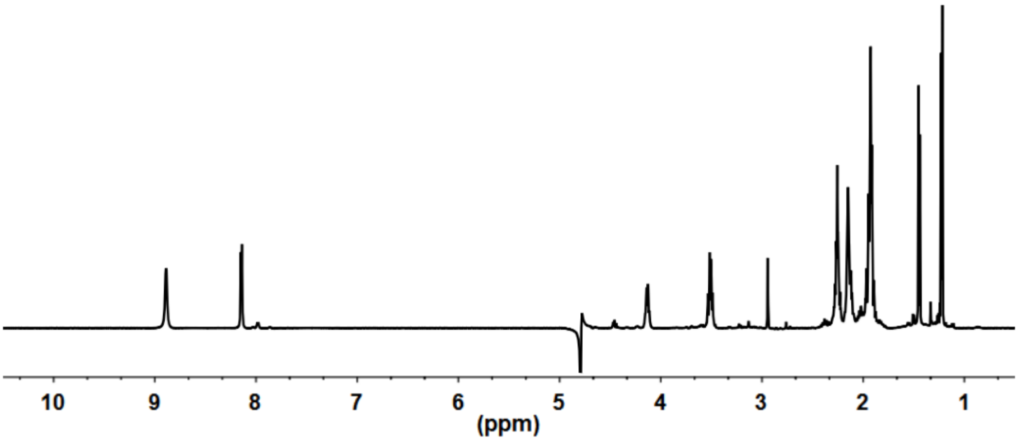
Peptide 7-C04-P1-A5 c(wpAwpA)



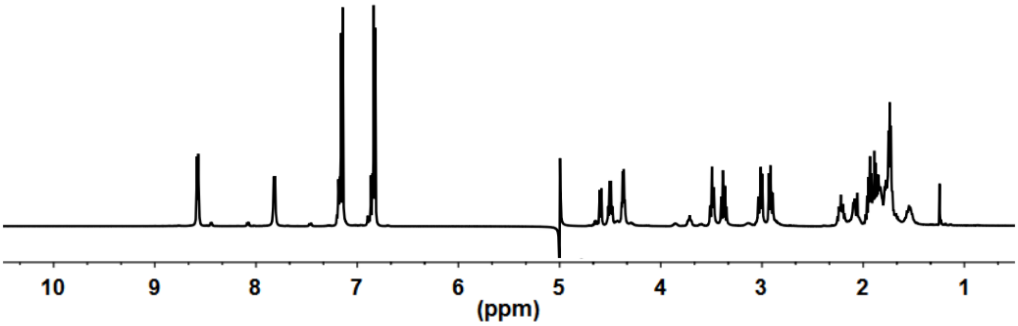
1D ^{1}H-NMR spectra for peptide 7-C04-A2 c(epAepA)



1D ^{1}H-NMR spectra for peptide 7-C04-A3 c(apAapA)



1D ^{1}H-NMR spectra for peptide 7-C04-A4 c(epYepY)



^{1}H-NMR chemical shifts for peptides 7-C04-A2, 7-C04-A3 and 7-C04-A4

	NH						Hα						Hβ					
Residue	7-C04-A2		7-04-A3		7-C04-A4		7-C04-A2		7-04-A3		7-C04-A4		7-C04-A2		7-04-A3		7-C04-A4	
	cc	tt	tt	cc	cc	tt	cc	tt	tt	cc	cc	tt	cc	tt	tt	cc	cc	tt
D-Xaa	7.90	8.31	8.12	7.96	7.81	8.07	4.53	4.70	4.65	4.40	4.36	4.70	1.97	1.67	1.22	1.26	1.73	1.93
													1.89					1.88
D-Pro	-	-	-	-	-	-	4.65	4.44	4.45	4.57	-	4.42	2.36	2.25	2.24	2.37	2.21	2.13
													2.10	1.94	1.92	2.12	1.76	1.68
Yaa	8.32	8.47	8.87	8.35	8.57	7.45	4.22	4.41	4.12	4.22	4.50	4.52	1.36	1.41	1.44	1.33	3.00	3.02
																	2.90	2.85

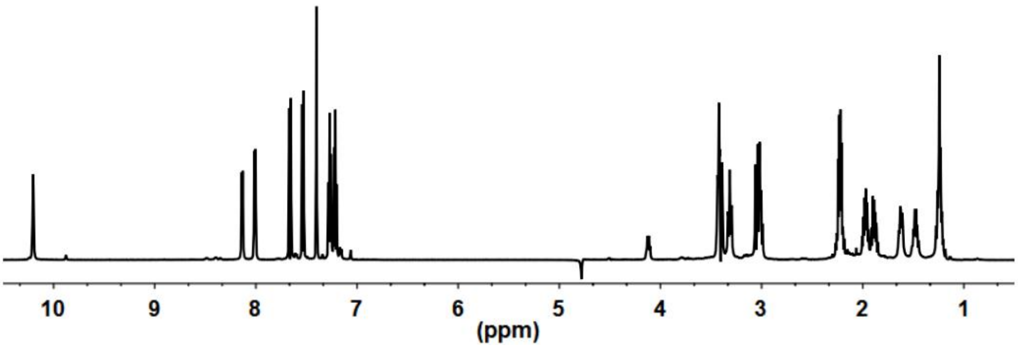
	Other					
Residue	7-C04-A2		7-04-A3		7-C04-A4	
	cc	tt	tt	cc	cc	tt
D-Xaa	γCH$_2$ 2.15	γCH$_2$ 2.14			γCH$_2$ 1.93, 1.88	γCH$_2$ 1.79
D-Pro	γCH$_2$ 1.95, 1.83	γCH$_2$ 2.12, 1.94	γCH$_2$ 2.14, 1.92	γCH$_2$ 1.94, 1.83	γCH$_2$ 1.83, 1.53	γCH$_2$ 1.90
	δCH$_2$ 3.58, 3.49	γCH$_2$ 3.77, 3.55	δCH$_2$ 3.71, 3.50	δCH$_2$ 3.59, 3.48	δCH$_2$ 3.49, 3.38	δCH$_2$ 3.71
Yaa					H$_{2,6}$ 7.16	H$_{2,6}$ 7.18
					H$_{3,5}$ 6.83	H$_{3,5}$ 6.85

^{13}C-NMR chemical shifts for peptides 7-C04-A2, 7-C04-A3 and 7-C04-A4

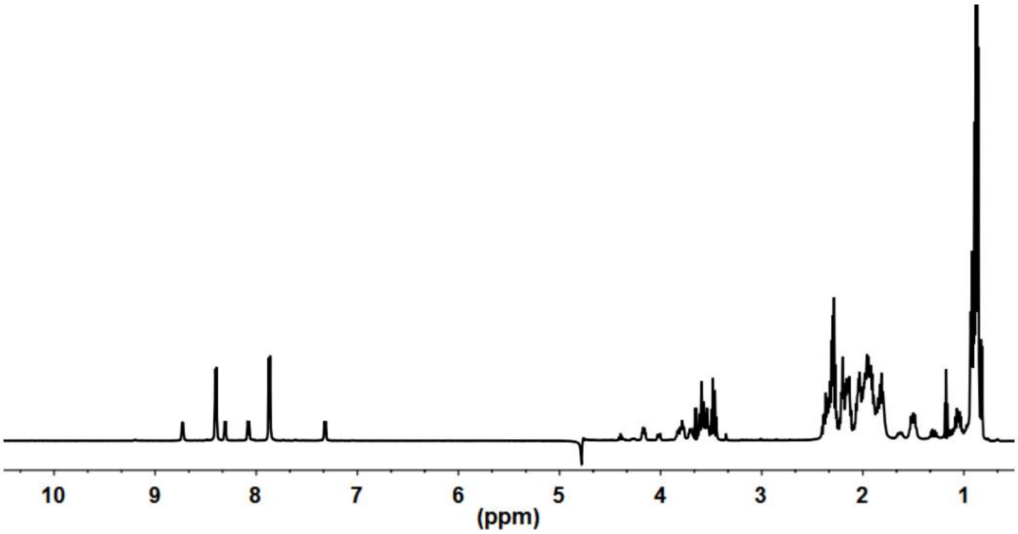
	Cα						Cβ					
Residue	7-C04-A2		7-04-A3		7-C04-A4		7-C04-A2		7-04-A3		7-C04-A4	
	cc	tt	tt	cc	cc	tt	cc	tt	tt	cc	cc	tt
D-Xaa	53.7	55.6	51.0	50.4	53.3	57.8	31.7	30.6	19.7	20.1	31.0	30.9
D-Pro	62.8	64.4	63.9	62.7	62.8	64.0	34.6	30.5	30.7	34.5	34.6	31.0
Yaa	52.2	52.5	54.1	52.1	57.4	57.8	19.8	18.9	18.9	20.4	39.4	39.2

	Other					
Residue	7-C04-A2		7-04-A3		7-C04-A4	
	cc	tt	tt	cc	cc	tt
D-Xaa	γCH_2 35.5	γCH_2 35.5			γCH_2 34.4	γCH_2 34.9
D-Pro	γCH_2 24.3	γCH_2 28.2	γCH_2 28.3	γCH_2 24.5	γCH_2 24.3	γCH_2 27.7
	δCH_2 49.8	δCH_2 50.5	δCH_2 50.4	δCH_2 49.9	δCH_2 49.6	δCH_2 50.4
Yaa					$C_{2,6}$ 133.3	$C_{2,6}$ 133.3
					$C_{3,5}$ 118.1	$C_{3,5}$ 118.1

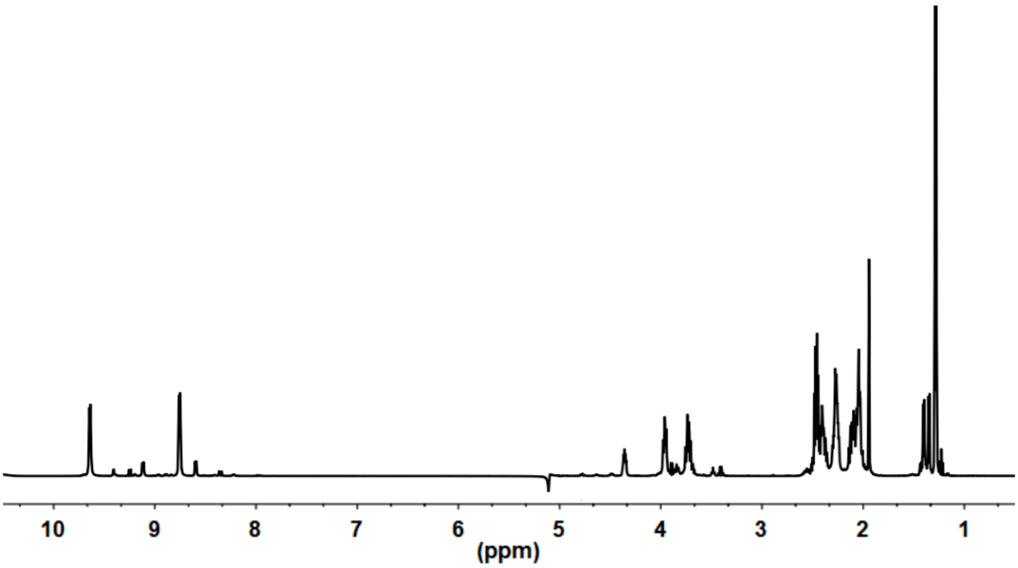
1D ^{1}H-NMR spectra for peptide 7-C04-P1 c(wpEwpE)



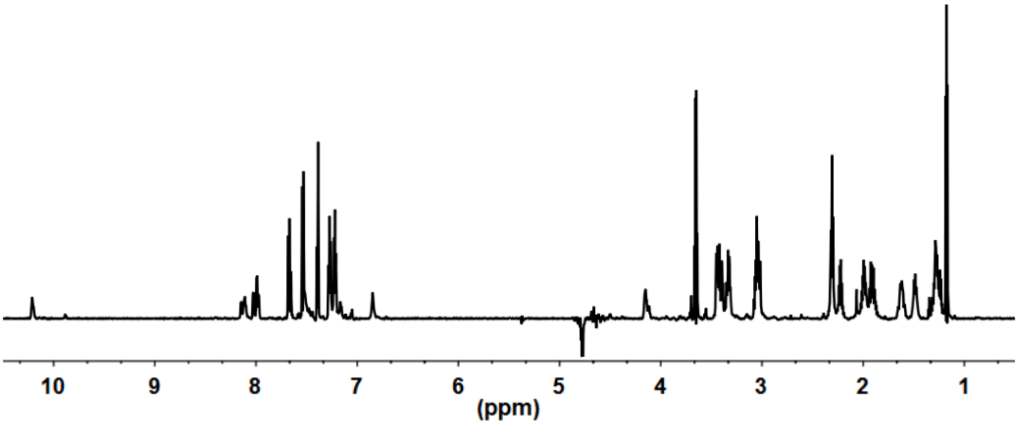
1D ^{1}H-NMR spectra for peptide 7-C04-P1-A1 c(ipEipE)



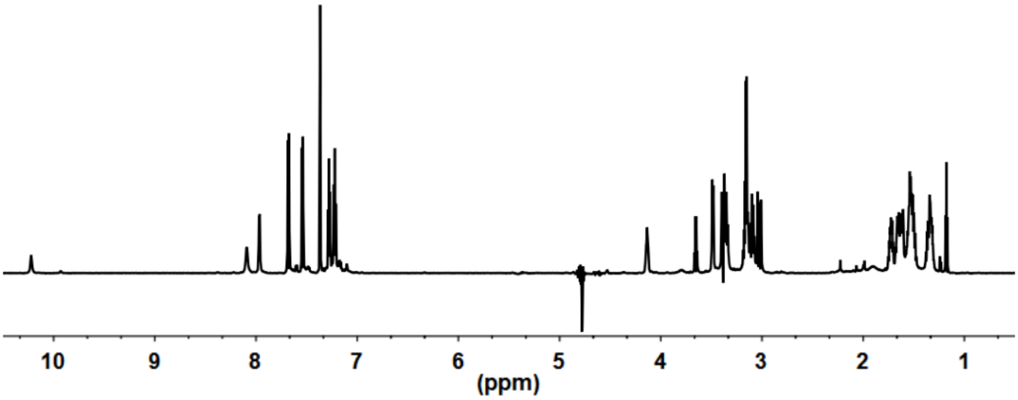
1D ^{1}H-NMR spectra for peptide 7-C04-P1-A2 c(apEapE)



1D ^{1}H-NMR spectra for peptide 7-C04-P1-A3 c(wpQwpQ)



1D ^{1}H-NMR spectra for peptide 7-C04-P1-A4 c(wpRwpR)



^{1}H-NMR chemical shifts for peptides 7-C04-P1, 7-C04-P1-A1, 7-C04-P1-A2, 7-C04-P1-A3, 7-C04-P1-A4.

	NH							
Residue	**7-C04-P1**	**7-C04-P1-A1**			**7-C04-P1A2**		**7-C04-P1-A3**	**7-C04-P1-A4**
		cc	**ct**		**tt**	**cc**		
D-Xaa	8.01	7.87	8.07	8.73	8.16	8.01	7.99	7.97
D-Pro	-	-	-	-	-	-	-	-
Yaa	8.14	8.40	8.30	7.31	8.99	8.50	8.11	8.09

	Hα							
Residue	**7-C04-P1**	**7-C04-P1-A1**			**7-C04-P1A2**		**7-C04-P1-A3**	**7-C04-P1-A4**
		cc	**ct**		**tt**	**cc**		
D-Xaa	4.49	4.40	4.48	4.02	4.69	4.33	4.50	4.52
D-Pro	3.43	4.79	5.00	4.53	4.46	4.59	3.45	3.48
Yaa	4.11	4.16	4.25	4.34	4.07	4.19	4.13	4.13

	Hβ							
Residue	**7-C04-P1**	**7-C04-P1-A1**			**7-C04-P1A2**		**7-C04-P1-A3**	**7-C04-P1-A4**
		cc	**ct**		**tt**	**cc**		
D-Xaa	3.39	1.80	1.82	2.02	1.22	1.30	3.40	3.38
	3.03						3.05	3.01
D-Pro	1.62	2.36	2.30	2.19	2.27	2.40	1.62	1.62
	1.24	2.14	2.13		1.94	2.15	1.28	1.35
Yaa	1.97	2.00	1.96	1.85	2.13	1.92	1.99	1.73
	1.87	1.93			1.99		1.91	1.66

Residue	Other							
	7-C04-P1	7-C04-P1-A1			7-C04-P1-A2		7-C04-P1-A3	7-C04-P1-A4
		cc	ct		tt	cc		
D-Xaa	H_2 7.40	γCH_2 1.50. 1.05	γCH_2 1.50. 1.11	γCH_2 1.62. 1.30			H_2 7.38	H_2 7.36
	H_4 7.66	γCH_3 0.88	γCH_3 0.88	γCH_3 0.88			H_4 7.67	H_4 7.67
	H_5 7.21	δCH_3 0.86	δCH_3 0.82	δCH_3 0.91			H_5 7.22	H_5 7.22
	H_6 7.27						H_6 7.27	H_6 7.28
	H_7 7.53						H_7 7.54	H_7 7.54
	NH 10,20						NH 10,22	NH 10,22
D-Pro	γCH_2 1.47, 1.24	γCH_2 1.96, 1.81	γCH_2 1.98, 1.87	γCH_2 1.91	γCH_2 2.13, 1.93	γCH_2 1.94, 1.84	γCH_2 1.49, 1.27	γCH_2 1.49, 1.32
	δCH_2 3.31, 3.01	δCH_2 3.59, 3.47	δCH_2 3.54	δCH_2 3.82, 3.70	δCH_2 3.71, 3.50	δCH_2 3.59, 3.47	δCH_2 3.33, 3.05	δCH_2 3.33, 3.09
Yaa	γCH_2 2.22	γCH_2 2.29	γCH_2 2.12	γCH_2 2.32	γCH_2 2.24	γCH_2 2.30	γCH_2 2.31	γCH_2 1.54 δCH_2 3.15

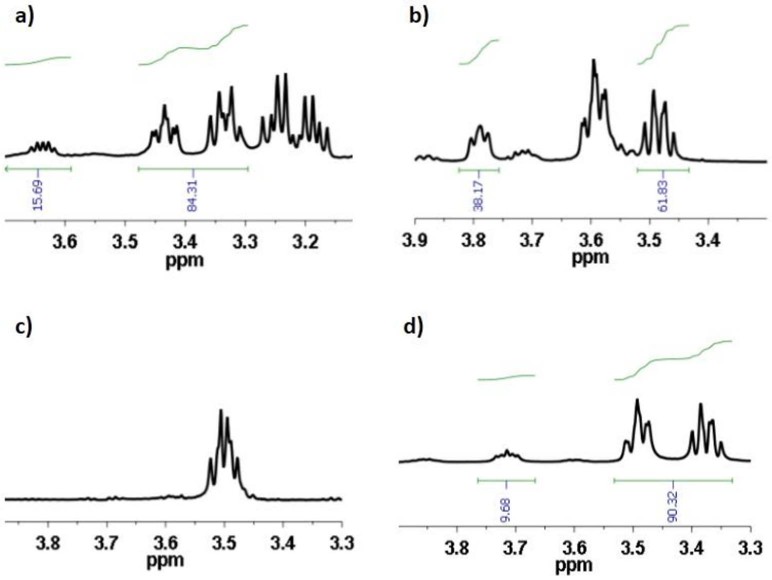
^{13}C-NMR chemical shifts for peptides 7-C04-P1, 7-C04-P1-A1, 7-C04-P1-A2, 7-C04-P1-A3, 7-C04-P1-A4.

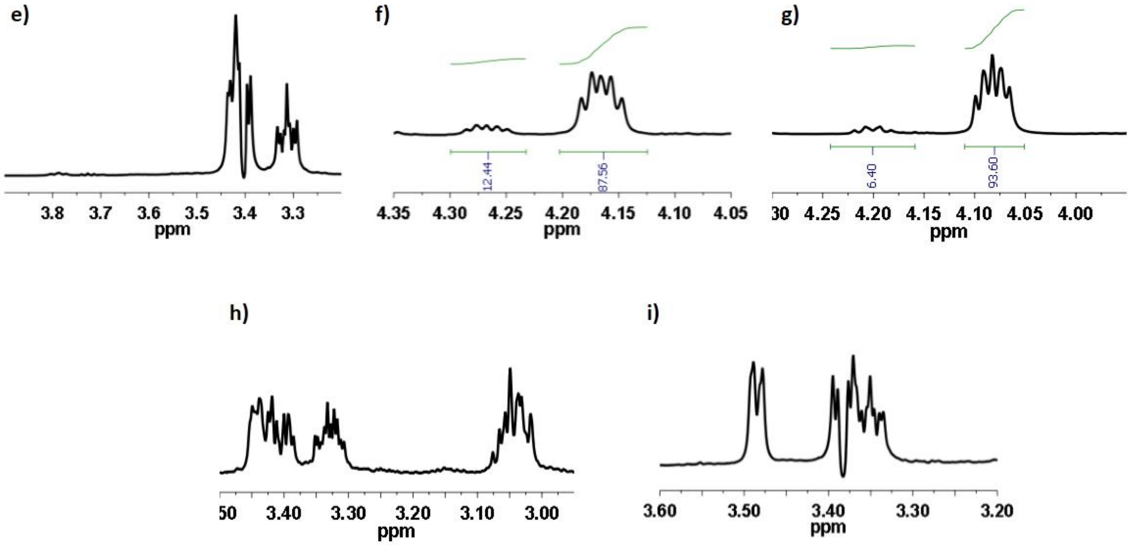
Residue	Cα							
	7-C04-P1	7-C04-P1-A1			7-C04-P1-A2		7-C04-P1-A3	7-C04-P1-A4
		cc	ct		tt	cc		
D-Xaa	56.3	58.2	58.0	64.4	51.1	50.8	56.2	56.2
D-Pro	62.6	63.2	63.5	64.3	63.9	62.7	62.6	63.6
Yaa	56.5	56.8	56.2	55.2	58.2	56.4	55.9	56.2

Residue	Cβ							
	7-C04-P1	7-C04-P1-A1			7-C04-P1-A2		7-C04-P1-A3	7-C04-P1-A4
		cc	ct		tt	cc		
D-Xaa	31.6	42.0	39.7	35.8	19.6	20.4	31.4	31.4
D-Pro	33.9	34.7	34.8	30.7	30.7	34.8	29.2	34.0
Yaa	30.1	30.6	30.9	30.5	29.8	28.3	34.0	30.4

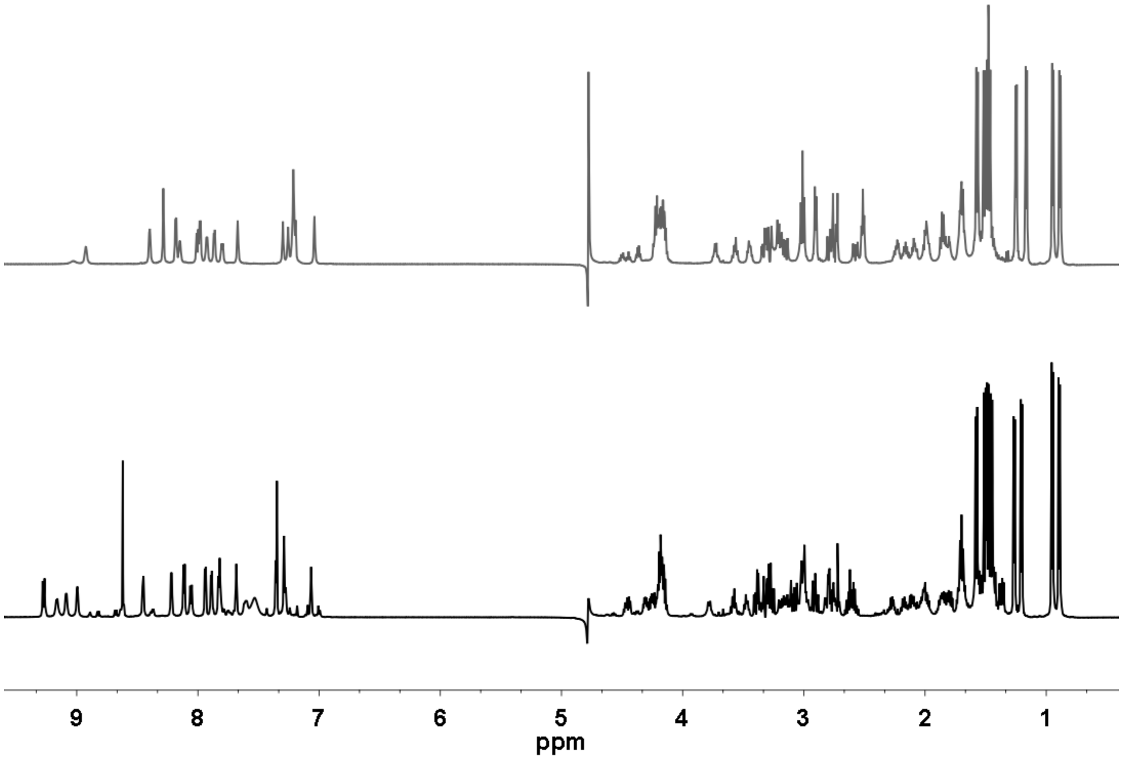
Residue	Other							
	7-C04-P1	7-C04-A1			7-C04-A2		7-C04-P1-A3	7-C04-P1-A4
		cc	ct		cc	tt		
D-Xaa	C_2 127.6	γCH2 27.3	γCH2 27.3	γCH2 28.8			C_2 127.5	C_2 127.5
	C_4 121.1	γCH_3 13.7	γCH_3 13.7	γCH_3 13.7			C_4 121.2	C_4 121.2
	C_5 124.6	δCH_3 16.4	δCH_3 16.9	δCH_3 17.3			C_5 124.6	C_5 124.6
	C_6 122.0						C_6 122.0	C_6 122.0
	C_7 114.6						C_7 114.5	C_7 114.5
D-Pro	γCH_2 24.4	γCH_2 24.2	γCH_2 24.0	γCH_2 27.5	γCH_2 28.4	γCH_2 24.7	γCH_2 24.4	γCH_2 24.4
	δCH_2 49.6	δCH_2 49.5	δCH_2 49.4	δCH_2 49.9	δCH_2 50.4	δCH_2 50.0	δCH_2 49.7	δCH_2 49.7
Yaa	γCH2 36.3	γCH2 36.5	γCH2 36.6	γCH2 35.5	γCH2 36.7	γCH2 36.2	γCH2 34.5	γCH2 27.0
								δCH_2 43.3

Zoom of the 1D ^{1}H-NMR spectra region used to determine *cis/cis* content in peptide a) 7-C04, b) 7-C04-A2, c) 7-C04-A3, d) 7-C04-A4, e) 7-C04-P1, f) 7-C04-P1-A1, g) 7-C04-P1-A2, h) 7-C04-P1-A3, i) 7-C04-P1-A4. *Quantification for peptide 7-C04-A3, 7-C04-P1, 7-C04-P1-A3 and 7-C04-P1-A4 could not be done because it was below the quantification threshold (5%).*

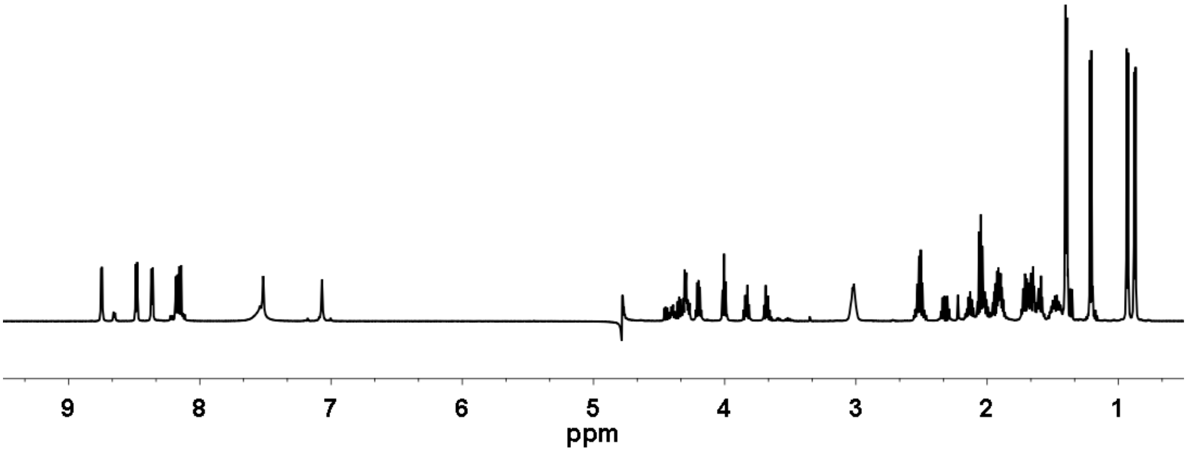




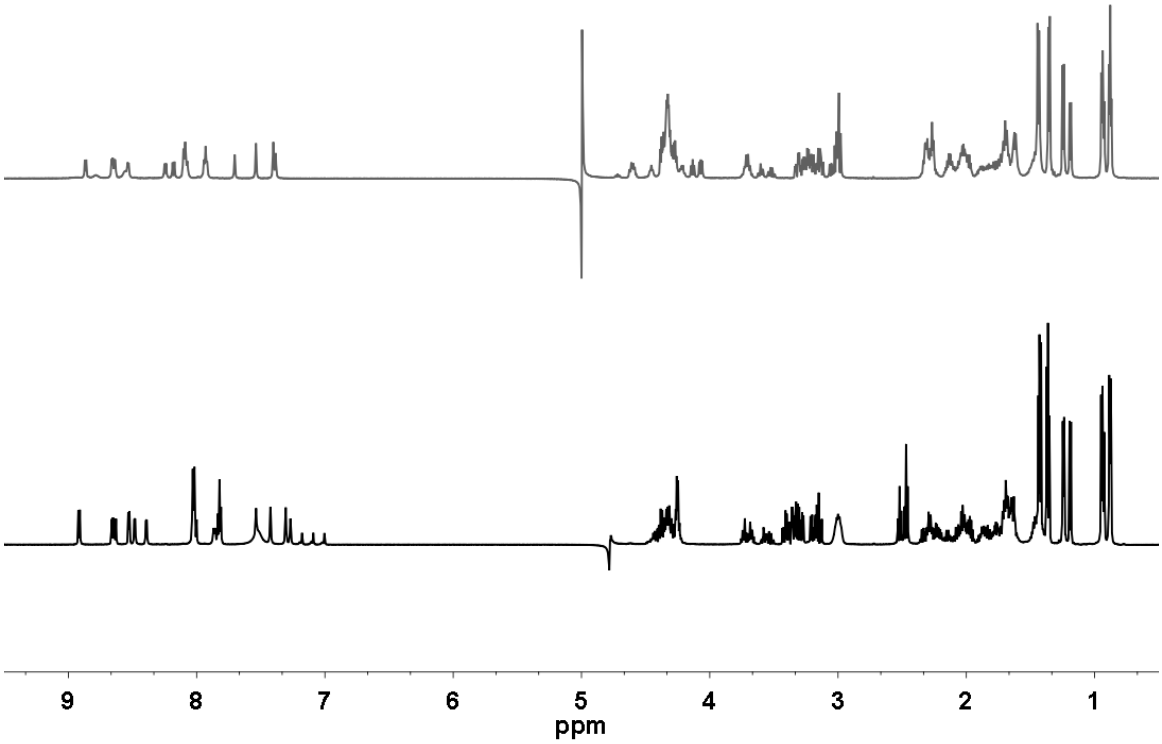
^{1}H NMR spectra for MiniAp-1 at 298 K in: water at pH 2-3 (black) or 10 mM phosphate buffer pH 7.4 (grey)



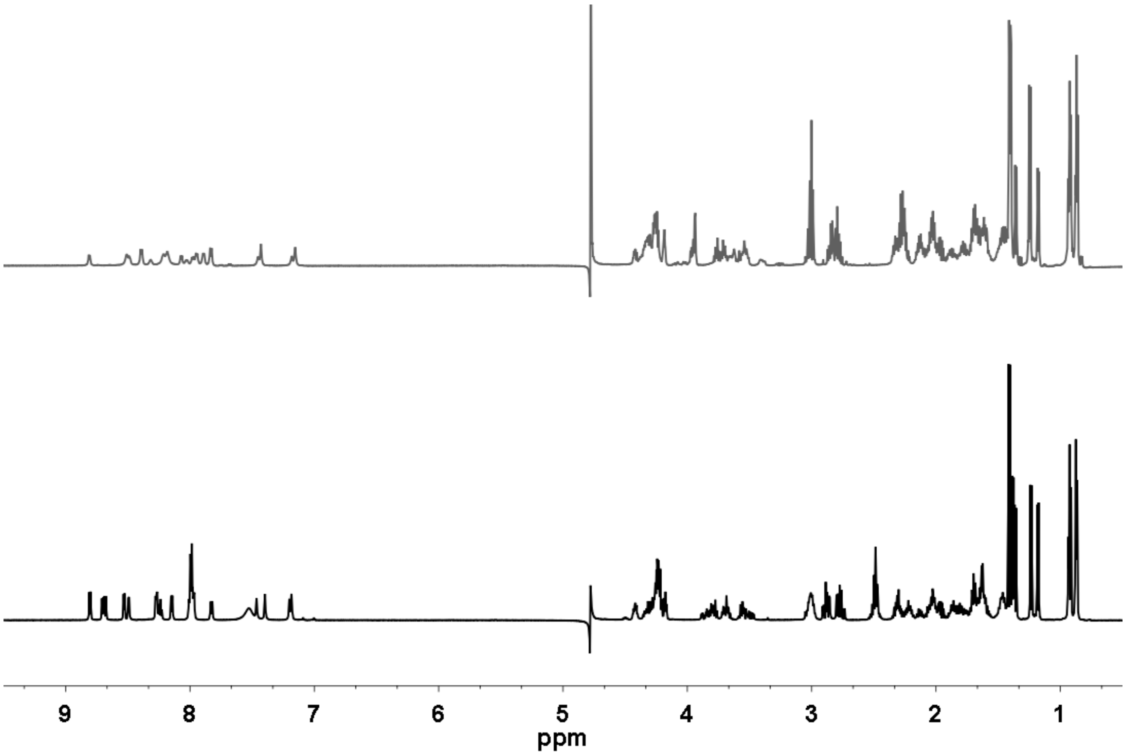
^{1}H NMR spectra for MiniAp-2 at 298 K in water at pH 2-3.



^{1}H NMR spectra for MiniAp-3 at 298 K in: water at pH 2-3 (black) or 10 mM phosphate buffer pH 7.4 (grey)



^{1}H NMR spectra for MiniAp-4 at 298 K in: water at pH 2-3 (black) or 10 mM phosphate buffer pH 7.4 (grey)



^{1}H NMR chemical shift for: A) MiniAp-1, B) *trans* MiniAp-3, C) *cis* MiniAp-3, D) *trans* MiniAp-4, E) *cis* MiniAp-4, E) MiniAp-2 at pH 2-3 and 298 K.

	NH						αCH						βCH					
Residue	A	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F
Cys							4.44						3.13					
													2.73					
Asn	9.27						4.90						3.08					
													2.78					
Cys/Dapa	9.09						4.70	4.36	4.32	4.21	4.24		3.31	3.41	3.33	3.81	3.86	
													2.77	3.31	3.15	3.70	3.50	
Lys	8.06	8.92	8.63	8.80	8.68		4.25	4.37	4.43	4.24	4.42	4.00	1.84	1.88	1.84	1.87	1.85	1.90
														1.77	1.75	1.80	1.75	
Ala	7.28	7.81	8.48	8.15	8.49	8.75	4.56	4.68	4.32	4.66	4.27	4.63	1.20	1.37	1.36	1.38	1.36	1.40
Pro							4.47	4.39	4.63	4.42	4.65	4.44	2.00	2.33	2.28	2.31	2.30	2.32
													1.78	1.98	2.16	1.96	2.28	1.92
Glu	9.16	7.86	8.66	8.00	8.70	8.48	4.30	4.43	4.47	4.41	4.50	4.40	2.27	2.24	2.21	2.23	2.20	2.13
													2.16	2.06	2.14	2.05	2.12	2.01
Thr	7.61	8.02	8.02	7.98	7.98	8.15	4.66	4.25	4.30	4.17	4.23	4.30	4.62	4.25	4.33	4.22	4.30	4.19
Ala	8.99	8.53	8.39	8.53	8.27	8.36	4.19	4.26	4.24	4.24	4.24	4.34	1.48	1.42	1.43	1.41	1.40	1.39
Leu	8.22	7.83	8.00	7.83	7.99	8.18	4.17	4.34	4.31	4.34	4.28	4.28	1.70	1.71	1.63	1.63	1.62	1.59
													1.57	1.64				1.65
Cys/Asp	7.81	8.02	8.02	7.97	7.99		4.70	4.61	4.60	4.70	4.73		3.18	3.28	3.37	2.88	2.87	
													2.81	3.19	3.15	2.77	2.76	
Ala	8.45						3.78						1.45					
Ala	7.89						4.15						1.50					
Ala	7.94						4.19						1.57					
Cys	8.11						4.44						3.00					
													2.90					
His	7.83						4.62						3.39					
													3.27					

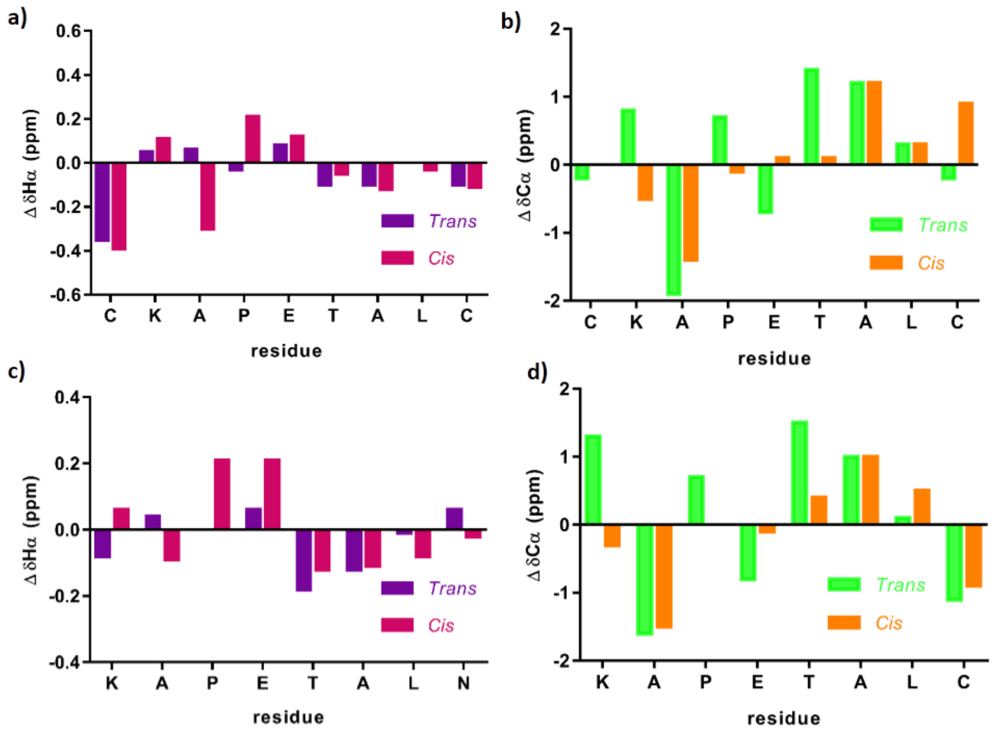
	others					
Residue	A	B	C	D	E	F
Cys						
Asn	NH_2 7.07, 7.68					
Cys/Dapa						
Lys	γCH_2 1.43, 1.52	γCH_2 1.41, 1.45	γCH2 1.46	γCH_2 1.46	γCH_2 1.46	γCH_2 1.47
	δCH_2 1.69	δCH_2 1.68	δCH_2 1.69	δCH_2 1.69	δCH_2 1.69	δCH_2 1.70
	εCH_2 3.00	εCH_2 3.00	εCH_2 3.01	εCH_2 3.00	εCH_2 3.01	εCH_2 3.02
	εNH_3 7.53	εNH_3 7.52	εNH_3 7.52	εNH_3 7.53	εNH_3 7.53	εNH_3 7.54
Ala						
Pro	γCH_2 2.11, 2.00	γCH_2 2.02	γCH_2 1.98, 1.81	γCH_2 2.02	γCH_2 1.98, 1.80	γCH_2 2.04
	δCH_2 3.57, 3.47	δCH_2 3.73, 3.67	δCH_2 3.56, 3.53	δCH_2 3.78, 3.67	δCH_2 3.55	δCH_2 3.83, 3.68
Glu	γCH_2 2.60	γCH_2 2.47	γCH_2 2.52	γCH_2 2.48	γCH_2 2.49	γCH_2 2.51
Thr	γCH_2 1.26	γCH_2 1.24	γCH_2 1.19	γCH_2 1.24	γCH_2 1.18	γCH_2 1.21
Ala						
Leu	γCH_2 1.70	γCH_2 1.63	γCH_2 1.65	γCH_2 1.62, 1.58	γCH_2 1.62	γCH_2 1.65
	δCH_3 0.95, 0.89	δCH_3 0.95, 0.89	δCH_3 0.94, 0.89	δCH_3 0.92, 0.87	δCH_3 0.93, 0.87	δCH_3 0.93, 0.87
Cys/Asp						
Ala						
Ala						
Ala						
Cys						
His	C_2H 8.62					
	C_4H 7.35					

^{13}C NMR chemical shift for: A) MiniAp-1, B) *trans* MiniAp-3, C) *cis* MiniAp-3, D) *trans* MiniAp-4, E) *cis* MiniAp-4, E) MiniAp-2 at 298K, pH 2-3.

	Cα						Cβ					
Residue	A	B	C	D	E	F	A	B	C	D	E	F
Cys	54.5						36.5					
Asn	51.2						38.4					
Cys/Dapa	54.3	55.2	55.4	55.5	55.5		34.7	40.6	41.2	42.6	43.0	
Lys	57.5	57.0	55.7	57.5	55.9	55.6	32.2	32.7	34.0	32.9	33.9	33.2
Ala	50.5	50.6	51.1	50.9	51.3	50.7	18.1	19.2	17.8	18.7	17.7	19.1
Pro	64.2	64.0	63.2	64.0	63.2	63.0	31.9	31.9	33.7	31.7	33.7	32.0
Glu	57.9	55.7	56.7	55.8	56.4	55.9	29.4	28.2	28.0	27.9	27.8	28.8
Thr	58.9	63.2	61.9	63.3	62.1	61.8	72.7	69.7	69.9	69.6	69.5	69.9
Ala	55.0	53.7	53.7	53.5	53.5	52.5	17.8	18.9	18.9	18.8	18.8	17.9
Leu	57.9	55.4	55.4	55.2	55.5	55.0	41.9	41.9	41.9	42.0	41.6	42.4
Cys/Asp	56.5	55.2	56.3	53.1	53.3		35.8	41.2	41.2	39.7	39.7	
Ala	55.7						17.5					
Ala	54.7						17.9					
Ala	54.5						18.4					
Cys	57.3						39.4					
His	55.9						28.8					

Residue	others A	B	C	D	E	F
Cys						
Asn						
Cys/Dapa						
Lys	γCH$_2$ 25.2	γCH$_2$ 24.7	γCH2 24.7	γCH$_2$ 24.8	γCH$_2$ 24.4	γCH$_2$ 23.8
	δCH$_2$ 28.8	δCH$_2$ 29.1	δCH$_2$ 29.1	δCH$_2$ 29.1	δCH$_2$ 29.1	δCH$_2$ 29.1
	εCH$_2$ 42.1	εCH$_2$ 42.1	εCH$_2$ 42.1	εCH$_2$ 42.0	εCH$_2$ 42.0	εCH$_2$ 41.9
Ala						
Pro	γCH$_2$ 27.6	γCH$_2$ 27.4	γCH$_2$ 24.5	γCH$_2$ 27.4	γCH$_2$ 24.5	γCH$_2$ 27.3
	δCH$_2$ 49.6	δCH$_2$ 50.5	δCH$_2$ 50.2	δCH$_2$ 50.5	δCH$_2$ 50.5	δCH$_2$ 50.4
Glu	γCH$_2$ 33.9	γCH$_2$ 32.9	γCH$_2$ 33.1	γCH$_2$ 32.9	γCH$_2$ 32.9	γCH$_2$ 32.8
Thr	γCH$_2$ 22.0	γCH$_2$ 21.8	γCH$_2$ 21.7	γCH$_2$ 21.7	γCH$_2$ 21.7	γCH$_2$ 21.5
Ala						
Leu	γCH$_2$ 27.2	γCH2 27.4	γCH$_2$ 27.0	γCH$_2$ 27.0	γCH$_2$ 27.0	γCH$_2$ 27.0
	δCH$_3$ 24.7, 23.8	δCH$_3$ 25.0, 23.1	δCH$_3$ 24.9, 23.1	δCH$_3$ 24.9, 23.2	δCH$_3$ 24.9, 23.2	δCH$_3$ 24.9, 23.1
Cys/Asp						
Ala						
Ala						
Ala						
Cys						
His	C$_2$ 136.2					
	C$_4$ 119.9					

Secondary $^1H_\alpha$ and $^{13}C_\alpha$ chemical shifts for: a) *trans* MiniAp-3 (purple and green), b) *cis* MiniAp-3 (pink and orange), c) *trans* MiniAp-4 (purple and green) and d) *cis* MiniAp-4 (pink and orange).



CONCLUSIONS

Conclusions related to the first objective:

1. Cyclic hexapeptides containing tryptophan residues in their sequences bind VEGF in the receptor binding epitope, with low mM affinities. Hence, these peptides can be used as a starting point to design VEGF-binders with higher affinity and specificity. NMR proved to be a suitable technique for the screening of low affinity VEGF-binders.
2. Cyclic hexapeptides with C2 backbone symmetry adopt a defined conformation containing two βII' turns with two hydrogen bonds between the residues preceding the D-proline residue.
3. Residue's configuration seems to slightly affect VEGF-binding. C2 backbone symmetry in cyclic hexapeptides is not required for VEGF-binding. More hydrophobic residues slightly improve VEGF-binding affinity. Finally, ring size expansion was key to find a cyclic dodecapeptide with improved VEGF-binding affinity and specificity.
4. Residue's configuration significantly affect the conformation of cyclic hexapeptides. Challenging types of β turns containing Xaa-Pro bonds in the *cis* conformation can be achieved with cyclic hexapeptides. These peptides can be useful to target newly discovered systems such as cyclophilins.

Conclusions related to the second objective

5. MiniAp-1 adopts a very similar conformation of that described for apamin, defined by a β-turn in the segment Asn2-Ala5 and a C-terminal α-helix comprising residues Ala9-His16.
6. Significant differences were observed for the two monocyclic peptides (MiniAp-3 and MiniAp-4), compared with the linear analogue (MiniAp-2); the last showing a clear random coil conformation. Although structural data for MiniAp-3 and MiniAp-4 suggested that they had no defined secondary structure, we could demonstrate that they had lower conformational flexibility than MiniAp-2. Most remarkably, while the linear analogue showed a 1:9 *cis/trans* Ala-Pro conformer population ratio, like the bicyclic peptides (apamin and MiniAp-1), the two monocyclic derivatives showed a 1:1 ratio. We hypothesized that the higher preference for the *cis* Ala-Pro rotamer in MiniAp-4 with respect to MiniAp-1 could partly account for the differences in permeability between the two analogues with highest resistance to proteases.

Conclusions related to the third objective:

7. A protocol for the expression and purification of soluble Sso7d-based cysteine mutants was set-up and further optimized. Cation exchange chromatography proved to be the best methodology to eliminate both the chaperone coeluting with the Sso7d proteins and the non-reacted dye after labeling reactions.
8. M11.1.3(W25C)-4-DMN2 proved to be the best biosensor to detect protein-protein interaction between the engineered Sso7d binder (M11.1.3) and MSA, as it presented low fluorescence background and high fluorescence intensity increase upon MSA binding.
9. Yeast surface display methodology was successfully used to engineer Sso7d-based hEGF binders. First hEGF binders were characterized and a K_D of ~3 µM was obtained. Further efforts need to be done to convert these hEGF binders into fluorescence-based biosensors.

REFERENCES

1. London, N., Raveh, B. & Schueler-Furman, O. Druggable protein–protein interactions – from hot spots to hot segments. *Current Opinion in Chemical Biology* **17**, 952-959 (2013).
2. Nevola, L. & Giralt, E. Modulating protein-protein interactions: the potential of peptides. *Chemical Communications* **51**, 3302-3315 (2015).
3. Buchwald, P. Small-molecule protein–protein interaction inhibitors: Therapeutic potential in light of molecular size, chemical space, and ligand binding efficiency considerations. *IUBMB Life* **62**, 724-731 (2010).
4. Hopkins, A.L. & Groom, C.R. The druggable genome. *Nat Rev Drug Discov* **1**, 727-730 (2002).
5. Fuller, J.C., Burgoyne, N.J. & Jackson, R.M. Predicting druggable binding sites at the protein–protein interface. *Drug Discovery Today* **14**, 155-161 (2009).
6. Clackson, T. & Wells, J. A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383-386 (1995).
7. Bogan, A.A. & Thorn, K.S. Anatomy of hot spots in protein interfaces1. *Journal of Molecular Biology* **280**, 1-9 (1998).
8. Carbonell, P., Nussinov, R. & del Sol, A. Energetic determinants of protein binding specificity: Insights into protein interaction networks. *PROTEOMICS* **9**, 1744-1753 (2009).
9. Keskin, O., Ma, B. & Nussinov, R. Hot Regions in Protein–Protein Interactions: The Organization and Contribution of Structurally Conserved Hot Spot Residues. *Journal of Molecular Biology* **345**, 1281-1294 (2005).
10. Reichmann, D. et al. The modular architecture of protein–protein binding interfaces. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 57-62 (2005).
11. Moza, B. et al. Long-range cooperative binding effects in a T cell receptor variable domain. *Proceedings of the National Academy of Sciences* **103**, 9867-9872 (2006).
12. London, N., Movshovitz-Attias, D. & Schueler-Furman, O. The Structural Basis of Peptide-Protein Binding Strategies. *Structure* **18**, 188-199 (2010).
13. Cheok, C.F., Verma, C.S., Baselga, J. & Lane, D.P. Translating p53 into the clinic. *Nat Rev Clin Oncol* **8**, 25-37 (2011).
14. Petros, A.M. et al. Discovery of a Potent Inhibitor of the Antiapoptotic Protein Bcl-xL from NMR and Parallel Synthesis. *Journal of Medicinal Chemistry* **49**, 656-663 (2006).
15. Braisted, A.C. et al. Discovery of a Potent Small Molecule IL-2 Inhibitor through Fragment Assembly. *Journal of the American Chemical Society* **125**, 3714-3715 (2003).
16. Kouadio, J.-L.K., Horn, J.R., Pal, G. & Kossiakoff, A.A. Shotgun Alanine Scanning Shows That Growth Hormone Can Bind Productively to Its Receptor through a Drastically Minimized Interface. *Journal of Biological Chemistry* **280**, 25524-25532 (2005).

17. Guharoy, M., Pal, A., Dasgupta, M. & Chakrabarti, P. PRICE (PRotein Interface Conservation and Energetics): a server for the analysis of protein–protein interfaces. *Journal of Structural and Functional Genomics* **12**, 33-41 (2011).

18. Lise, S., Buchan, D., Pontil, M. & Jones, D.T. Predictions of Hot Spot Residues at Protein-Protein Interfaces Using Support Vector Machines. *PLoS ONE* **6**, e16774 (2011).

19. Tuncbag, N., Keskin, O. & Gursoy, A. HotPoint: hot spot prediction server for protein interfaces. *Nucleic Acids Research* **38**, W402-W406 (2010).

20. Morrow, J.K. & Zhang, S. Computational Prediction of Hot Spot Residues. *Current pharmaceutical design* **18**, 1255-1265 (2012).

21. Murray, C.W. & Blundell, T.L. Structural biology in fragment-based drug design. *Current Opinion in Structural Biology* **20**, 497-507 (2010).

22. Valkov, E., Sharpe, T., Marsh, M., Greive, S. & Hyvönen, M. in Fragment-Based Drug Discovery and X-Ray Crystallography (eds. Davies, G.T. & Hyvönen, M.) 145-179 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012).

23. Erlanson, D.A., McDowell, R.S. & O'Brien, T. Fragment-Based Drug Discovery. *Journal of Medicinal Chemistry* **47**, 3463-3482 (2004).

24. Yan, C., Wu, F., Jernigan, R.L., Dobbs, D. & Honavar, V. Characterization of Protein–Protein Interfaces. *The protein journal* **27**, 59-70 (2008).

25. Wells, J.A. & McClendon, C.L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001-1009 (2007).

26. McCafferty, J. The long and winding road to antibody therapeutics. *mAbs* **2**, 459-460 (2010).

27. Mullard, A. 2012 FDA drug approvals. *Nat Rev Drug Discov* **12**, 87-90 (2013).

28. Mullard, A. 2013 FDA drug approvals. *Nat Rev Drug Discov* **13**, 85-89 (2014).

29. Mullard, A. 2014 FDA drug approvals. *Nat Rev Drug Discov* **14**, 77-81 (2015).

30. Mullard, A. 2015 FDA drug approvals. *Nat Rev Drug Discov* **15**, 73-76 (2016).

31. Mujić-Delić, A., de Wit, R.H., Verkaar, F. & Smit, M.J. GPCR-targeting nanobodies: attractive research tools, diagnostics, and therapeutics. *Trends in Pharmacological Sciences* **35**, 247-255 (2014).

32. De Meyer, T., Muyldermans, S. & Depicker, A. Nanobody-based products as research and diagnostic tools. *Trends in Biotechnology* **32**, 263-270 (2014).

33. Whitehead, T.A., Baker, D. & Fleishman, S.J. in Methods in Enzymology (ed. Amy, E.K.) 1-19 (Academic Press, 2013).

34. Fleishman, S.J. et al. Computational Design of Proteins Targeting the Conserved Stem Region of Influenza Hemagglutinin. *Science* **332**, 816-821 (2011).

35. Jardine, J. et al. Rational HIV Immunogen Design to Target Specific Germline B Cell Receptors. *Science* **340**, 711-716 (2013).
36. Correia, B.E. et al. Proof of principle for epitope-focused vaccine design. *Nature* **507**, 201-206 (2014).
37. King, C. et al. Removing T-cell epitopes with computational protein design. *Proceedings of the National Academy of Sciences* **111**, 8577-8582 (2014).
38. Merrifield, R.B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **85**, 2149-2154 (1963).
39. Thomas, B., Oleg, M. & Fernando, A. From Production of Peptides in Milligram Amounts for Research to Multi-Tons Quantities for Drugs of the Future. *Current Pharmaceutical Biotechnology* **5**, 29-43 (2004).
40. Higueruelo, A.P., Jubb, H. & Blundell, T.L. Protein–protein interactions as druggable targets: recent technological advances. *Current Opinion in Pharmacology* **13**, 791-796 (2013).
41. Vlieghe, P., Lisowski, V., Martinez, J. & Khrestchatisky, M. Synthetic therapeutic peptides: science and market. *Drug Discovery Today* **15**, 40-56 (2010).
42. Ahrens, V.M., Bellmann-Sickert, K. & Beck-Sickinger, A.G. Peptides and peptide conjugates: therapeutics on the upward path. *Future Medicinal Chemistry* **4**, 1567-1586 (2012).
43. Wójcik, P. & Berlicki, Ł. Peptide-based inhibitors of protein–protein interactions. *Bioorganic & Medicinal Chemistry Letters* **26**, 707-713 (2016).
44. Prades, R. et al. Applying the Retro-Enantio Approach To Obtain a Peptide Capable of Overcoming the Blood–Brain Barrier. *Angewandte Chemie International Edition* **54**, 3967-3972 (2015).
45. Liu, M. et al. A Left-Handed Solution to Peptide Inhibition of the p53–MDM2 Interaction. *Angewandte Chemie International Edition* **49**, 3649-3652 (2010).
46. Mándity, I.M. & Fülöp, F. An overview of peptide and peptoid foldamers in medicinal chemistry. *Expert Opinion on Drug Discovery* **10**, 1163-1177 (2015).
47. Cabrele, C., Martinek, T.A., Reiser, O. & Berlicki, Ł. Peptides Containing β-Amino Acid Patterns: Challenges and Successes in Medicinal Chemistry. *Journal of Medicinal Chemistry* **57**, 9718-9739 (2014).
48. Horne, W.S. Peptide and peptoid foldamers in medicinal chemistry. *Expert Opinion on Drug Discovery* **6**, 1247-1262 (2011).
49. Koyack, M.J. & Cheng, R.P. in Protein Design: Methods and Applications (eds. Guerois, R. & de la Paz, M.L.) 95-109 (Humana Press, Totowa, NJ, 2006).
50. Weiss, H.M. et al. ADME Investigations of Unnatural Peptides: Distribution of a 14C-Labeled β 3-Octaarginine in Rats. *Chemistry & Biodiversity* **4**, 1413-1437 (2007).

51. Laursen, J.S., Engel-Andreasen, J. & Olsen, C.A. β-Peptoid Foldamers at Last. *Accounts of Chemical Research* **48**, 2696-2704 (2015).

52. Hara, T., Durell, S.R., Myers, M.C. & Appella, D.H. Probing the Structural Requirements of Peptoids That Inhibit HDM2−p53 Interactions. *Journal of the American Chemical Society* **128**, 1995-2004 (2006).

53. Bock, J.E., Gavenonis, J. & Kritzer, J.A. Getting in Shape: Controlling Peptide Bioactivity and Bioavailability Using Conformational Constraints. *ACS Chemical Biology* **8**, 488-499 (2013).

54. Bhat, A., Roberts, L.R. & Dwyer, J.J. Lead discovery and optimization strategies for peptide macrocycles. *European Journal of Medicinal Chemistry* **94**, 471-479 (2015).

55. Gao, M., Cheng, K. & Yin, H. Targeting protein−protein interfaces using macrocyclic peptides. *Peptide Science* **104**, 310-316 (2015).

56. Tsomaia, N. Peptide therapeutics: Targeting the undruggable space. *European Journal of Medicinal Chemistry* **94**, 459-470 (2015).

57. Mahon, A.B. & Arora, P.S. Design, synthesis and protein-targeting properties of thioether-linked hydrogen bond surrogate helices. *Chemical Communications* **48**, 1416-1418 (2012).

58. Henchey, L.K., Jochim, A.L. & Arora, P.S. Contemporary strategies for the stabilization of peptides in the α-helical conformation. *Current Opinion in Chemical Biology* **12**, 692-697 (2008).

59. Cabezas, E. & Satterthwait, A.C. The Hydrogen Bond Mimic Approach: Solid-Phase Synthesis of a Peptide Stabilized as an α-Helix with a Hydrazone Link. *Journal of the American Chemical Society* **121**, 3862-3875 (1999).

60. Verdine, G.L. & Hilinski, G.J. in Methods in Enzymology (eds. Wittrup, K.D. & Gregory, L.V.) 3-33 (Academic Press, 2012).

61. Guerlavais, V. & Sawyer, T.K. in Annual Reports in Medicinal Chemistry (ed. Manoj, C.D.) 331-345 (Academic Press, 2014).

62. Walensky, L.D. & Bird, G.H. Hydrocarbon-Stapled Peptides: Principles, Practice, and Progress. *Journal of Medicinal Chemistry* **57**, 6275-6288 (2014).

63. Barthe, P., Roumestand, C., Rochette, S. & Vita, C. Synthesis and NMR solution structure of an α-helical hairpin stapled with two disulfide bridges. *Protein Science* **9**, 942-955 (2000).

64. Taylor, J.W. The synthesis and study of side-chain lactam-bridged peptides. *Peptide Science* **66**, 49-75 (2002).

65. Fairlie, D.P. & de Araujo, A.D. Stapling peptides using cysteine crosslinking. *Peptide Science*, n/a-n/a (2016).

66. Liang, L. & Astruc, D. The copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) “click” reaction and its applications. An overview. *Coordination Chemistry Reviews* **255**, 2933-2945 (2011).
67. Lau, Y.H., de Andrade, P., Wu, Y. & Spring, D.R. Peptide stapling techniques based on different macrocyclisation chemistries. *Chemical Society Reviews* **44**, 91-102 (2015).
68. Bird, G.H., Gavathiotis, E., LaBelle, J.L., Katz, S.G. & Walensky, L.D. Distinct BimBH3 (BimSAHB) Stapled Peptides for Structural and Cellular Studies. *ACS Chemical Biology* **9**, 831-837 (2014).
69. Bernal, F. et al. A Stapled p53 Helix Overcomes HDMX-Mediated Suppression of p53. *Cancer Cell* **18**, 411-422 (2010).
70. Chu, Q. et al. Towards understanding cell penetration by stapled peptides. *MedChemComm* **6**, 111-119 (2015).
71. Nevola, L. et al. Light-Regulated Stapled Peptides to Inhibit Protein–Protein Interactions Involved in Clathrin-Mediated Endocytosis. *Angewandte Chemie International Edition* **52**, 7704-7708 (2013).
72. Robinson, J.A., DeMarco, S., Gombert, F., Moehle, K. & Obrecht, D. The design, structures and therapeutic potential of protein epitope mimetics. *Drug Discovery Today* **13**, 944-951 (2008).
73. Robinson, J.A. β-Hairpin Peptidomimetics: Design, Structures and Biological Activities. *Accounts of Chemical Research* **41**, 1278-1288 (2008).
74. Martínez, L. et al. Hydrogen Bonded Squaramide-Based Foldable Module Induces Both β- and α-Turns in Hairpin Structures of α-Peptides in Water. *Organic Letters* **17**, 2980-2983 (2015).
75. Ganesh Kumar, M., Mali, S.M., Raja, K.M.P. & Gopi, H.N. Design of Stable β-Hairpin Mimetics through Backbone Disulfide Bonds. *Organic Letters* **17**, 230-233 (2015).
76. Park, J.W. & Lee, Y.G. in HCI International 2013 - Posters’ Extended Abstracts: International Conference, HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part I (ed. Stephanidis, C.) 69-73 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).
77. Makwana, K.M. & Mahalakshmi, R. Trp-Trp Cross-Linking: A Structure–Reactivity Relationship in the Formation and Design of Hyperstable Peptide β-Hairpin and α-Helix Scaffolds. *Organic Letters* **17**, 2498-2501 (2015).
78. Robinson, J.A. Protein epitope mimetics as anti-infectives. *Current Opinion in Chemical Biology* **15**, 379-386 (2011).
79. Craik, D.J., Fairlie, D.P., Liras, S. & Price, D. The Future of Peptide-based Drugs. *Chemical Biology & Drug Design* **81**, 136-147 (2013).
80. Kim, Y.-W., Grossmann, T.N. & Verdine, G.L. Synthesis of all-hydrocarbon stapled [alpha]-helical peptides by ring-closing olefin metathesis. *Nat. Protocols* **6**, 761-771 (2011).

81. Mahon, A.B. & Arora, P.S. End-capped α-helices as modulators of protein function. *Drug Discovery Today: Technologies* **9**, e57-e62 (2012).

82. Venkatachalam, C.M. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers* **6**, 1425-1436 (1968).

83. White, C.J. & Yudin, A.K. Contemporary strategies for peptide macrocyclization. *Nat Chem* **3**, 509-524 (2011).

84. Clark, R.J. & Craik, D.J. Invited reviewnative chemical ligation applied to the synthesis and bioengineering of circular peptides and proteins. *Peptide Science* **94**, 414-422 (2010).

85. Hemu, X., Qiu, Y. & Tam, J.P. Peptide macrocyclization through amide-to-amide transpeptidation. *Tetrahedron* **70**, 7707-7713 (2014).

86. Marcelino, A.M.C. & Gierasch, L.M. Roles of β-Turns in Protein Folding: From Peptide Models to Protein Engineering. *Biopolymers* **89**, 380-391 (2008).

87. Weißhoff, H. et al. Mimicry of βII′-turns of proteins in cyclic pentapeptides with one and without d-amino acids. *European Journal of Biochemistry* **259**, 776-789 (1999).

88. Chatterjee, J., Mierke, D. & Kessler, H. N-Methylated Cyclic Pentaalanine Peptides as Template Structures. *Journal of the American Chemical Society* **128**, 15164-15172 (2006).

89. Pelay-Gimeno, M., Glas, A., Koch, O. & Grossmann, T.N. Structure-Based Design of Inhibitors of Protein–Protein Interactions: Mimicking Peptide Binding Epitopes. *Angewandte Chemie International Edition* **54**, 8896-8927 (2015).

90. Fernández-Llamazares, A.I. et al. The Backbone N-(4-Azidobutyl) Linker for the Preparation of Peptide Chimera. *Organic Letters* **15**, 4572-4575 (2013).

91. Valverde, I.E., Lecaille, F., Lalmanach, G., Aucagne, V. & Delmas, A.F. Synthesis of a Biologically Active Triazole-Containing Analogue of Cystatin A Through Successive Peptidomimetic Alkyne–Azide Ligations. *Angewandte Chemie* **124**, 742-746 (2012).

92. Skropeta, D., Jolliffe, K.A. & Turner, P. Pseudoprolines as Removable Turn Inducers: Tools for the Cyclization of Small Peptides. *The Journal of Organic Chemistry* **69**, 8804-8809 (2004).

93. Sonntag, L.-S., Schweizer, S., Ochsenfeld, C. & Wennemers, H. The “Azido Gauche Effect”Implications for the Conformation of Azidoprolines. *Journal of the American Chemical Society* **128**, 14697-14703 (2006).

94. Dumy, P. et al. Pseudo-Prolines as a Molecular Hinge: Reversible Induction of cis Amide Bonds into Peptide Backbones. *Journal of the American Chemical Society* **119**, 918-925 (1997).

95. Loughlin, W.A., Tyndall, J.D.A., Glenn, M.P., Hill, T.A. & Fairlie, D.P. Update 1 of: Beta-Strand Mimetics. *Chemical Reviews* **110**, PR32-PR69 (2010).

96. Cheng, P.-N., Pham, J.D. & Nowick, J.S. The Supramolecular Chemistry of β-Sheets. *Journal of the American Chemical Society* **135**, 5477-5492 (2013).

97. Madala, P.K., Tyndall, J.D.A., Nall, T. & Fairlie, D.P. Update 1 of: Proteases Universally Recognize Beta Strands In Their Active Sites. *Chemical Reviews* **110**, PR1-PR31 (2010).

98. Hammond, M.C., Harris, B.Z., Lim, W.A. & Bartlett, P.A. β Strand Peptidomimetics as Potent PDZ Domain Ligands. *Chemistry & Biology* **13**, 1247-1251 (2006).

99. Yin, H. & Hamilton, A.D. Strategies for Targeting Protein–Protein Interactions With Synthetic Agents. *Angewandte Chemie International Edition* **44**, 4130-4163 (2005).

100. Sibanda, B.L. & Thornton, J.M. [beta]-Hairpin families in globular proteins. *Nature* **316**, 170-174 (1985).

101. Nowick, J.S., Smith, E.M. & Pairish, M. Artificial [small beta]-sheets. *Chemical Society Reviews* **25**, 401-415 (1996).

102. Venkatraman, J., Shankaramma, S.C. & Balaram, P. Design of Folded Peptides. *Chemical Reviews* **101**, 3131-3152 (2001).

103. Schneider, J.P. & Kelly, J.W. Templates That Induce .alpha.-Helical, .beta.-Sheet, and Loop Conformations. *Chemical Reviews* **95**, 2169-2187 (1995).

104. Chatterjee, B. et al. Designed Peptides with Homochiral and Heterochiral Diproline Templates as Conformational Constraints. *Chemistry – A European Journal* **14**, 6192-6204 (2008).

105. Gibbs, A.C., Bjorndahl, T.C., Hodges, R.S. & Wishart, D.S. Probing the Structural Determinants of Type II' β-Turn Formation in Peptides and Proteins. *Journal of the American Chemical Society* **124**, 1203-1213 (2002).

106. Haque, T.S., Little, J.C. & Gellman, S.H. Stereochemical Requirements for β-Hairpin Formation: Model Studies with Four-Residue Peptides and Depsipeptides. *Journal of the American Chemical Society* **118**, 6975-6985 (1996).

107. Fasan, R. et al. Structure–Activity Studies in a Family of β-Hairpin Protein Epitope Mimetic Inhibitors of the p53–HDM2 Protein–Protein Interaction. *ChemBioChem* **7**, 515-526 (2006).

108. Holland-Nell, K. & Meldal, M. Maintaining Biological Activity by Using Triazoles as Disufide Bond Mimetics. *Angewandte Chemie* **123**, 5310-5312 (2011).

109. Almeida, A.M., Li, R. & Gellman, S.H. Parallel β-Sheet Secondary Structure Is Stabilized and Terminated by Interstrand Disulfide Cross-Linking. *Journal of the American Chemical Society* **134**, 75-78 (2012).

110. Hanold, L.E. et al. Inhibiting EGFR Dimerization Using Triazolyl-Bridged Dimerization Arm Mimics. *PLoS ONE* **10**, e0118796 (2015).

111. Lingard, H. et al. Diphenylacetylene-Linked Peptide Strands Induce Bidirectional β-Sheet Formation. *Angewandte Chemie International Edition* **53**, 3650-3653 (2014).

112. Nowick, J.S. et al. An Artificial β-Sheet Comprising a Molecular Scaffold, a β-Strand Mimic, and a Peptide Strand1. *Journal of the American Chemical Society* **118**, 2764-2765 (1996).

113. Phillips, S.T., Rezac, M., Abel, U., Kossenjans, M. & Bartlett, P.A. “@-Tides”: The 1,2-Dihydro-3(6H)-pyridinone Unit as a β-Strand Mimic. *Journal of the American Chemical Society* **124**, 58-66 (2002).

114. Lingard, H. et al. Diphenylacetylene-Linked Peptide Strands Induce Bidirectional β-Sheet Formation. *Angewandte Chemie* **126**, 3724-3727 (2014).

115. Koch, O., Cole, J., Block, P. & Klebe, G. Secbase: Database Module To Retrieve Secondary Structure Elements with Ligand Binding Motifs. *Journal of Chemical Information and Modeling* **49**, 2388-2402 (2009).

116. Pauling, L., Corey, R.B. & Branson, H.R. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences* **37**, 205-211 (1951).

117. Cooley, R.B., Arp, D.J. & Karplus, P.A. Evolutionary Origin of a Secondary Structure: π-Helices as Cryptic but Widespread Insertional Variations of α-Helices That Enhance Protein Functionality. *Journal of Molecular Biology* **404**, 232-246 (2010).

118. Fodje, M.N. & Al-Karadaghi, S. Occurrence, conformational features and amino acid propensities for the π-helix. *Protein Engineering* **15**, 353-358 (2002).

119. Barlow, D.J. & Thornton, J.M. Helix geometry in proteins. *Journal of Molecular Biology* **201**, 601-619 (1988).

120. Berman, H.M. et al. The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242 (2000).

121. Bullock, B.N., Jochim, A.L. & Arora, P.S. Assessing Helical Protein Interfaces for Inhibitor Design. *Journal of the American Chemical Society* **133**, 14220-14223 (2011).

122. Toniolo, C. et al. Preferred conformations of peptides containing α,α-disubstituted α-amino acids. *Biopolymers* **22**, 205-215 (1983).

123. Leduc, A.-M. et al. Helix-stabilized cyclic peptides as selective inhibitors of steroid receptor–coactivator interactions. *Proceedings of the National Academy of Sciences* **100**, 11273-11278 (2003).

124. Galande, A.K. et al. Potent Inhibitors of LXXLL-Based Protein–Protein Interactions. *ChemBioChem* **6**, 1991-1998 (2005).

125. Muppidi, A. et al. Design of antiviral stapled peptides containing a biphenyl cross-linker. *Bioorganic & Medicinal Chemistry Letters* **24**, 1748-1751 (2014).

126. Spokoyny, A.M. et al. A Perfluoroaryl-Cysteine SNAr Chemistry Approach to Unprotected Peptide Stapling. *Journal of the American Chemical Society* **135**, 5946-5949 (2013).

127. Wysoczanski, P. et al. NMR Solution Structure of a Photoswitchable Apoptosis Activating Bak Peptide Bound to Bcl-xL. *Journal of the American Chemical Society* **134**, 7644-7647 (2012).

128. Phelan, J.C., Skelton, N.J., Braisted, A.C. & McDowell, R.S. A General Method for Constraining Short Peptides to an α-Helical Conformation. *Journal of the American Chemical Society* **119**, 455-460 (1997).

129. Sia, S.K., Carr, P.A., Cochran, A.G., Malashkevich, V.N. & Kim, P.S. Short constrained peptides that inhibit HIV-1 entry. *Proceedings of the National Academy of Sciences* **99**, 14664-14669 (2002).

130. Harrison, R.S. et al. Downsizing human, bacterial, and viral proteins to short water-stable alpha helices that maintain biological potency. *Proceedings of the National Academy of Sciences* **107**, 11686-11691 (2010).

131. Baskin, J.M. et al. Copper-free click chemistry for dynamic in vivo imaging. *Proceedings of the National Academy of Sciences* **104**, 16793-16797 (2007).

132. Cantel, S. et al. Synthesis and Conformational Analysis of a Cyclic Peptide Obtained via i to i+4 Intramolecular Side-Chain to Side-Chain Azide−Alkyne 1,3-Dipolar Cycloaddition. *The Journal of Organic Chemistry* **73**, 5663-5674 (2008).

133. Kawamoto, S.A. et al. Design of Triazole-Stapled BCL9 α-Helical Peptides to Target the β-Catenin/B-Cell CLL/lymphoma 9 (BCL9) Protein–Protein Interaction. *Journal of Medicinal Chemistry* **55**, 1137-1146 (2012).

134. Lau, Y.H., de Andrade, P., McKenzie, G.J., Venkitaraman, A.R. & Spring, D.R. Linear Aliphatic Dialkynes as Alternative Linkers for Double-Click Stapling of p53-Derived Peptides. *ChemBioChem* **15**, 2680-2683 (2014).

135. Schafmeister, C.E., Po, J. & Verdine, G.L. An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Peptides. *Journal of the American Chemical Society* **122**, 5891-5892 (2000).

136. Bird, G.H. et al. Stapled HIV-1 peptides recapitulate antigenic structures and engage broadly neutralizing antibodies. *Nat Struct Mol Biol* **21**, 1058-1067 (2014).

137. Hilinski, G.J. et al. Stitched α-Helical Peptides via Bis Ring-Closing Metathesis. *Journal of the American Chemical Society* **136**, 12314-12322 (2014).

138. Brown, C.J. et al. Stapled Peptides with Improved Potency and Specificity That Activate p53. *ACS Chemical Biology* **8**, 506-512 (2013).

139. Díaz-Perlas, C., Sánchez-Navarro, M., Moreno, M., Teixidó, M. & Giralt, E. Phage Display as a Tool to Discover BBB-Shuttle Peptides: Panning Against a Human Blood-Brain Barrier Cellular Model. *Peptide Science*, n/a-n/a (2016).

140. Visser, C.C. et al. Targeting liposomes with protein drugs to the blood–brain barrier in vitro. *European Journal of Pharmaceutical Sciences* **25**, 299-305 (2005).

141. Oller-Salvia, B., Sanchez-Navarro, M., Giralt, E. & Teixido, M. Blood-brain barrier shuttle peptides: an emerging paradigm for brain delivery. *Chemical Society Reviews* **45**, 4690-4707 (2016).

142. Pardridge, W.M. Molecular Trojan horses for blood–brain barrier drug delivery. *Current Opinion in Pharmacology* **6**, 494-500 (2006).

143. Oller-Salvia, B., Teixidó, M. & Giralt, E. From venoms to BBB shuttles: Synthesis and blood–brain barrier transport assessment of apamin and a nontoxic analog. *Peptide Science* **100**, 675-686 (2013).

144. Pazos, E., Vazquez, O., Mascarenas, J.L. & Eugenio Vazquez, M. Peptide-based fluorescent biosensors. *Chemical Society Reviews* **38**, 3348-3359 (2009).

145. Loving, G. & Imperiali, B. A Versatile Amino Acid Analogue of the Solvatochromic Fluorophore 4-N,N-Dimethylamino-1,8-naphthalimide: A Powerful Tool for the Study of Dynamic Protein Interactions. *Journal of the American Chemical Society* **130**, 13630-13638 (2008).

146. Risau, W. Mechanisms of angiogenesis. *Nature* **386**, 671-674 (1997).

147. Hanahan, D. & Folkman, J. Patterns and Emerging Mechanisms of the Angiogenic Switch during Tumorigenesis. *Cell* **86**, 353-364 (1996).

148. Ferrara, N. & Kerbel, R.S. Angiogenesis as a therapeutic target. *Nature* **438**, 967-974 (2005).

149. Shibuya, M. in Advances in Cancer Research (eds. George, F.V.W. & George, K.) 281-316 (Academic Press, 1995).

150. Shibuya, M. & Claesson-Welsh, L. Signal transduction by VEGF receptors in regulation of angiogenesis and lymphangiogenesis. *Experimental Cell Research* **312**, 549-560 (2006).

151. Senger, D. et al. Tumor cells secrete a vascular permeability factor that promotes accumulation of ascites fluid. *Science* **219**, 983-985 (1983).

152. Leung, D., Cachianes, G., Kuang, W., Goeddel, D. & Ferrara, N. Vascular endothelial growth factor is a secreted angiogenic mitogen. *Science* **246**, 1306-1309 (1989).

153. Carmeliet, P. et al. Synergism between vascular endothelial growth factor and placental growth factor contributes to angiogenesis and plasma extravasation in pathological conditions. *Nat Med* **7**, 575-583 (2001).

154. Aase, K. et al. Vascular Endothelial Growth Factor-B–Deficient Mice Display an Atrial Conduction Defect. *Circulation* **104**, 358-364 (2001).

155. Dhondt, J. et al. Neuronal FLT1 receptor and its selective ligand VEGF-B protect against retrograde degeneration of sensory neurons. *The FASEB Journal* **25**, 1461-1473 (2011).

156. McColl, B.K. et al. Proprotein convertases promote processing of VEGF-D, a critical step for binding the angiogenic receptor VEGFR-2. *The FASEB Journal* **21**, 1088-1098 (2007).

157. Alitalo, K. & Carmeliet, P. Molecular mechanisms of lymphangiogenesis in health and disease. *Cancer Cell* **1**, 219-227 (2002).

158. Muller, Y.A. et al. Vascular endothelial growth factor: Crystal structure and functional mapping of the kinase domain receptor binding site. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 7192-7197 (1997).

159. Wiesmann, C. et al. Crystal Structure at 1.7 Å Resolution of VEGF in Complex with Domain 2 of the Flt-1 Receptor. *Cell* **91**, 695-704.

160. Muller, Y.A., Christinger, H.W., Keyt, B.A. & de Vos, A.M. The crystal structure of vascular endothelial growth factor (VEGF) refined to 1.93 Å resolution: multiple copy flexibility and receptor binding. *Structure* **5**, 1325-1338 (1997).

161. Fairbrother, W.J., Champe, M.A., Christinger, H.W., Keyt, B.A. & Starovasnik, M.A. 1H, 13C, and 15N backbone assignment and secondary structure of the receptor-binding domain of vascular endothelial growth factor. *Protein Science : A Publication of the Protein Society* **6**, 2250-2260 (1997).

162. Sawano, A., Takahashi, T., Yamaguchi, S., Aonuma, M. & Shibuya, M. Flt-1 but not KDR/Flk-1 tyrosine kinase is a receptor for placenta growth factor, which is related to vascular endothelial growth factor. *Cell Growth Differ* **7**, 213-221 (1996).

163. Wu, Y. et al. The vascular endothelial growth factor receptor (VEGFR-1) supports growth and survival of human breast carcinoma. *International Journal of Cancer* **119**, 1519-1529 (2006).

164. Kendall, R.L. & Thomas, K.A. Inhibition of vascular endothelial cell growth factor activity by an endogenously encoded soluble receptor. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 10705-10709 (1993).

165. Ambati, B.K. et al. Corneal avascularity is due to soluble VEGF receptor-1. *Nature* **443**, 993-997 (2006).

166. Sa-nguanraksa, D. & O-charoenrat, P. The Role of Vascular Endothelial Growth Factor A Polymorphisms in Breast Cancer. *International Journal of Molecular Sciences* **13**, 14845 (2012).

167. Hurwitz , H. et al. Bevacizumab plus Irinotecan, Fluorouracil, and Leucovorin for Metastatic Colorectal Cancer. *New England Journal of Medicine* **350**, 2335-2342 (2004).

168. Sandler , A. et al. Paclitaxel–Carboplatin Alone or with Bevacizumab for Non–Small-Cell Lung Cancer. *New England Journal of Medicine* **355**, 2542-2550 (2006).

169. Peak, S.J. & Levin, V.A. Role of bevacizumab therapy in the management of glioblastoma. *Cancer management and research* **2**, 97-104 (2010).

170. Wada, S. et al. Rationale for Antiangiogenic Cancer Therapy with Vaccination Using Epitope Peptides Derived from Human Vascular Endothelial Growth Factor Receptor 2. *Cancer Research* **65**, 4939-4946 (2005).

171. Bais, C. et al. PlGF Blockade Does Not Inhibit Angiogenesis during Primary Tumor Growth. *Cell* **141**, 166-177 (2010).

172. Van de Veire, S. et al. Further Pharmacological and Genetic Evidence for the Efficacy of PlGF Inhibition in Cancer and Eye Disease. *Cell* **141**, 178-190 (2010).

173. Ferrara, N. VEGF: an update on biological and therapeutic aspects. *Current opinion in biotechnology* **11**, 617-624 (2000).

174. Rubio, S. et al. Diminished oligomerization in the synthesis of new anti-angiogenic cyclic peptide using solution instead of solid-phase cyclization. *Peptide Science* **106**, 368-375 (2016).

175. Reille-Seroussi, M. et al. Vascular Endothelial Growth Factor Peptide Ligands Explored by Competition Assay and Isothermal Titration Calorimetry. *Biochemistry* **54**, 5147-5156 (2015).

176. Gautier, B. et al. Targeting the Proangiogenic VEGF-VEGFR Protein-Protein Interface with Drug-like Compounds by In Silico and In Vitro Screening. *Chemistry & Biology* **18**, 1631-1639 (2011).

177. Ovchinnikov, Y.A.I., V. T. . in The proteins (ed. Hill, H.N.a.R.L.) (Academic Press, INC., London, 1982).

178. Cacho, R.A., Jiang, W., Chooi, Y.-H., Walsh, C.T. & Tang, Y. Identification and Characterization of the Echinocandin B Biosynthetic Gene Cluster from Emericella rugulosa NRRL 11440. *Journal of the American Chemical Society* **134**, 16781-16790 (2012).

179. Prompanya, C. et al. A New Cyclic Hexapeptide and a New Isocoumarin Derivative from the Marine Sponge-Associated Fungus Aspergillus similanensis KUFA 0013. *Marine Drugs* **13**, 1432 (2015).

180. Wang, C.K., Northfield, S.E., Huang, Y.-H., Ramos, M.C. & Craik, D.J. Inhibition of tau aggregation using a naturally-occurring cyclic peptide scaffold. *European Journal of Medicinal Chemistry* **109**, 342-349 (2016).

181. Belda, I. (Universitat Ramón Llull, 2001).

182. Chou, P.Y. & Fasman, G.D. in Advances in Enzymology and Related Areas of Molecular Biology 45-148 (John Wiley & Sons, Inc., 2006).

183. Kyte, J. & Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105-132 (1982).

184. Glaser, F., Steinberg, D.M., Vakser, I.A. & Ben-Tal, N. Residue frequencies and pairing preferences at protein–protein interfaces. *Proteins: Structure, Function, and Bioinformatics* **43**, 89-102 (2001).

185. Spengler, J., Jiménez, J.C., Burger, K., Giralt, E. & Albericio, F. Abbreviated nomenclature for cyclic and branched homo- and hetero-detic peptides. *The Journal of Peptide Research* **65**, 550-555 (2005).

186. Vila, X. (Universitat de Barcelona, Barcelona, 2014).

187. Mías, R.A.R. (Universitat de Barcelona, 2006).

188. Pan, B. et al. Solution structure of a phage-derived peptide antagonist in complex with vascular endothelial growth factor1. *Journal of Molecular Biology* **316**, 769-787 (2002).

189. Fairbrother, W.J. et al. Novel Peptides Selected to Bind Vascular Endothelial Growth Factor Target the Receptor-Binding Site. *Biochemistry* **37**, 17754-17764 (1998).

190. Peterson, K.J. et al. A Fluorescence Polarization Assay for Identifying Ligands that Bind to Vascular Endothelial Growth Factor. *Analytical biochemistry* **378**, 8 (2008).

191. Spera, S. & Bax, A. Empirical correlation between protein backbone conformation and C.alpha. and C.beta. 13C nuclear magnetic resonance chemical shifts. *Journal of the American Chemical Society* **113**, 5490-5492 (1991).

192. Wishart, D.S. Interpreting protein chemical shift data. *Progress in Nuclear Magnetic Resonance Spectroscopy* **58**, 62-87 (2011).

193. Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. & Sykes, B.D. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *Journal of Biomolecular NMR* **5**, 67-81 (1995).

194. Schubert, M., Labudde, D., Oschkinat, H. & Schmieder, P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. *Journal of Biomolecular NMR* **24**, 149-154 (2002).

195. Shoichet, B.K. Virtual screening of chemical libraries. *Nature* **432**, 862-865 (2004).

196. Inc., C.C.G. (1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2016).

197. Woody, R.W. in Peptides, polypeptides and proteins (ed. Wiley) 338-350 (Blout, E.R., Bovey, F.A., Goodman, M., & Lotan, N. Eds., New York, 1974).

198. Gierasch, L.M., Deber, C.M., Madison, V., Niu, C.-H. & Blout, E.R. Conformations of (X-L-Pro-Y)2 cyclic hexapeptides. Preferred .beta.-turn conformers and implications for .beta. turns in proteins. *Biochemistry* **20**, 4730-4738 (1981).

199. Malešević, M., Schumann, M., Jahreis, G., Fischer, G. & Lücke, C. Design of Cyclic Peptides Featuring Proline Predominantly in the cis Conformation under Physiological Conditions. *ChemBioChem* **13**, 2122-2127 (2012).

200. Mantovani, F., Zannini, A., Rustighi, A. & Del Sal, G. Interaction of p53 with prolyl isomerases: Healthy and unhealthy relationships. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1850**, 2048-2060 (2015).

201. Blair, L.J., Baker, J.D., Sabbagh, J.J. & Dickey, C.A. The emerging role of peptidyl-prolyl isomerase chaperones in tau oligomerization, amyloid processing and Alzheimer's disease. *Journal of neurochemistry* **133**, 1-13 (2015).

202. Rostam, M.A. et al. Peptidyl-prolyl isomerases: Functionality and potential therapeutic targets in cardiovascular disease. *Clinical and Experimental Pharmacology and Physiology* **42**, 117-124 (2015).

203. Lin, Y.-J., Chu, L.-K. & Horng, J.-C. Effects of the Terminal Aromatic Residues on Polyproline Conformation: Thermodynamic and Kinetic Studies. *The Journal of Physical Chemistry B* **119**, 15796-15806 (2015).

204. D.A. Case, V.B., J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E., Duke, H.G., A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko,, T.S. Lee, S.L., T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui,, R. Salomon-Ferrer, G.S., C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. & Kollman, W.a.P.A. (University of California, San Francisco, 2014).

205. Wickstrom, L., Okur, A. & Simmerling, C. Evaluating the Performance of the ff99SB Force Field Based on NMR Scalar Coupling Data. *Biophysical Journal* **97**, 853-856 (2009).

206. Doshi, U. & Hamelberg, D. Reoptimization of the AMBER Force Field Parameters for Peptide Bond (Omega) Torsions Using Accelerated Molecular Dynamics. *The Journal of Physical Chemistry B* **113**, 16590-16595 (2009).

207. McHugh, S.M., Rogers, J.R., Yu, H. & Lin, Y.-S. Insights into How Cyclic Peptides Switch Conformations. *Journal of Chemical Theory and Computation* **12**, 2480-2488 (2016).

208. King, G.F. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert Opinion on Biological Therapy* **11**, 1469-1484 (2011).

209. de Boer, A.G. & Gaillard, P.J. Drug Targeting to the Brain. *Annual Review of Pharmacology and Toxicology* **47**, 323-355 (2007).

210. Brightman, M.W. & Reese, T.S. JUNCTIONS BETWEEN INTIMATELY APPOSED CELL MEMBRANES IN THE VERTEBRATE BRAIN. *The Journal of Cell Biology* **40**, 648-677 (1969).

211. Malakoutikhah, M., Teixidó, M. & Giralt, E. Shuttle-Mediated Drug Delivery to the Brain. *Angewandte Chemie International Edition* **50**, 7998-8014 (2011).

212. Wei, X. et al. Retro-Inverso Isomer of Angiopep-2: A Stable d-Peptide Ligand Inspires Brain-Targeted Drug Delivery. *Molecular Pharmaceutics* **11**, 3261-3268 (2014).

213. Wei, X. et al. A D-Peptide Ligand of Nicotine Acetylcholine Receptors for Brain-Targeted Drug Delivery. *Angewandte Chemie International Edition* **54**, 3023-3027 (2015).

214. LabbÉ-JulliÉ, C. et al. Binding and toxicity of apamin. *European Journal of Biochemistry* **196**, 639-645 (1991).

215. Wu, J. et al. Apamin-Mediated Actively Targeted Drug Delivery for Treatment of Spinal Cord Injury: More Than Just a Concept. *Molecular Pharmaceutics* **11**, 3210-3222 (2014).

216. Fiori, S., Pegoraro, S., Rudolph-Böhner, S., Cramer, J. & Moroder, L. Synthesis and conformational analysis of apamin analogues with natural and non-natural cystine/selenocystine connectivities. *Biopolymers* **53**, 550-564 (2000).

217. Zadran, S. et al. Fluorescence resonance energy transfer (FRET)-based biosensors: visualizing cellular dynamics and bioenergetics. *Applied Microbiology and Biotechnology* **96**, 895-902 (2012).

218. Burr, M. & Koshland, D.E. USE OF „REPORTER GROUPS" IN STRUCTURE-FUNCTION STUDIES OF PROTEINS. *Proceedings of the National Academy of Sciences of the United States of America* **52**, 1017-1024 (1964).

219. Wang, H.-Q., Wu, Z., Tang, L.-J., Yu, R.-Q. & Jiang, J.-H. Fluorescence protection assay: a novel homogeneous assay platform toward development of aptamer sensors for protein detection. *Nucleic Acids Research* (2011).

220. Zhao, W. et al. Cell-surface sensors for real-time probing of cellular environments. *Nat Nano* **6**, 524-531 (2011).

221. Paige, J.S., Nguyen-Duc, T., Song, W. & Jaffrey, S.R. Fluorescence Imaging of Cellular Metabolites with RNA. *Science* **335**, 1194-1194 (2012).

222. Goguen, B.N., Loving, G.S. & Imperiali, B. Development of a fluorogenic sensor for activated Cdc42. *Bioorganic & Medicinal Chemistry Letters* **21**, 5058-5061 (2011).

223. Choulier, L. et al. A peptide-based fluorescent ratiometric sensor for quantitative detection of proteins. *Analytical Biochemistry* **401**, 188-195 (2010).

224. Pazos, E. et al. Rational design of a cyclin A fluorescent peptide sensor. *Organic & Biomolecular Chemistry* **9**, 7629-7632 (2011).

225. Renard, M. et al. Knowledge-based Design of Reagentless Fluorescent Biosensors from Recombinant Antibodies. *Journal of Molecular Biology* **318**, 429-442 (2002).

226. Miranda, F.F., Brient-Litzler, E., Zidane, N., Pecorari, F. & Bedouelle, H. Reagentless fluorescent biosensors from artificial families of antigen binding proteins. *Biosensors and Bioelectronics* **26**, 4184-4190 (2011).

227. Gulyani, A. et al. A biosensor generated via high-throughput screening quantifies cell edge Src dynamics. *Nat Chem Biol* **7**, 437-444 (2011).

228. Kummer, L. et al. Knowledge-Based Design of a Biosensor to Quantify Localized ERK Activation in Living Cells. *Chemistry & Biology* **20**, 847-856 (2013).

229. Abe, R. et al. "Quenchbodies": Quench-Based Antibody Probes That Show Antigen-Dependent Fluorescence. *Journal of the American Chemical Society* **133**, 17386-17394 (2011).

230. Brient-Litzler, E., Plückthun, A. & Bedouelle, H. Knowledge-based design of reagentless fluorescent biosensors from a designed ankyrin repeat protein. *Protein Engineering Design and Selection* **23**, 229-241 (2010).

231. Vivian, J.T. & Callis, P.R. Mechanisms of Tryptophan Fluorescence Shifts in Proteins. *Biophysical Journal* **80**, 2093-2109.

232. Loving, G.S., Sainlos, M. & Imperiali, B. Monitoring protein interactions and dynamics with solvatochromic fluorophores. *Trends in Biotechnology* **28**, 73-83.

233. Chao, G. et al. Isolating and engineering human antibodies using yeast surface display. *Nat. Protocols* **1**, 755-768 (2006).

234. Clackson, T., Hoogenboom, H.R., Griffiths, A.D. & Winter, G. Making antibody fragments using phage display libraries. *Nature* **352**, 624-628 (1991).

235. Roberts, R.W. & Szostak, J.W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proceedings of the National Academy of Sciences* **94**, 12297-12302 (1997).

236. Hanes, J. & Plückthun, A. In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences* **94**, 4937-4942 (1997).

237. Binz, H.K. et al. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat Biotech* **22**, 575-582 (2004).

238. Koide, A., Bailey, C.W., Huang, X. & Koide, S. The fibronectin type III domain as a scaffold for novel binding proteins1. *Journal of Molecular Biology* **284**, 1141-1151 (1998).

239. Liao, X., Rabideau, A.E. & Pentelute, B.L. Delivery of Antibody Mimics into Mammalian Cells via Anthrax Toxin Protective Antigen. *ChemBioChem* **15**, 2458-2466 (2014).

240. de Picciotto, S., Imperiali, B., Griffith, L.G. & Wittrup, K.D. Equilibrium and dynamic design principles for binding molecules engineered for reagentless biosensors. *Analytical Biochemistry* **460**, 9-15 (2014).

241. Spicer, C.D. & Davis, B.G. Selective chemical protein modification. *Nat Commun* **5** (2014).

242. Rashidian, M., Dozier, J.K. & Distefano, M.D. Enzymatic Labeling of Proteins: Techniques and Approaches. *Bioconjugate Chemistry* **24**, 1277-1294 (2013).

243. Mayor, U. et al. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421**, 863-867 (2003).

244. Neuweiler, H. et al. Downhill versus Barrier-Limited Folding of BBL 2: Mechanistic Insights from Kinetics of Folding Monitored by Independent Tryptophan Probes. *Journal of Molecular Biology* **387**, 975-985 (2009).

245. Zhang, L. et al. Mapping hydration dynamics around a protein surface. *Proceedings of the National Academy of Sciences* **104**, 18461-18466 (2007).

246. Alder, N.N., Sutherland, J., Buhring, A.I., Jensen, R.E. & Johnson, A.E. Quaternary Structure of the Mitochondrial TIM23 Complex Reveals Dynamic Association between Tim23p and Other Subunits. *Molecular Biology of the Cell* **19**, 159-170 (2008).
247. Nguyen, A.H., Nguyen, V.T., Kamio, Y. & Higuchi, H. Single-Molecule Visualization of Environment-Sensitive Fluorophores Inserted into Cell Membranes by Staphylococcal γ-Hemolysin. *Biochemistry* **45**, 2570-2576 (2006).
248. Amiss, T.J., Sherman, D.B., Nycz, C.M., Andaluz, S.A. & Pitner, J.B. Engineering and rapid selection of a low-affinity glucose/galactose-binding protein for a glucose biosensor. *Protein Science* **16**, 2350-2359 (2007).
249. Dattelbaum, J.D. et al. Analysis of allosteric signal transduction mechanisms in an engineered fluorescent maltose biosensor. *Protein Science* **14**, 284-291 (2005).
250. Shvadchak, V.V., Klymchenko, A.S., de Rocquigny, H. & Mély, Y. Sensing peptide–oligonucleotide interactions by a two-color fluorescence label: application to the HIV-1 nucleocapsid protein. *Nucleic Acids Research* **37**, e25 (2009).
251. Zhu, J. & Pei, D. A LuxP-Based Fluorescent Sensor for Bacterial Autoinducer II. *ACS Chemical Biology* **3**, 110-119 (2008).
252. Walkup, G.K. & Imperiali, B. Fluorescent Chemosensors for Divalent Zinc Based on Zinc Finger Domains. Enhanced Oxidative Stability, Metal Binding Affinity, and Structural and Functional Characterization. *Journal of the American Chemical Society* **119**, 3443-3450 (1997).
253. Savalli, N., Kondratiev, A., Toro, L. & Olcese, R. Voltage-dependent conformational changes in human Ca2+- and voltage-activated K+ channel, revealed by voltage-clamp fluorometry. *Proceedings of the National Academy of Sciences* **103**, 12619-12624 (2006).
254. Mannuzzu, L.M., Moronne, M.M. & Isacoff, E.Y. Direct Physical Measure of Conformational Rearrangement Underlying Potassium Channel Gating. *Science* **271**, 213-216 (1996).
255. Simard, J.R. et al. A new screening assay for allosteric inhibitors of cSrc. *Nat Chem Biol* **5**, 394-396 (2009).
256. Yeh, R.-H., Yan, X., Cammer, M., Bresnick, A.R. & Lawrence, D.S. Real Time Visualization of Protein Kinase Activity in Living Cells. *Journal of Biological Chemistry* **277**, 11527-11532 (2002).
257. Tsutsumi, H. et al. Fluorogenically Active Leucine Zipper Peptides as Tag–Probe Pairs for Protein Imaging in Living Cells. *Angewandte Chemie* **121**, 9328-9330 (2009).
258. Nalbant, P., Hodgson, L., Kraynov, V., Toutchkine, A. & Hahn, K.M. Activation of Endogenous Cdc42 Visualized in Living Cells. *Science* **305**, 1615-1619 (2004).
259. Lanfermeijer, F.C., Detmers, F.J.M., Konings, W.N. & Poolman, B. On the binding mechanism of the peptide receptor of the oligopeptide transport system of Lactococcus lactis. *The EMBO Journal* **19**, 3649-3656 (2000).

260. de Picciotto, S. et al. Design principles for reagentless biosensors: Specific fluorophore/analyte binding and minimization of fluorophore/scaffold interactions. *Journal of Molecular Biology* (2016).

261. Baumann, H., Knapp, S., Lundback, T., Ladenstein, R. & Hard, T. Solution structure and DNA-binding properties of a thermostable protein from the archaeon Sulfolobus solfataricus. *Nat Struct Mol Biol* **1**, 808-819 (1994).

262. Baumann, H., Knapp, S., Karshikoff, A., Ladenstein, R. & Härd, T. DNA-binding Surface of the Sso7d Protein fromSulfolobus solfataricus. *Journal of Molecular Biology* **247**, 840-846 (1995).

263. Edmondson, S.P. & Shriver, J.W. in Methods in Enzymology 129-145 (Academic Press, 2001).

264. Knapp, S. et al. Thermal Unfolding of the DNA-binding Protein Sso7d from the HyperthermophileSulfolobus solfataricus. *Journal of Molecular Biology* **264**, 1132-1144 (1996).

265. Gera, N., Hussain, M., Wright, R.C. & Rao, B.M. Highly Stable Binding Proteins Derived from the Hyperthermophilic Sso7d Scaffold. *Journal of Molecular Biology* **409**, 601-616 (2011).

266. Gera, N., Hill, A.B., White, D.P., Carbonell, R.G. & Rao, B.M. Design of pH Sensitive Binding Proteins from the Hyperthermophilic Sso7d Scaffold. *PLoS ONE* **7**, e48928 (2012).

267. Ackerman, M. et al. Highly Avid Magnetic Bead Capture: An Efficient Selection Method for de novo Protein Engineering Utilizing Yeast Surface Display. *Biotechnology progress* **25**, 774-783 (2009).

268. Loving, G. & Imperiali, B. Thiol-Reactive Derivatives of the Solvatochromic 4-N,N-Dimethylamino-1,8-naphthalimide Fluorophore: A Highly Sensitive Toolset for the Detection of Biomolecular Interactions. *Bioconjugate Chemistry* **20**, 2133-2141 (2009).

269. Green, A.A. & Hughes, W.L. in Methods in Enzymology 67-90 (Academic Press, 1955).

270. de Nobel, H., Pike, J., Lipke, P.N. & Kurjan, J. Genetics of a-agglutunin function in Saccharomyces cerevisiae. *Molecular and General Genetics MGG* **247**, 409-415 (1995).

271. Griffiths, A.D. & Duncan, A.R. Strategies for selection of antibodies by phage display. *Current Opinion in Biotechnology* **9**, 102-108 (1998).

272. Cherf, G.M. & Cochran, J.R. Applications of yeast surface display for protein engineering. *Methods in molecular biology (Clifton, N.J.)* **1319**, 155-175 (2015).

273. Hwang, T.L. & Shaka, A.J. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A* **112**, 275-279 (1995).

274. Bax, A. & Davis, D.G. MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *Journal of Magnetic Resonance (1969)* **65**, 355-360 (1985).

275. Bax, A. & Freeman, R. Investigation of complex networks of spin-spin coupling by two-dimensional NMR. *Journal of Magnetic Resonance (1969)* **44**, 542-561 (1981).

276. Jeener, J., Meier, B.H., Bachmann, P. & Ernst, R.R. Investigation of exchange processes by two-dimensional NMR spectroscopy. *The Journal of Chemical Physics* **71**, 4546-4553 (1979).

277. A. T. Brünger, P.D.A., G. M. Clore, W. L. DeLano, P. Gros, R. W., Grosse-Kunstleve, J.S.J., J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, & L. M. Rice, T.S., G. L. Warren. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-921 (1998).

278. Berjanskii, M.V., Neal, S. & Wishart, D.S. PREDITOR: a web server for predicting protein torsion angle restraints. *Nucleic Acids Research* **34**, W63-W69 (2006).

279. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**, 283-291 (1993).

280. Patriksson, A. & van der Spoel, D. A temperature predictor for parallel tempering simulations. *Physical Chemistry Chemical Physics* **10**, 2073-2077 (2008).

281. Mongan, J., Simmerling, C., McCammon, J.A., Case, D.A. & Onufriev, A. Generalized Born model with a simple, robust molecular volume correction. *Journal of chemical theory and computation* **3**, 156-169 (2007).

282. Weiser, J., Shenkin, P.S. & Still, W.C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *Journal of Computational Chemistry* **20**, 217-230 (1999).

283. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. & Haak, J.R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684-3690 (1984).

284. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).

285. Hunter, J.D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **9**, 90-95 (2007).

286. Wishart, D.S. et al. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *Journal of Biomolecular NMR* **6**, 135-140 (1995).

287. Sadowsky, J.D. et al. Chimeric (α/β + α)-Peptide Ligands for the BH3-Recognition Cleft of Bcl-xL: Critical Role of the Molecular Scaffold in Protein Surface Recognition. *Journal of the American Chemical Society* **127**, 11966-11968 (2005).

288. Banks, P. & Harvey, M. Considerations for Using Fluorescence Polarization in the Screening of G Protein-Coupled Receptors. *Journal of Biomolecular Screening* **7**, 111-117 (2002).

289. Zhang, R. et al. Fluorescence polarization assay and inhibitor design for MDM2/p53 interaction. *Analytical Biochemistry* **331**, 138-146 (2004).

290. Roehrl, M.H.A., Wang, J.Y. & Wagner, G. A General Framework for Development and Data Analysis of Competitive High-Throughput Screens for Small-Molecule Inhibitors of Protein−Protein Interactions by Fluorescence Polarization. *Biochemistry* **43**, 16056-16066 (2004).

SUMMARY IN SPANISH

INTRODUCCIÓN

Reconocimiento molecular de superficies proteicas

En química y biología, el reconocimiento molecular hace referencia a la interacción específica entre dos o más moléculas a través de uniones complementarias no covalentes, como pueden ser: puentes de hidrogeno, coordinación de metales, fuerzas de van der Waals, e interacciones hidrofóbicas o electroestáticas de tipo π-π. Los procesos de reconocimiento entre entidades biológicas, entre las cuales se incluyen las interacciones entre: receptores y ligandos, antígenos y anticuerpos, ácidos nucleicos y proteínas, proteínas y proteínas, encimas y sustratos y ácidos nucleicos entre ellos; guían las interacciones selectivas esenciales para la vida.

Las proteínas tienen un papel muy importante en la actividad celular, incluyendo funciones de tipo: estructural, mecánicas, bioquímicas y de señalización celular. Los procesos de reconocimiento molecular de las interacciones proteína-proteína gobiernan la afinidad y la especificad de la formación del complejo y determinan su función biológica. Ejemplos de vías medidas por interacciones entre proteínas incluyen: expresión génica, proliferación, comunicación intracelular y apoptosis. Numerosas enfermedades tienen lugar debido al mal funcionamiento de vías gobernadas por interacciones particulares entre proteínas. Hasta la fecha, solo unas pocas interacciones proteína-proteína han sido sujeto de una iniciativa de desarrollo de fármacos. Aun así, debido a las numerosas oportunidades que presenta la modulación de las interacciones entre proteínas, este campo ha ido adquiriendo un creciente interés científico durante la última década, y se han logrado varios avances en grupos de investigación tanto en la vertiente académica como en la industrial.

Las interacciones entre proteínas se caracterizan fundamentalmente como estables o transitorias. Las interacciones estables son aquellas en las que los entes individuales no se observan "*in vivo*". Por el contrario, las interacciones transitorias, que controlan la mayoría de los procesos celulares, son reversibles y típicamente requieren de una serie de condiciones que promueva la interacción, tales como fosforilación, cambios conformacionales o localización en áreas discretas de la célula. En las interacciones transitorias los individuos que forman la unión y el complejo están en equilibrio dinámico, aunque con una gran diversidad de afinidades y constantes cinéticas. Las proteínas se unen entre ellas mediante la combinación de enlaces hidrofóbicos, fuerzas van der Waals, y puentes salinos en regiones específicas de cada proteína. Estas regiones pueden ser pequeñas hendeduras o grandes superficies y pueden estar constituidas por solo unos pocos residuos o abarcar cientos de aminoácidos, y la fuerza de la unión está influenciada por el tamaño de estas regiones donde se produce la interacción. En este sentido, las interacciones proteína-proteína, puede clasificarse estructuralmente como mediadas tanto por la interacción de dos dominios proteicos (dominio-dominio) o por la interacción entre una secuencia lineal de residuos de uno de los individuos y el dominio del otro (péptido-dominio) (**Figura i**). En las interacciones de tipo péptido-dominio, las cuales constituyen el 40% de las interacciones proteína-proteína conocidas, el péptido que interacciona suele adoptar una estructura secundaria, típicamente α-helice, aunque también puede ser completamente desordenado.

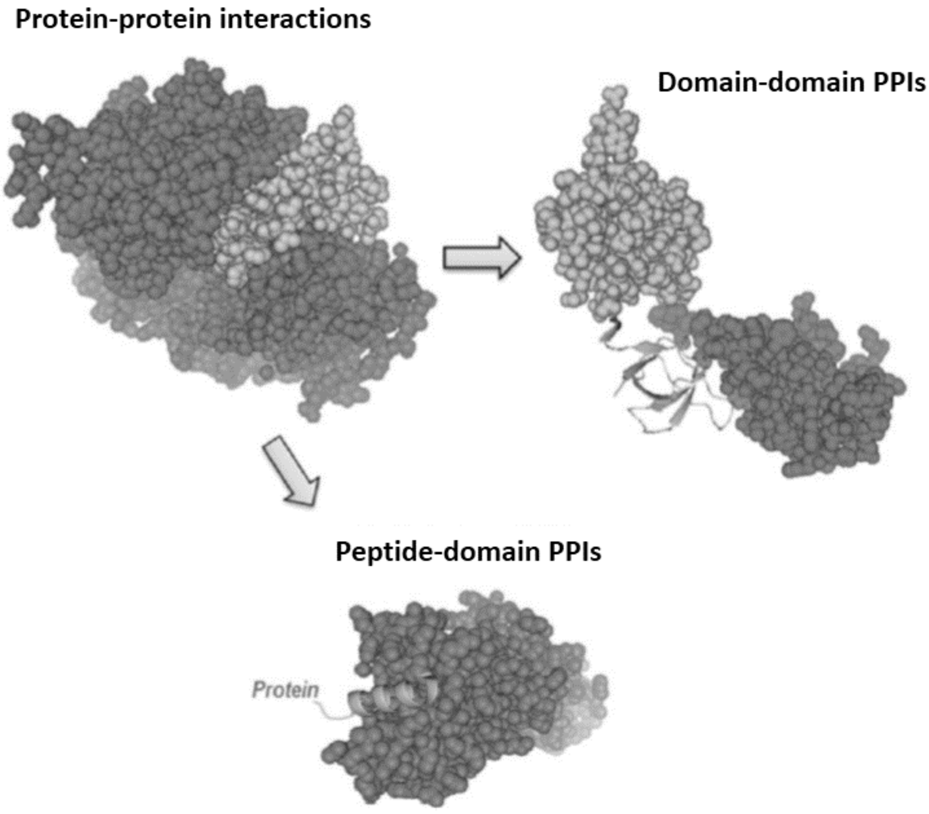


Figura i. *Representación esquemática de las interacciones proteína-proteína que pueden ser mediadas por dominios o mediadas por péptidos.*

Durante algún tiempo, los “targets” basados en interacciones proteína-proteína se consideraban como objetivos de alto riesgo, difíciles e incluso imposibles de ser tratadas con fármacos, debido al hecho que la mayoría de las interfaces involucradas en interacciones proteína-proteína son planas y no presentan cavidades hidrofóbicas bien definidas que pudieran acomodar una pequeña molécula como ligando.

Los residuos que se encuentran en las interfaces de las interacciones proteína-proteína no contribuyen por igual en la interacción. Hay una colección de pequeñas cavidades de unión repartidas a lo largo de toda la superficie de interacción. Se ha demostrado que solo algunas de estas interacciones, conocidas como "hot spots" son esenciales para la afinidad debido a que son el contribuyente mayoritario en la energía de interacción total. Un "hot spot" se define como un residuo sobre el cual la sustitución por alanina provoca un descenso significante en la energía libre de unión ($\Delta\Delta$Gbinding > 1.5 kcal/mol). La composición de aminoácidos de los "hot spots" se encuentra enriquecida en aminoácidos voluminosos como triptófano, arginina y tirosina. La prevalencia de estos "hot spots" tanto en interacciones mediadas por dominio-domino como en las mediadas por péptido-dominio hacen que sea teóricamente factible la inhibición de estas interacciones proteicas mediante ligandos con un tamaño suficiente para interaccionar simultáneamente con múltiples "hot spots". Se estima que más del 50% de las interacciones dominio-dominio entre proteínas están mediadas por un segmento compuesto por "hot spot" que contribuye a la mayor parte de la energía de interacción.

Ejemplos representativos de interacciones proteína-proteína que han sido exitosamente moduladas son las de: el dominio de p53 con MDM2, las proteínas Bcl-xL con los péptidos BH3 y las IL-2 con sus receptores IL2Rα.

Péptidos como herramientas químicas para el reconocimiento de superficies proteicas

La modulación de interacciones proteína-proteína se ha llevado a cabo mediante diferentes aproximaciones, en particular basándose en el uso de pequeñas moléculas, ya sea de origen natural u obtenidas mediante síntesis orgánica. Sendas favorables se desarrollaron en los ochenta para abordar el uso de pequeñas moléculas como inhibidores de interacciones proteína-proteína. Basándose en la estructura de los complejos, varios grupos de investigación desarrollaron métodos computacionales para la predicción de "hot spots". En promedio, estos métodos tiene una tasa de éxito del 80% en sus predicciones y combinándolos se puede incluso mejorar la predicción. Otras aproximaciones como la aproximación basada en el descubrimiento de ligandos basados en fragmentos permiten acceder a un mayor espacio químico en menor tiempo. Aunque los “hits” obtenidos de los fragmentos iniciales son típicamente ligandos de unión débil, dos o más fragmentos se pueden unir covalentemente para obtener ligandos de tamaño moderado con elevada afinidad. Aun así, la mayor dificultad que se encuentra en el desarrollo de moduladores basados en moléculas pequeñas deriva de las extensas superficies de contacto involucradas en las interacciones entre proteínas (~1500-3000 $Å^2$) cuando se comparan con las que se observan en interacciones entre pequeñas moléculas y proteínas (~300-1000 $Å^2$). Además, como ya hemos comentado, estas superficies tienden a ser planas y carecen de cavidades hidrofóbicas, lo cual dificulta aún más el uso de pequeñas moléculas como moduladoras. Es por eso que los investigadores han mostrado una creciente atención hacia la posibilidad de usar alternativas más naturales como proteínas, anticuerpos o péptidos capaces de explorar extensas superficies proteicas.

Estas extensas superficies con topologías complejas se han modulado de forma efectiva con fármacos biológicos basados en proteínas. Debido a su mayor tamaño y su conformación tridimensional bien definida, los fármacos biológicos unen a superficies proteicas con elevada afinidad y remarcada selectividad. El desarrollo de anticuerpos monoclonales que compiten con interacciones proteína-proteína ha adquirido una creciente importancia en la última década. Esta aproximación ha sido muy exitosa teniendo en cuenta que cada año decenas de nuevos anticuerpos entran en ensayos clínicos o son aprobados por la FDA para su uso como fármacos o para el diagnóstico de enfermedades. Aun así, la mayor limitación de estos fármacos biológicos sigue siendo la dificultad de introducir estas proteínas en el interior de las células. Por eso, su aplicación sigue estando limitada al tratamiento de dianas extracelulares tales como receptores de membrana y proteínas solubles.

Los péptidos, que se distinguen de las proteínas basándose en su menor tamaño (50 aminoácidos o menos), median varias funciones biológicas esenciales, tales como transducción, regulación del ritmo cardiaco, consumo de alimentos y crecimiento. Péptidos naturales como la insulina, la oxitocina y la somatostatina; así como péptidos sintéticos son actualmente exitosos fármacos que están en el mercado, lo cual destaca el gran potencial de mercado de estos fármacos.

Los péptidos terapéuticos normalmente derivan de tres fuentes: (i) péptidos naturales producidos por plantas, animales o humanos; (ii) péptidos asilados de librerías genéticas y recombinantes; y (iii) péptidos descubiertos a través de librerías químicas. La producción de péptidos terapéuticos sintéticos es ahora posible en la industria farmacéutica gracias a los recientes avances en la síntesis de péptidos en fase sólida, inicialmente desarrollada por Merrifield. Hoy en día, la síntesis de péptidos en fase sólida en las primeras etapas de la investigación preclínica y la producción de ingredientes activos farmacéuticos basados en péptidos.

Los péptidos presentan varias ventajas sobre las pequeñas moléculas y las proteínas: (i) flexibilidad, la cual se traduce en adaptabilidad a superficies extensas; (ii) fácil modularidad, que incrementa la diversidad estructural y en consecuencia otorga una mayor selectividad y potencia; (iii) tamaño, que se traduce en limitación de su acumulación en tejidos; y (iv) completa biocompatibilidad, que significa baja toxicidad en humanos. Las dos últimas características son enormemente deseables debido al creciente interés en las interacciones proteína-proteína como dianas terapéuticas.

Sin embargo, el progreso hacia el desarrollo de péptidos terapéuticos para modular interacciones proteicas se ve obstaculizado por los siguientes inconvenientes de estas moléculas: baja estabilidad frente a la degradación por encimas proteolíticas del sistema digestivo y del plasma sanguíneo; rápida eliminación del sistema circulatorio; dificultad para atravesar barreras fisiológicas; y potencial inmunogenicidad. A pesar de estas limitaciones, un gran número de péptidos han llegado a ser exitosos moduladores de interacciones proteína-proteína y el gran esfuerzo realizado para abordar este cuello de botella que perjudica su uso como fármacos es impresionante. Además de avances que aumentan su permeabilidad, varias modificaciones químicas se han introducido en estas moléculas para minimizar su degradación proteolítica y mejorar su biodisponibilidad.

Hoy en día, el aumento de la estabilidad de la conformación activa así como la mejora de la estabilidad frente a proteólisis son los mayores objetivos de la introducción de modificaciones en péptidos. Entre ellas encontramos dos grupos mayoritarios: (i) las modificaciones del esqueleto y (ii) las ciclaciones (**Figura ii**).

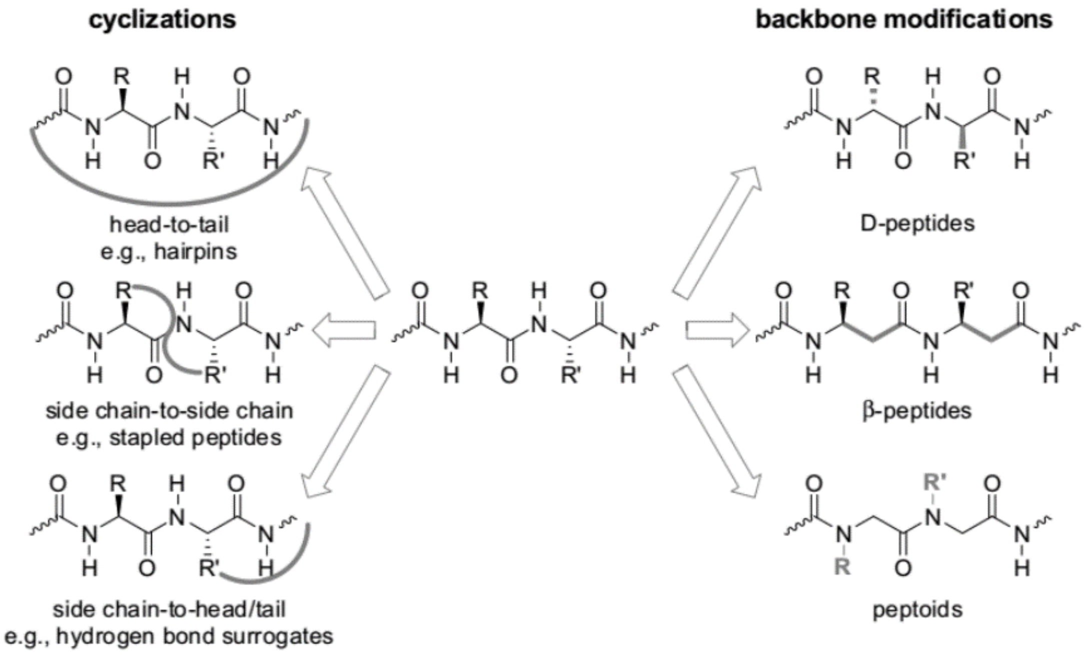


Figura ii. *Modificaciones de péptidos más frecuentes para su uso como inhibidores de interacciones proteína-proteína.*

El primer grupo de modificaciones en péptidos se basa en realizar cambios en el esqueleto, normalmente alterna las propiedades del compuesto en más profundidad y permite la obtención de secuencias y estructuras tridimensionales que difieren significativamente del fragmento original. Las posibilidades más usadas son las siguientes: (i) variaciones de la estereoquímica (uso de D-aminoácidos), (ii) extensión del esqueleto (uso de β-aminoácidos) y (iii) cambio de la posición de las cadenas laterales del carbono alfa a la amida del átomo de nitrógeno (peptoides).

El extensivo análisis de la información estructural disponible sobre interacciones proteína-proteína demuestra que la mayoría de estas uniones está normalmente mediada por varios dominios estructurales, entre los cuales las hélices α son el motivo estructural más presente en estas interacciones. En el caso de los ligandos peptídicos, se cree que su conformación determina su bioactividad y biodisponibilidad. Por eso, es intuitivo empezar a diseñar péptidos imitando cuidadosamente elementos estructurales como giros, hélices α o láminas β presentes en las interfaces proteicas.

La aproximación basada en la ciclación para la obtención de péptidos cíclicos es uno de los métodos mayormente usados para el desarrollo de inhibidores de interacciones proteína-proteína. Varias estrategias como la sustitución de enlaces de hidrogeno, el grapado o las horquillas se desarrollaron para estabilizar giros, hélices y conformaciones extendidas en péptidos. La combinación de las metodologías de ciclación con otras estrategias para fijar la conformación deseada de los péptidos es un objetivo que hace tiempo se persigue en ciencias químicas y no solo relativo a inhibir interacciones proteícas. Estos péptidos que contienen ciertas constricciones pueden tener actividades biológicas similares a las proteínas, permitiendo su uso como sondas biológicas, fármacos, kits de diagnóstico y vacunas.

Péptidos cíclicos como agentes terapéuticos y herramientas bioquímicas

Algunos ejemplos de péptidos cíclicos usados en clínica son la gramicidina, la tirocidina, la ciclosporina o la vancomicina. Normalmente los péptidos cíclicos presentan una mejor actividad biológica que sus precursores lineales debido a la rigidez conformacional, que disminuye el termino entrópico de la energía libre de Gibbs. Otro beneficio de las estructuras cíclicas es su mayor resistencia a la hidrolisis provocada por exopeptidasas debido a la pérdida de sus extremos amino y carboxi terminal. Los péptidos cíclicos pueden ser incluso resistentes a endopeptidasas, debido a la mayor restricción de su estructura. Además, algunos péptidos cíclicos, aunque no todos, son capaces atravesar membranas celulares.

Se prevé que los péptidos cíclicos puedan superar algunas de las limitaciones que presentan los péptidos lineales abordando problemas biológicos como la hidrolisis enzimática o la pobre biodisponibilidad oral. El creciente interés por los fármacos peptídicos, así como la eficacia tanto *“in vitro”* como “*in vivo*” de los péptidos cíclicos, proporciona amplias oportunidades para el desarrollo de péptido cíclicos para el tratamiento de varias enfermedades.

CAPÍTULO 1. PÉPTIDOS HEXACÍCLICOS PARA EL RECONOCIMIENTO DE LA SUPERFICIE PROTEICA DE VEGF

El factor de crecimiento endotelial vascular (VEGF) es una proteína señalizadora implicada en la vasculogénesis (formación *de novo* del sistema circulatorio embrionario) así como en la angiogénesis (crecimiento de vasos sanguíneos provenientes de vasos preexistentes). La angiogénesis es una de las señas de identidad de varias enfermedades relacionadas con cáncer, Parkinson, alzhéimer, etc. El desarrollo de fármacos que unen a VEGF es una difícil tara debido a la complejidad de la superficie proteica de esta proteína.

Por tal de demostrar el potencial de los péptidos cíclicos para reconocer superficies proteicas desafiantes, en nuestro laboratorio se diseñó una librería, llamada EXORIS (**Figura iii**) compuesta por hexapéptidos cíclicos simétricos del tipo $c(A_1\text{-}A_2\text{-}A_3)_2$, donde A_1, A_2 y A_3 son residuos pertenecientes a un set reducido de aminoácidos naturales, seleccionados cuidadosamente considerando sus propiedades fisicoquímicas, sintéticas y parámetros estadísticos como tendencia a adoptar una determinada estructura secundaria, hidrofobicidad y presencia en sitios activos.

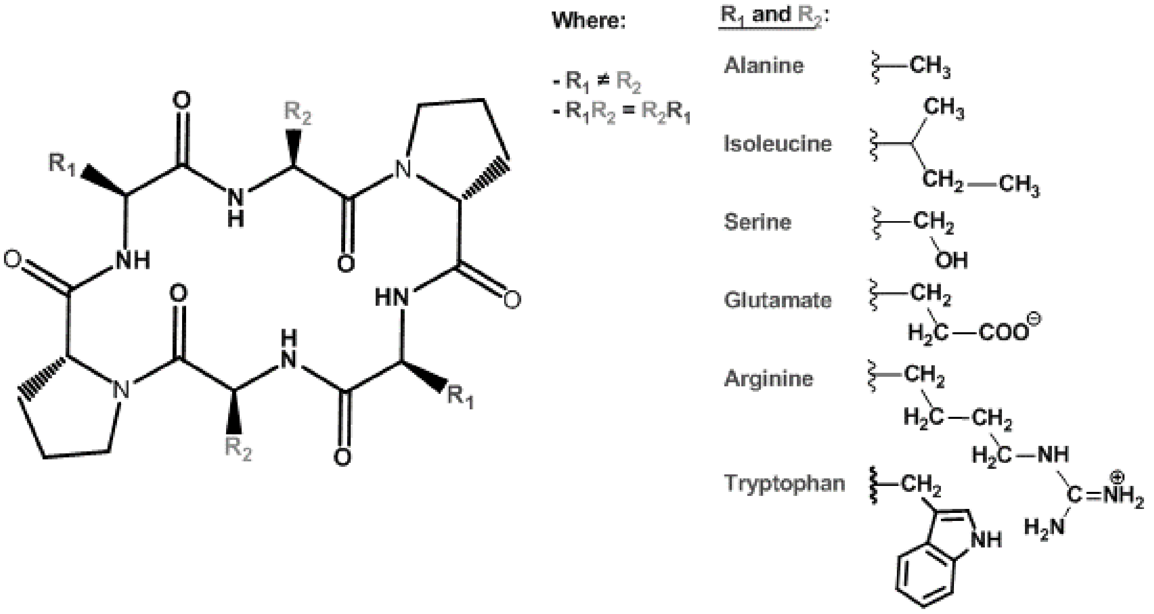


Figura iii. *Representación esquemática de la librería EXORIS*

El cribado inicial de esta librería frente a VEGF se llevó a cabo por espectroscopia de resonancia magnética nuclear (RMN). Para ello nos aprovechamos de la metodología previamente desarrollada por el Dr. Ricard Rodríguez, quien usó la VEGF marcada selectivamente con ^{13}C-metil-metionina y obtuvo un espectro de RMN de alta calidad donde se observaban solo cinco señales específicas correspondientes a las metioninas presentes en VEGF (posiciones 18, 55, 78, 81 y 94), lo cual simplificaba significativamente el uso de experimentos de RMN para el cribado de librerías. En nuestro caso, esta estrategia de marcado se benéfica de la afortunada localización de estas cinco metioninas en la estructura de VEGF. Dos de las cinco metioninas presentes en VEGF (Met 18 y Met 81) se sabe que tienen un papel importante en la formación y estabilización de complejos.

La evaluación de los péptidos EXORIS como posibles ligandos que unan a VEGF se realizó comparando los espectros de ^{1}H-^{13}C-HSQC de VEGF en ausencia y en presencia de 0.5 mM de péptido cíclico como se muestra en la **Figura iv**.

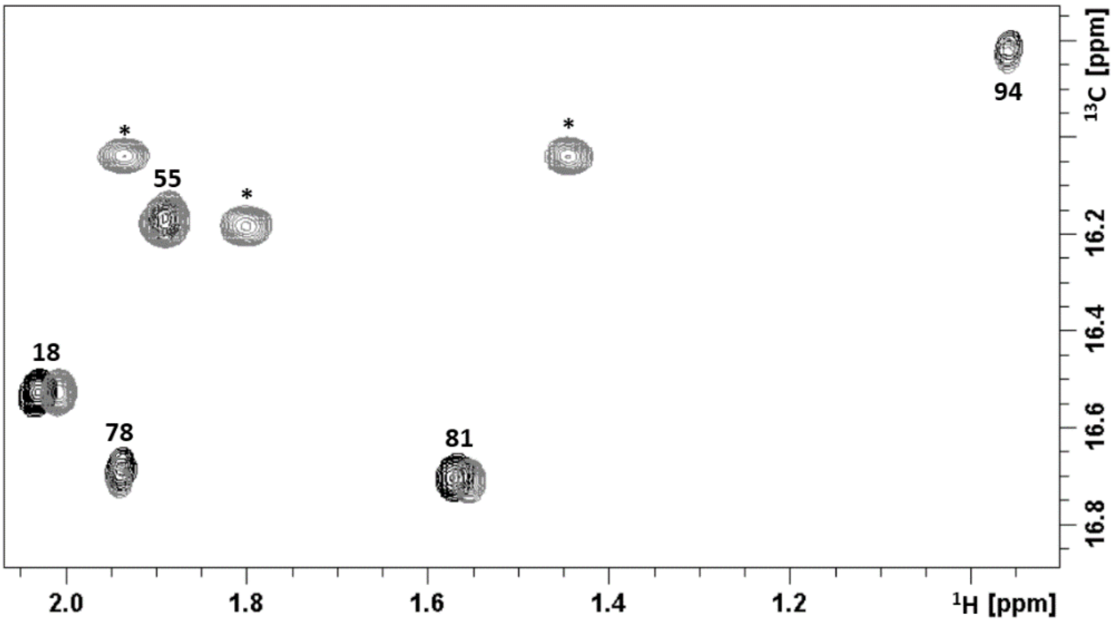


Figura iv. *Superposición de los espectros de ^{1}H-^{13}C-HSQC de VEGF en ausencia (negro) y en presencia de 0.5 mM de péptido cíclico (EpWEpW) (gris).*

La mayoría de los péptidos EXORIS causaron efectos prácticamente despreciables en el espectro de VEGF. Sólo aquellos péptidos que contienen triptófano produjeron cambios pequeños pero reproducibles en los desplazamientos químicos de las resonancias correspondientes a las Met18 y Met81 de VEGF. Este comportamiento sugiere que estos péptidos que contienen triptófano unen a VEGF, probablemente en la región de unión a los receptores, pero con afinidades bajas.

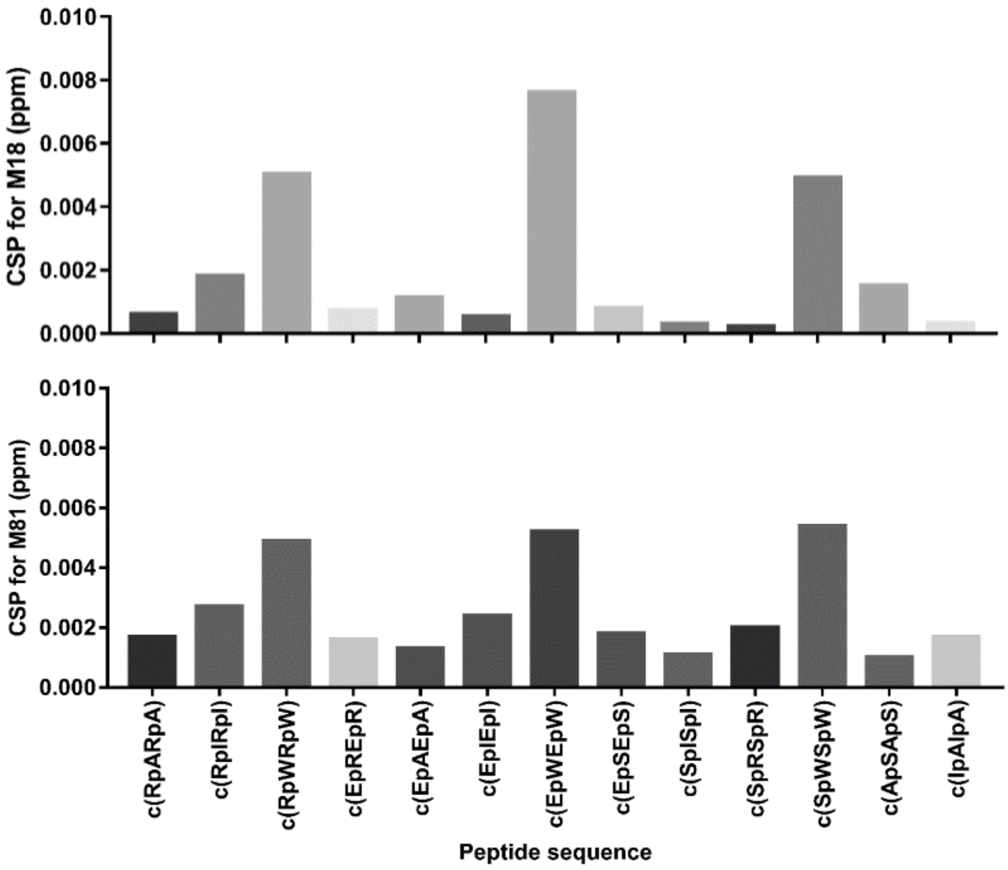


Figura v. *Representación gráfica de las cambios en los desplazamientos químicos de las resonancias de VEGF correspondientes a la Met 18 (arriba) y la Met81 (abajo) después de la adición de 0.5 mM de los péptidos EXORIS.*

Para confirmar la importancia del triptófano en la unión de los péptidos EXORIS a VEGF se seleccionó un péptido con triptófano c(EpWEpE) y uno sin c(EpSEpS) para realizar experimentos de diferencia de transferencia de saturación (STD). En presencia de proteína el péptido c(EpWEpW) presentó intensas señales STD, especialmente en la región aromática del espectro. Las señales STD observadas son indicativas de unión puesto que se producen por la transferencia de magnetización desde la superficie de la proteína saturada al ligando en su forma unida. En ausencia de VEGF no se observó la presencia de señales de STD, confirmando que el efecto STD era producido por la interacción del péptido con la proteína.

Cuando el mismo experimento de STD se llevó a cabo con el péptido c(EpSEpS) no se observaron señales STD en presencia de VEGF, confirmando que en este caso, la magnetización no se transfirió de la proteína al péptido, y por consecuencia confirmando la falta de unión de este péptido a VEGF.

En vistas de los resultados, se llevó a cabo un estudio más exhaustivo del péptido c(EpWEpW), puesto que éste dio lugar a los mayores cambios en las resonancias de las Met18 y Met81. Primero se estudió la zona de unión de este péptido en la superficie de VEGF. Para ello se compararon los experimento de ^{1}H-^{15}N-HSQC con VEGF marcada uniformemente con ^{15}N en ausencia y en presencia del péptido c(EpWEpW). Este tipo de experimentos permite detectar aquellos residuos de VEGF que son selectivamente perturbados por la unión del ligando. Mapeando los residuos más afectados por la presencia del ligando en la estructura tridimensional de VEGF se consigue información sobre la región de unión del péptido.

Los residuos más afectados por la adición del péptido c(EpWEpW) se observaron cerca de la hélice α situada en el extremo N-terminal y en la vecindad de las hebras β5 y β6, las cuales se sabe que son importantes en interacciones intermoleculares. Los aminoácidos cercanos a la hebra β2 y así como los que se encuentran cerca de β4 y β7 también se vieron afectados por la presencia del péptido cíclico (**Figura vi**).

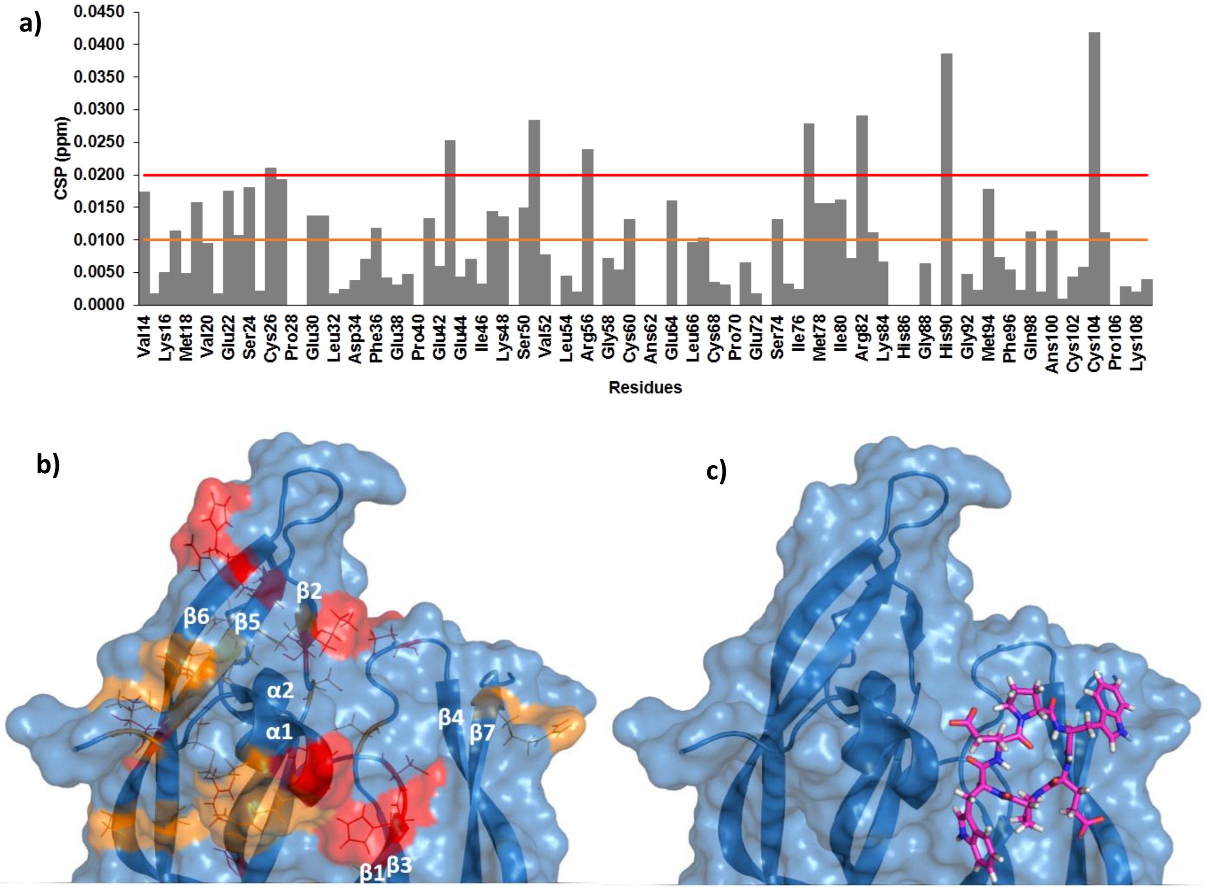


__Figura vi.__ a) Grafico de barras mostrando los resultados de los cambios en el desplazamiento químico de los residuos de VEGF tras la adición del péptido c(EpWEpW). b) Representación de la estructura en 3D de la zona de interacción de VEGF con sus receptores. Los residuos más afectados por la presencia del péptido cíclico se muestran resaltados en rojo y naranja. c) Resultados del modo de unión del péptido cíclico obtenidos por simulación virtual.

Basándonos en estos resultados se realizaron también estudios de simulación virtual (docking) para predecir el modo de unión de este péptido a VEGF, obteniéndose resultados que concordaban con los obtenidos mediante RMN.

Con el fin de determinar un valor de afinidad para el péptido c(EpWEpW), se realizó una titración por RMN añadiendo concentraciones crecientes de péptido frente a una concentración fija de VEGF. Los cambios en los desplazamientos químicos de las resonancias de la Met18 se representaron frente a la concentración de ligando para obtener una curva que se ajustó a la ecuación "One site-specific binding" mediante el software GraphPad Prism. A pesar de no conseguir un plató en la curva de saturación, ni siquiera con concentraciones elevadas de péptido (5 mM), se obtuvo una K_D estimada alrededor de 10 mM, confirmando la débil unión del péptido cíclico a VEGF.

Para completar la caracterización de los péptidos EXORIS se realizó también un estudio conformacional mediante RMN. Los parámetros de RMN susceptibles a la estructura, tales como desplazamientos químicos, coeficientes de temperatura de los protones de las amidas, las constantes de acoplamiento $^3J_{HNH\alpha}$ y las correlaciones en el espacio (NOE), mostraron la tendencia de estos hexapéptidos cíclicos a adoptar una estructura en forma de dos giros βII' consecutivos, situando las prolinas en la posición i+1 del giro. Esta conformación se vería estabilizada por la presencia de dos puentes de hidrogeno entre los CO(i) del esqueleto y los NH(i+3) (**Figura vii**).

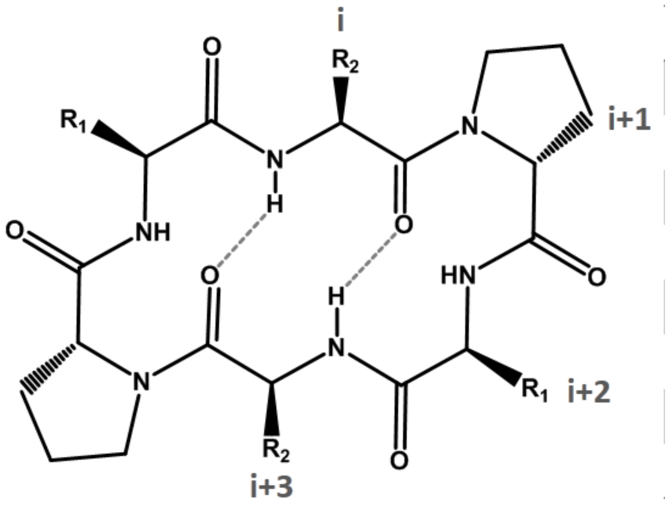


Figura vii. *Estructura propuesta para los péptidos EXORIS según los resultados de RMN. Las líneas discontinuas muestran los puentes de hidrogeno.*

Los resultados obtenidos para el péptido c(EpWEpW) nos alentaron diseñar serie de péptidos relacionados y a evaluar su capacidad de una para unir a VEGF. Estos nuevos péptidos se diseñaros explorando tres distintas modificaciones del péptido c(EpWEpW): 1) rompiendo la simetría C2, 2) mediante sustitución de aminoácidos y 3) mediante variaciones en la estereoquímica (**Figura viii**).

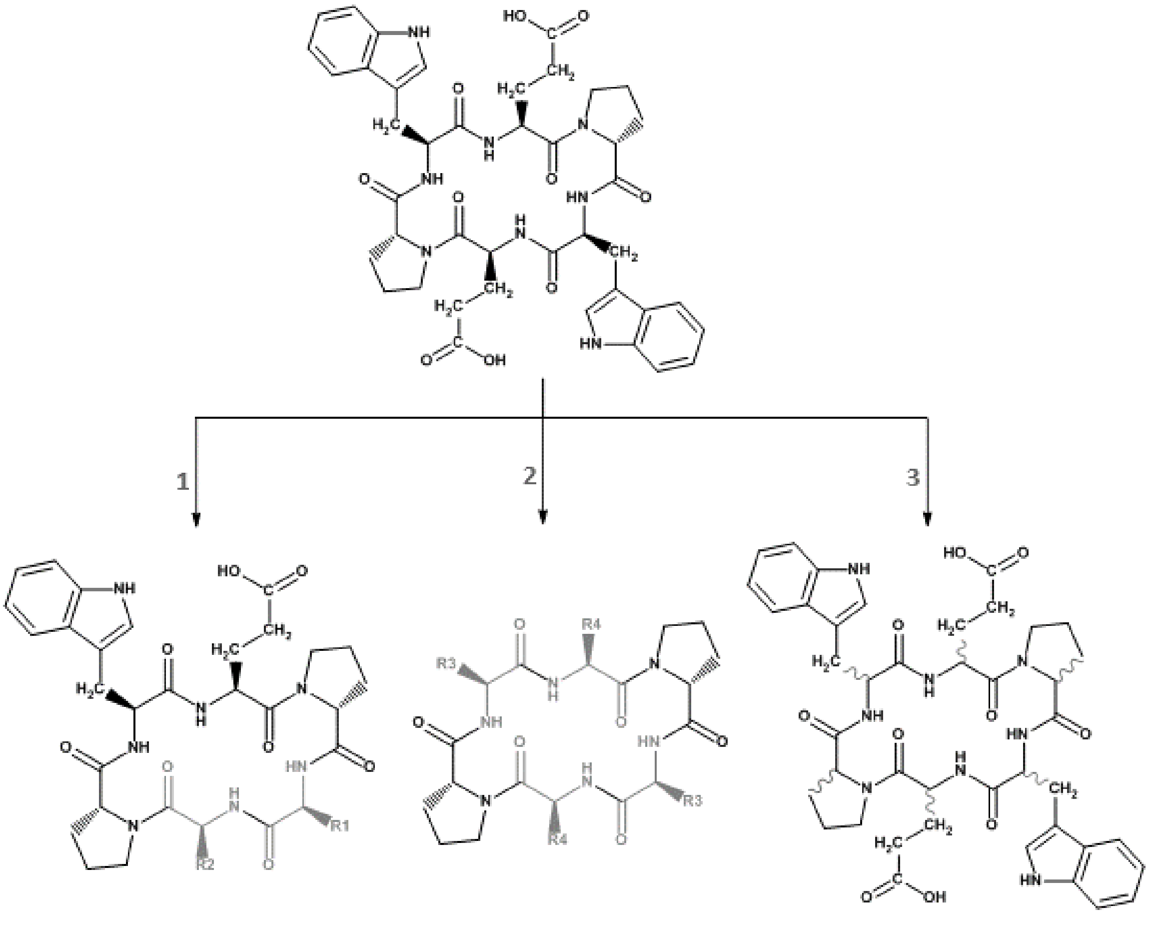


Figura viii. *Representación esquematica del proceso de diseño para la optimización del péptido c(EpWEpW). R1 y R2 son aminoácidos de la base reducida usada para la síntesis de los péptidos EXORIS. R3 son aminoácidos aromáticos naturales y no-naturales y R4 son aminoácidos naturales y no-naturales cargados negativamente.*

La evaluación de estos nuevos péptidos se realizó tanto por métodos computacionales como experimentalmente. Ninguno de los péptidos diseñados y evaluados presentó una afinidad de unión a VEGF significativamente menor a la obtenida con el péptido c(EpWEpW).

Los péptidos homoquirales c(EPWEPW) y c(epwepw) no se consiguieron ciclar en forma de hexapeptidos cíclicos. En lugar de eso se obtuvieron los correspondientes dímeros dando lugar a los dodecapeptidos homoquierales $c(EPWEPW)_2$ y $c(epwepw)_2$. Éstos se evaluaron frente a VEGF para evaluar su unión a esta proteína. Para el péptido homoquiral $c(epwepw)_2$ se obtuvieron resultados muy interesantes, pues se observó una interacción a VEGF del orden µM y se confirmó la unión de este péptido a la zona de unión de VEGF a sus receptores. Estos resultados abren la puerta al uso de este péptido u derivados para inhibir la unión de VEGF con sus receptores.

Dado que las variaciones en la estereoquímica de los hexapeptidos cíclicos derivados del péptido c(EpWEpE) dieron lugar a diferentes cambios en los desplazamientos químicos de las resonancias de VEGF, se decidió llevar a cabo también un estudio conformacional para comprobar cómo afectan a la conformación de los hexapéptidos cíclicos los cambios en la estereoquímica. Una selección de péptidos derivados del compuesto c(EpWEpW) fueron estudiados tanto por RMN como por métodos computacionales tales como dinámica molecular, observándose una gran variedad de conformaciones que adoptaban distintos tipos de giros β. Los resultados más interesantes se obtuvieron al modificar las posiciones de los residuos glutámico y triptófano en el péptido c(epWepW) para dar lugar al péptido c(wpEwpE), el cual presentó una conformación donde los enlaces amida de las dos prolinas se observaban en la conformación *cis* en lugar de la típica conformación *trans* observada normalmente en este tipo de péptidos. Además este péptido adoptaba una conformación en forma de dos giros βVI invertidos. Este tipo de péptidos con enlaces amida en *cis* pueden tener un especial interés biológico puesto que se han propuesto como un potencial motivo estructural de unión a diferentes PPIases, las cuales están adquiriendo un potencial interés biológico.

CAPÍTULO 2. ANÁLISIS CONFORMACIONAL DE ANÁLOGOS DE APAMINA

La apamina es una toxina que se encuentra en el veneno de las abejas y que actúa bloqueando ciertos canales de potasio mediados por calcio en el sistema nervioso central. Este péptido presenta una estructura bicíclica muy bien definida estabilizada mediante dos puentes disulfuro que unen una hélice α a un giro β, que apuntaba a una elevada estabilidad frente a la hidrolisis proteolítica. Estudios realizados en nuestro laboratorio por el Dr. Benjamí Oller-Salvia, en colaboración con la Dr. Macarena Sánchez, dieron lugar a la obtención de un péptido cíclico derivado de la apamina, con mayor capacidad para cruzar la barrera hematoencefálica que su precursor y que además presentaba una gran resistencia a la hidrolisis proteolítica (**Figura ix**).

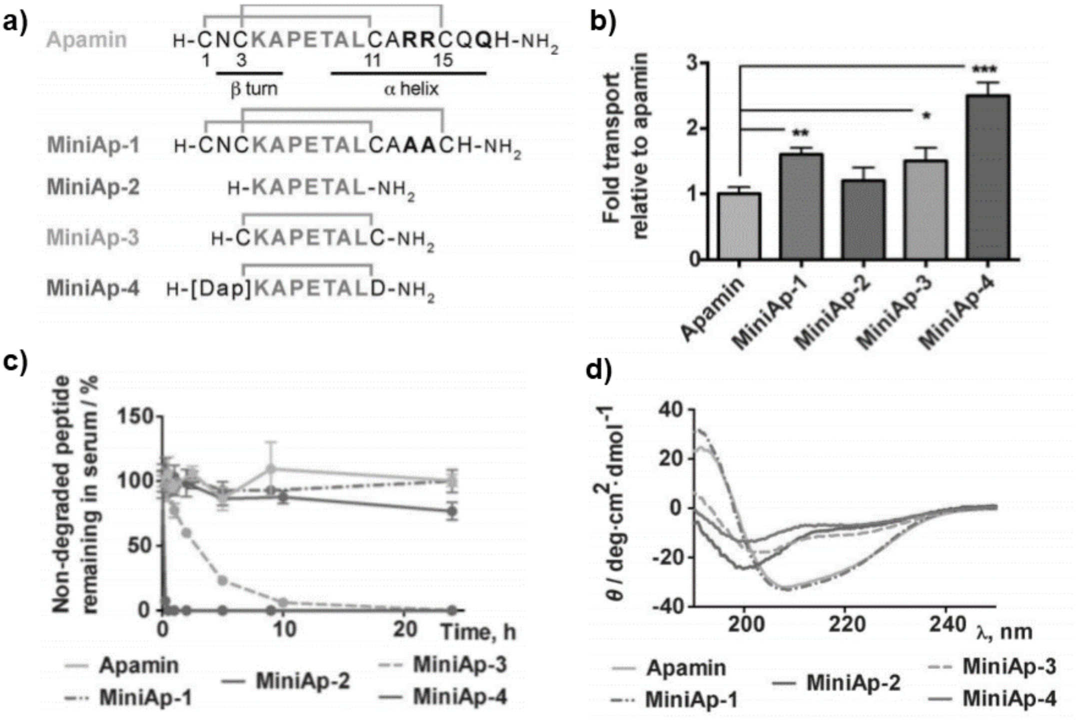


__Figura ix.__ Comparación de los diferentes analogos de la apamina. a) Secuencias de los péptidos. b) Transporte relativo de estos péptidos en un modelo basado en células bovinas. c) Estabilidad en serum humano. d) Espectros de dicroísmo circular.

Tras realizar un estudio preliminar por dicroísmo circular para comparar las distintas conformaciones entre la apamina y los análogos de ésta, sintetizados por el Dr. Oller-Salvia, se decidió ir más allá y realizar un análisis conformacional de estos péptidos por RMN.

Primero se realizó el estudio conformacional del péptido MiniAp-1. Los parámetros de RMN sensibles a la estructura tales como desplazamientos químicos secundarios de ^{1}H y ^{13}C, los coeficientes de temperatura de los NHs de amida y las constantes de acoplamiento $^3J_{HNH\alpha}$, sugirieron una conformación muy parecida a la descrita para la apamina original (**Figura x**).

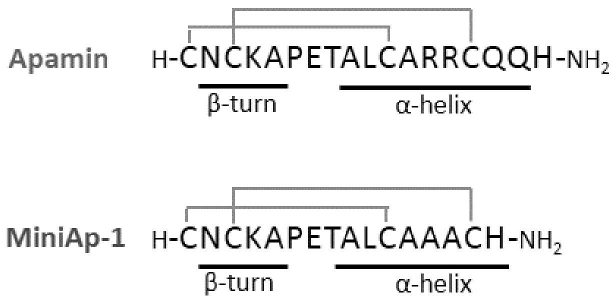
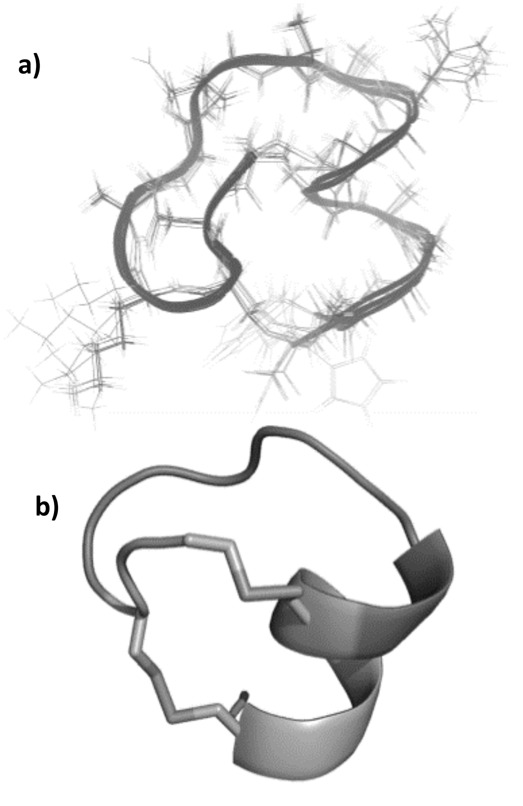


Figura x. *Representación esquemática comparando la apmina y la MiniAp-1.*

Basándose en los parámetros obtenidos por RMN el Dr. Jesús García obtuvo una representación gráfica en 3D donde se observó la gran similitud entre MiniAp-1 y la apamina original.



c)

Restraints	
NOE	
Sequential ($\|i-j\| = 1$)	21
Medium range ($1 < \|i-j\| \leq 4$)	25
Long range ($\|i-j\| > 4$)	6
Dihedral angle restraints	
ϕ	8
ψ	8
Quality	
Ramachandran Analysis	
Most favoured regions	12 (92.3%)
Additional allowed regions	1 (7.7%)
Generously allowed regions	0 (0%)
Disallowed regions	0 (0%)

Figura xi. *Estructura de MiniAp-1. a) Conjunto de las 10 mejores estructuras de MiniAp-1. Valores de RMSD de 0.24 y 1.31 Å se calcularon para la superposición del esqueleto y de las cadenas pesadas respectivamente. b) Representación de MiniAp-1. c) Estadística estructural de MiniAp-1.*

En cuanto a los otros tres análogos de la apamina, se observaron diferencias significativas para los dos péptidos monocíclicos (MiniAp-3 y MiniAp-4), comparados don el análogo lineal (MiniAp-2); mostrando este último una clara conformación totalmente flexible, tal y como se anticipó en el espectro de dicroísmo circular.

Aunque la información estructural obtenida para los péptidos MiniAp-3 y MiniAp-4 por RMN sugería que estos péptidos no presentaban una estructura secundaria definida, la mayor dispersión de los desplazamientos químicos correspondientes a los protones amida de estos dos péptidos y las notables diferencias en los desplazamientos de ^{13}C indicaron que éstos presentaban una menor flexibilidad conformacional que la observada para MiniAp-2 (**Figura xii**).

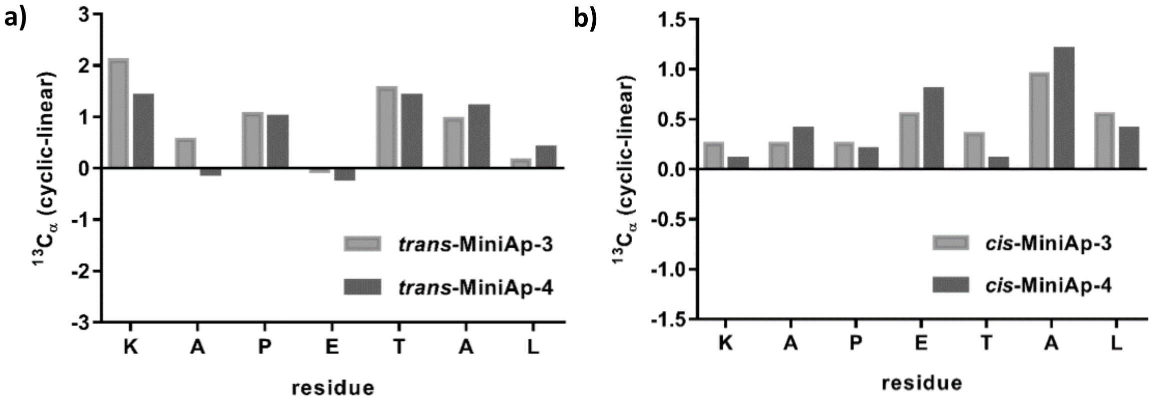


Figura xii. *Comparación de los desplazamientos químicos de 13C del péptido MiniAp-2 y las conformaciones trans (a) y cis (b) de los análogos cíclicos MiniAp-3 y MiniAp4.*

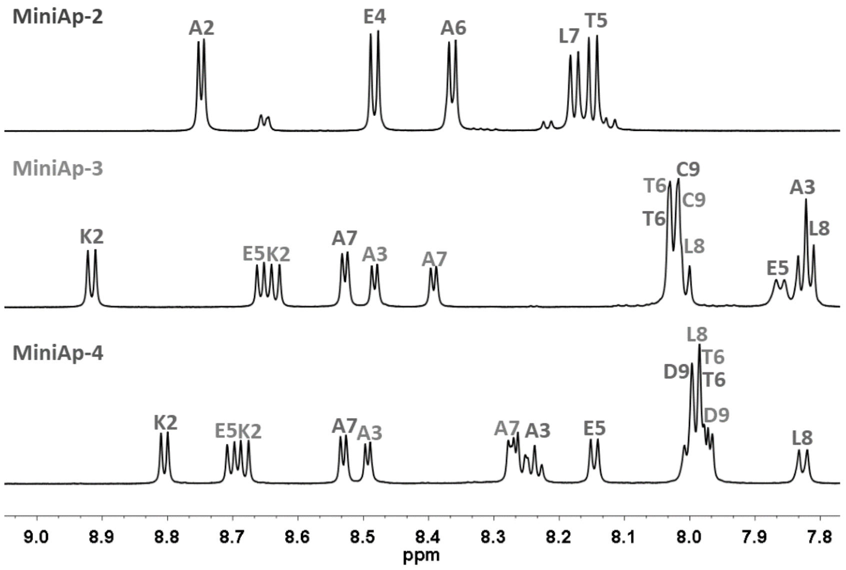


Figura xiii. *Comparación de la región del espectro de protón correspondiente a las amidas para MiniAp-2, MiniAp-3 y MiniAp-4.*

Más remarcable es el hecho de que el análogo lineal (MiniAp-2) presentó un ratio entre las poblaciones *cis/trans* del enlace amida de la prolina de 9:1, del mismo modo que se observó para el péptido bicíclico MiniAp-1; en cambio, los dos derivados monocíclicos presentaron un ratio de 1:1 tal y como se muestra en la **Figura xii**.

Las especies *cis* y *trans* de los tres derivados se asignaron basándose en la observación de intensas señales de NOE entre los protones α de Ala2 (MiniAp-2) o Ala3 (MiniAp-3 y -4) y los protones δ de las correspondientes prolinas, en el caso de la conformación *trans;* y entre los protones α de la Ala2 (MiniAp-2) o Ala3 (MiniAp-3 y -4) y los protones α de las correspondientes prolinas para la conformación *cis.* Estos resultados fueron confirmados mediante la diferencia entre los carbonos β y γ de la prolina: *trans*-Pro C_β - C_γ ($\Delta\delta = 4.7$ ppm para MiniAp-2 y 4.5 ppm para MiniAp-3 y MiniAp-4), y *cis*-Pro C_β - C_γ ($\Delta\delta = 9.6$ ppm para MiniAp-2 y 9.2 para MiniAp-3 y MiniAp-4).

Con estos resultados, nuestra hipótesis se basó en el hecho que la mayor preferencia por el rotámetro *cis* Ala-Pro en MiniAp-4 respecto a MiniAp-1 podría ser la explicación de las diferencias en permeabilidad entre los dos análogos con mayor resistencia a proteasas.

CAPÍTULO 3. DESARROLLO DE BIOSENSORES BASADOS EN FLUORESCENCIA PARA LA DETECCIÓN DE PROTEÍNAS

Además de aplicaciones terapéuticas, la habilidad de péptidos y proteínas para reconocer receptores u otras proteínas de forma específica, se puede explotar en varias aplicaciones biotecnológicas y farmacéuticas incluyendo el uso de biosensores o biomarcadores basados en estructuras peptídicas o proteicas. Las estrategias de bioconjugación ofrecen una forma eficiente de transformar información basada en interacciones entre polipéptidos y analitos en una señal que pueda ser medida, la cual puede usarse para la fabricación de nuevos biosensores polipeptídicos. Muchos fluoróforos sensibles pueden responder rápidamente a cambios ambientales, por eso, los fluoróforos sensibles al ambiente se han usado ampliamente como marcadores para conjugarlos a péptidos y proteínas con el fin de crear sensores moleculares polipeptídicos.

Actualmente los biosensores basados en fluorescencia se diseñan siguiendo tres pasos. Primero se consigue el ligando de unión, el cual se diseña a partir de distintos “scaffolds” frente a la diana terapéutica deseada, mediante el uso de tecnologías de evolución directa de librerías. El segundo paso consiste en escoger cuidadosamente el lugar donde introducir la sonda fluorescente, teniendo en cuenta que el residuo escogido debe: estar expuesto al disolvente, no perturbar la interacción entre el scaffold y el analito y estar situado cerca del lugar de unión. Finalmente el fluoróforo se conjuga al lugar específico del ligando de unión mediante técnicas químicas o enzimáticas establecidas.

La proteína ultra-estable Sso7d ha sido usada en varias ocasiones para generar proteínas muy estables capaces de unir a un amplio rango de dianas terapéuticas. Esta proteína tiene un tamaño relativamente pequeño de unos 63 con una estructura en forma de barril β incompleta con cinco láminas β y una hélice α en el extremo C-terminal. La Sso7d presenta una elevada estabilidad con una temperatura de fusión de 100^{o}C y no contiene cisteínas.

En este trabajo, realizado durante mi estancia en el laboratorio de la Prof. Barbara Imperiali en el MIT, primeramente se realizó un trabajo de optimización del protocolo de expresión, purificación y marcaje de un “scaffold” basado en Sso7d previamente diseñado por el Dr. Seymour De Picciotto, mediante “yeast surface display” (YSD). Este mutante de Sso7d, desde ahora M11.1.3, se diseñó para unir específicamente a la proteína MSA, para ser usado como prueba de concepto para demostrar que los “scaffolds” rígidos como el de la proteína Sso7d pueden minimizar la pre-activación del fluorofóro en ausencia de la diana terapéutica (**Figura xiv**).

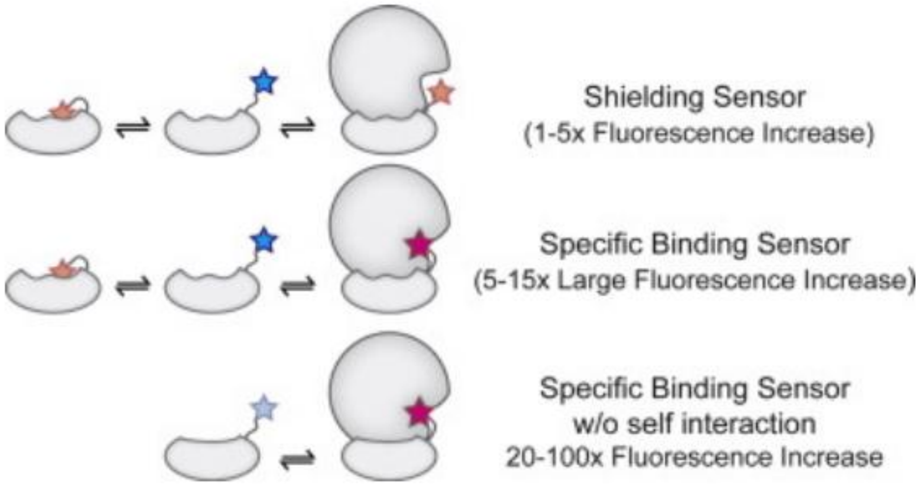


Figura xiv. *Representación de los tres mecanismos de activación de los biosensors. En el primer y el segundo ejemplo se muestra el proceso de preactivación del fluoróforo.*

Inicialmente se realizaron pruebas de expresión y purificación para los mutantes M11.1.3(I23C), M11.1.3(W25C) y M11.1.3(G26C). Se comprobó que la expresión de estos mutantes iba siempre acompañada de una chaperona. Después de varias pruebas de purificación concluimos que la purificación mediante intercambio catiónico era la técnica más adecuada para la eliminación de esta chaperona. Tras obtener los mutantes deseados puros, se realizaron pruebas de marcaje con distintos fluoróforos derivados de 4-DMN (**Figura xv**).

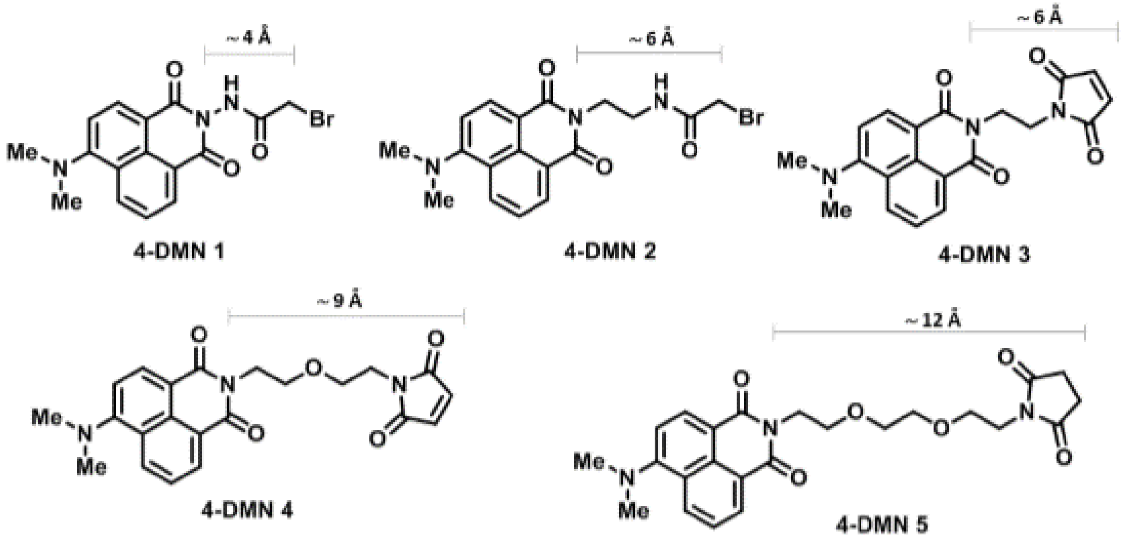


Figura xv. *Agentes de union a cisteína basados en 4-DMN para el marcaje de los mutantes de Sso7d.*

Después de realizar varias pruebas se observó que la mejor forma de purificación para eliminar el exceso de fluorofóro después del marcaje de las proteínas era, nuevamente, mediante cromatografía de intercambio catiónico.

Con esto se consiguió desarrollar un protocolo completo para la expresión, purificación y marcaje de proteínas basadas en Sso7d (**Figura xvi**).

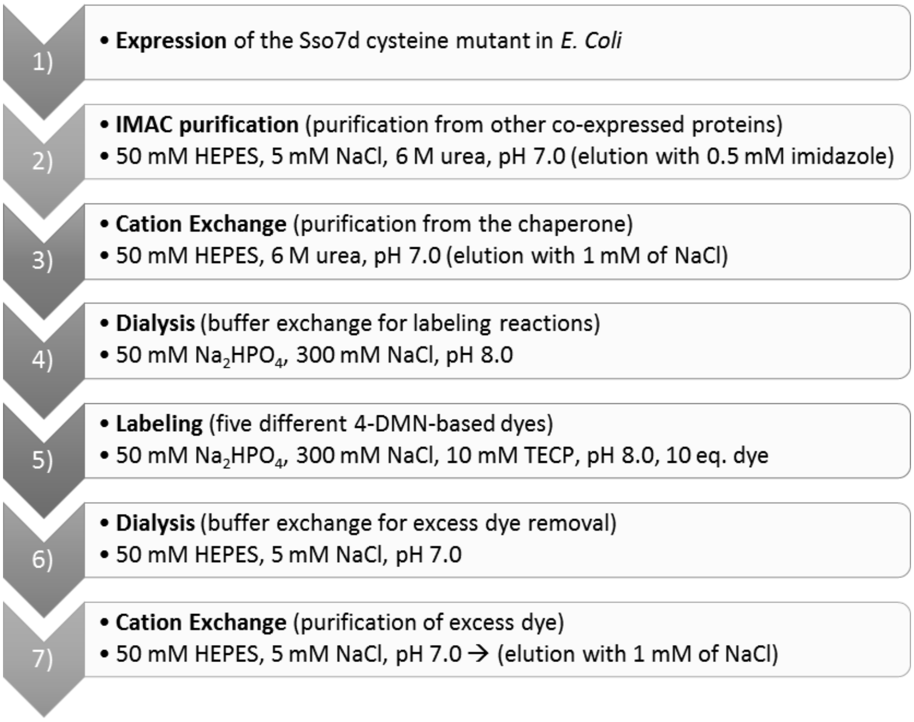


Figura xvi. *Protocolo optimizado para la expresión, purificación y marcaje de cisteínas en proteínas basadas en Sso7d.*

Una vez hecho esto, se decidió probar el diseño de ligandos basados en Sso7d capaces de unir a una diana relevante como el factor de crecimiento epidérmico (hEGF), la cual está presente en numerosos tipos de cáncer. Para el diseño de estos ligandos se usó la metodología basada en YSD, siguiendo el procedimiento mostrado en la **Figura xvii**. Para ello se usaron dos librerías basadas en Sso7d en las cuales nuevo posiciones de la secuencia de la proteína se mutan al azar usando 11 (Librería 11) o 18 (Librería 18) de los 20 aminoácidos naturales (**Figura xviii**).

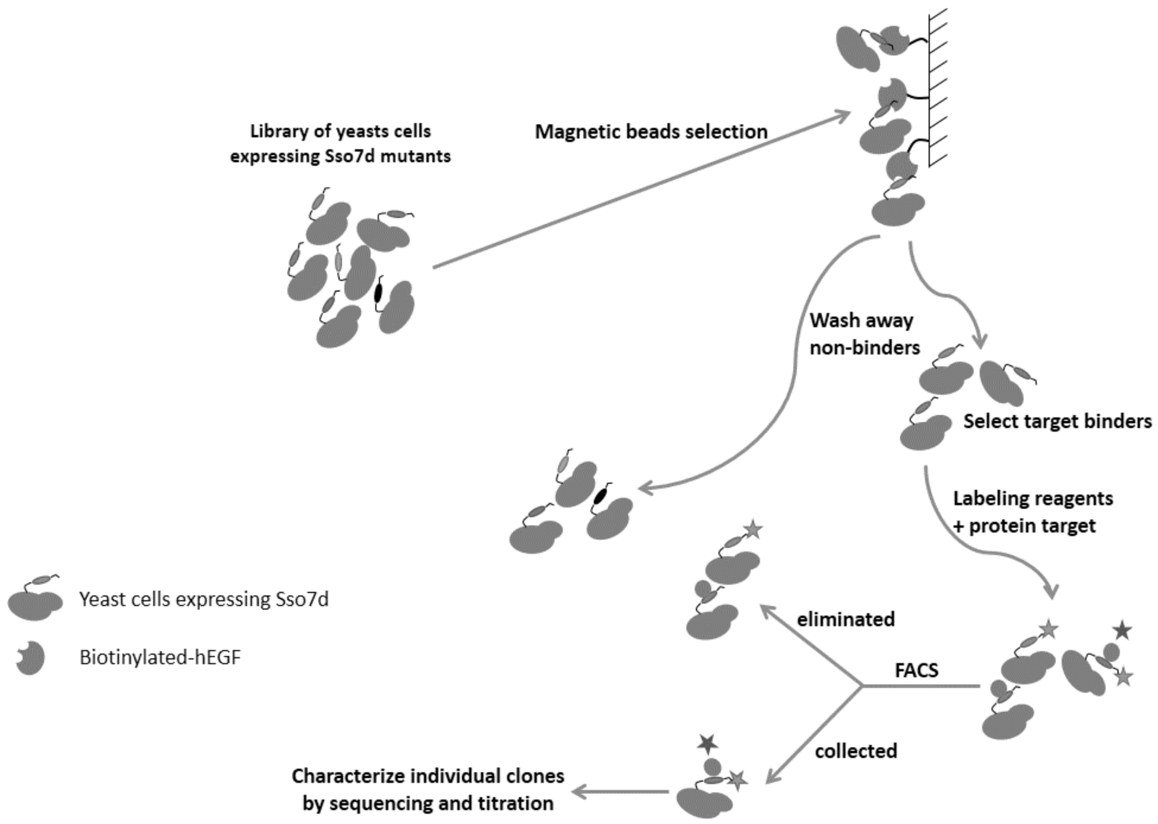
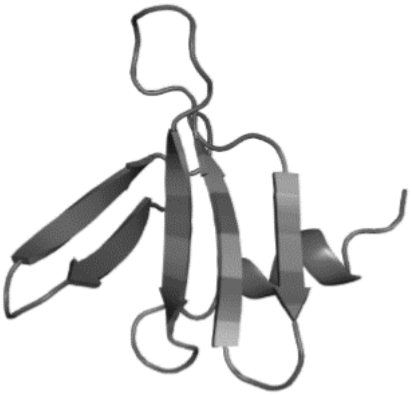


Figura xvii. *Representación esquemática del proceso seguido para la obtención de ligandos capaces de unir a hEGF.*



Sequence#	Details	Sequence
1-19	Invariant	ATVKFKYKGEEKQVDISKI
20-26	Strand 1	KKVWRVG
27-33	Strand 2	KMISFTY
34-40	Invariant	DEGGGKT
41-47	Strand 3	GRGAVSE
48-61	Invariant	KDAPKELLQMLEKQ

Figura xviii. *Representación de la estructura donde se observan las nueve posiciones de la secuencia que se mutan aleatoriamente durante el proceso de desarrollo mediante YSD.*

La técnica de YSD se basa en la expresión de las proteínas de interés como proteínas de fusión a Aga-2 en la superficie de células de levaduras (**Figura xix**).

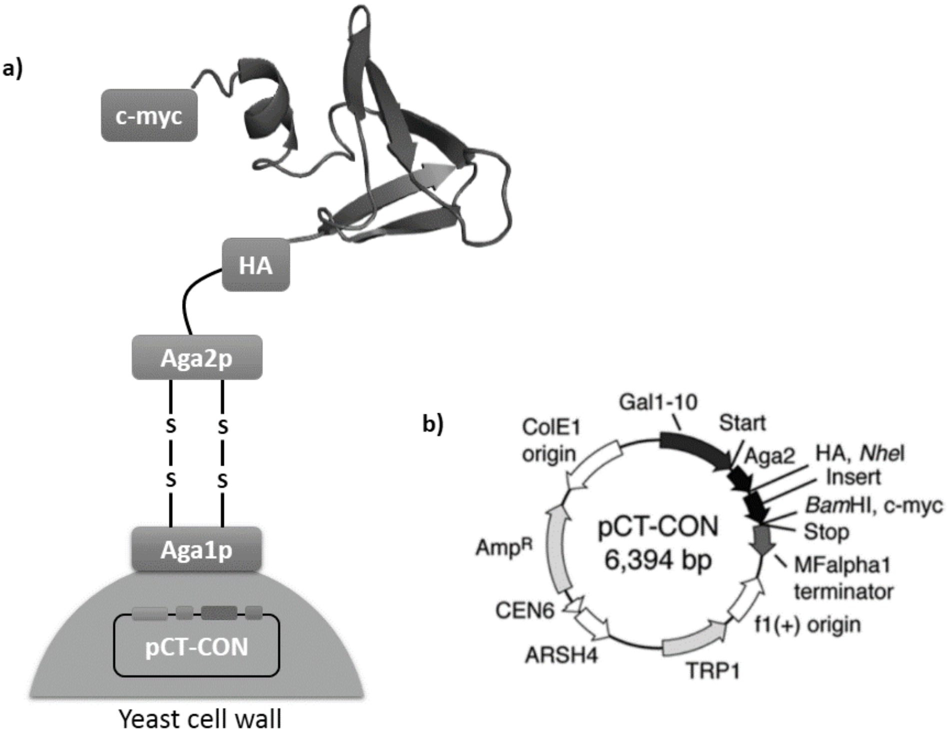
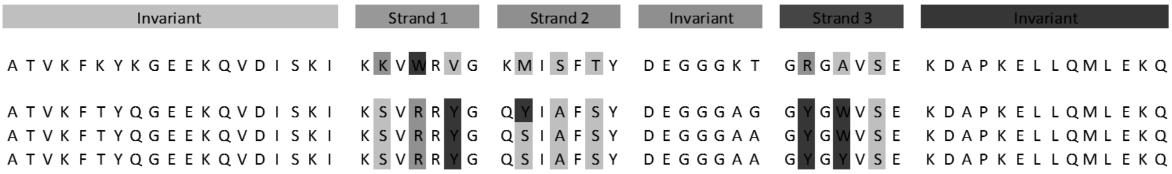


Figura xix. *Representación esquemática de YSD. a) Representación del scaffold de Sso7d expresado en la superficie de la celula. b) Mapa del vector usado en esta técnica. La etiqueta INSERT indica la posición donde se inserta el gen de Sso7d.*

Después de dos rondas de selección mediante perlas magnéticas y dos rondas de selección mediante citometría de flujo, se consiguieron aislar los primeros ligandos basados en Sso7d capaces de unir a hEGF. El DNA de estos ligandos se aisló para ser secuenciado e identificar así los primeros ligandos de hEGF que presentaron una unión del orden de 3 µM frente a hEGF (**Figura xx**).



Invariant	Strand 1	Strand 2	Invariant	Strand 3	Invariant
ATVKFKYKGEEKQVDISKI	KKVWRVG	KMISFTY	DEGGGKT	GRGAVSE	KDAPKELLQMLEKQ
ATVKFTYQGEEKQVDISKI	KSVRRYG	QYIAFSY	DEGGGAG	GYGWVSE	KDAPKELLQMLEKQ
ATVKFTYQGEEKQVDISKI	KSVRRYG	QSIAFSY	DEGGGAA	GYGWVSE	KDAPKELLQMLEKQ
ATVKFTYQGEEKQVDISKI	KSVRRYG	QSIAFSY	DEGGGAA	GYGYVSE	KDAPKELLQMLEKQ

***Figura xx**. Secuencias de los primeros ligandos de hEGF obtenidos mediante YSD*

CONCLUSIONES

1. Los hexapéptidos cíclicos que contienen triptófano en su secuencia unen a VEGF en la zona de unión a receptores, con afinidades bajas. Por eso estos péptidos pueden usarse como punto de inicio en el diseño de ligandos de unión a VEGF con afinidades más altas y mayor especificidad. La resonancia magnética nuclear demostró ser una técnica adecuada para el cribado de ligandos de unión a VEGF con afinidades bajas.
2. Los hexapéptidos cíclicos con simetría C2 adoptan una conformación que contien dos giros βII' consecutivos, estabilizados mediante dos puentes de hidrogeno entre los aminoácidos que preceden a la prolina.
3. La estereoquímica de los residuos en los hexapeptidos cíclicos afecta ligeramente la unión de éstos a VEGF. La simetría C2 no es necesaria para la unión a VEGF. Residuos más hidrofóbicos mejoran ligeramente la unión de estos péptidos a VEGF. Finalmente la expansión del tamaño del anillo de estos péptidos cíclicos demostró ser clave en la obtención de un dodecapéptido cíclico capaz de unir a VEGF con afinidades del orden µM y con elevada especificidad.
4. La estereoquímica de los hexapeptidos cíclicos afecta significativamente a su conformación. Desafiantes tipos de giros β con los enlaces amida de las prolinas en *cis* se pueden obtener con hexapéptidos cíclicos. Estos péptidos pueden tener un gran interés biológico para tratar nuevas dianas como las ciclofilinas.
5. El péptido MiniAp-1 adopta una conformación similar a la descrita para la apamina original, definida por un giro β y una hélice α unidas mediante puentes disulfuro.
6. El péptido MiniAp-2 presentó una conformación totalmente flexible, mientras que los péptidos cíclicos MiniAp-3 y MiniAp-4 presentaron una estructura más rígida aunque no se pudo demostrar la presencia de ninguna estructura secundaria definida.
7. Se estableció un protocolo optimizado para la expresión, purificación y marcaje de los mutantes basados en Sso7d. La cromatografía de intercambio catiónico demostró ser la mejor técnica para la eliminación de la chaperona que eluía juntamente a las proteínas deseadas.
8. El mutante M11.1.3(W25C)-4-DMN2 resulto ser el mejor candidato a biosensor entre todos los mutantes basados en M11.1.3 debido a que presentó la intensidad de

fluorescencia más baja en ausencia de MSA y un mayor incremento en la intensidad de fluorescencia tras unir a MSA.

9. La metodología basada en YSD se usó exitosamente para la obtención de ligandos de hEGF basados en la Sso7d. Los primeros ligandos de hEGF fueron caracterizados observando una afinidad de alrededor de 3 µM frente a hEGF. Aun se deben hacer algunos esfuerzos para convertir estos ligandos en biosensores capaces de detecta hEGF *in vivo.*