

**THE POTENTIAL OF PEPTIDE CHEMISTRY:
FROM SMALL MOLECULES UP TO
INNOVATIVE PROTEIN ENGINEERING**

Ph.D. Thesis

DEBORAH SEMENTA

2017

TUTOR:
Prof. ETTORE NOVELLINO

CO-TUTOR:
Assoc. Prof. BRADLEY L. PENTELUTE



The potential of peptide chemistry: from small molecules up to innovative protein engineering
Ph.D. Thesis, Department of Pharmacy, University of Naples Federico II (Italy)

Copyright: Deborah Sementa, 2017

All rights reserved. No part of this publication may be reproduced, stored or transmitted in any forms or by any means, without written permission of the author.

It is written based on three-years research conducted by the author, both at the Department of Pharmacy in Naples (IT) and at the Department of Chemistry at Massachusetts Institute of Technology in Cambridge (MA).

The research described in this thesis was supported by European Social Fund, P.O. Campania.

UNIVERSITY OF NAPLES FEDERICO II

DEPARTMENT OF PHARMACY



Ph.D. in Pharmaceutical Science – XXIX cycle

**The Potential of Peptide Chemistry:
from Small Molecules up to
Innovative Protein Engineering**

Ph.D. Thesis

DEBORAH SEMENTA

Tutor:

Prof. ETTORE NOVELLINO

Ph.D. Coordinator:

Prof. MARIA VALERIA D'AURIA

Co-Tutor:

Assoc. Prof. Bradley L. Pentelute

April 2017

DEFENSE COMMITTEE

Commissioners

Prof. GIOVANNI APPENDINO,
University of Piemonte Orientale

Prof. ROMANO SILVESTRI,
University of Rome La Sapienza

Prof. ROBERTA D'EMMANUELE DI VILLA BIANCA,
University of Naples Federico II

Prof. FLEMMING HOFMANN LARSEN,
University of Copenhagen

List of abbreviations

1-Nal	L-3-(1-Naphthyl)-alanine
2-Nal	L-3-(2-Naphthyl)-alanine
2Cl-Trt	2-chlorotrityl
6-Aha	Aminohexanoic acid
ACE	Acetyl
ADP	Adenosine diphosphate
AIDS	Acquired Immune Deficiency Syndrome
AMP	Adenosine monophosphate
AMBS	4-(aminomethyl)benzoic acid
AT	Anthrax toxin
ATP	Adenosine triphosphate
BOC	Di- <i>tert</i> -butyl dicarbonate
CCL	CC chemokine ligand
CCR	CC chemokine receptor
CHO	Chinese hamster ovary
CV	Column volume
CXCL	CXC chemokine ligand
CXCR	CXC chemokine receptor
ddH ₂ O	Double distilled water
DCM	Dichloromethane
DIEA	<i>N,N</i> -diisopropylethylamine
DMF	<i>N,N</i> -dimethylformamide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DOTA	1,4,7,10-Tetraazacyclododecane-1,4,7,10-tetraacetic acid
dsDNA	double stranded DNA
DTA	Diphtheria toxin A

DTPA	2-[Bis[2-[bis(carboxymethyl)amino]ethyl]amino]acetic acid
E. coli	Escherichia coli
ECL	Extracellular loop
EDT	1,2-ethanedithiol
EDTA	Ethylenediaminetetraacetic acid
EF	Edema Factor
Fmoc	9-fluorenylmethyloxycarbonyl
GPCR	G protein-coupled receptor
HATU	<i>O</i> -(7-azabenzotriazolo-1-yl)- <i>N,N,N',N'</i> tetramethyl-uroniumhexafluorophosphate
HBTU	<i>O</i> -benzotriazole- <i>N,N,N',N'</i> tetramethyl-uroniumhexafluorophosphate
HIV	Human immunodeficiency virus
HOBt	<i>N</i> -hydroxybenzotriazole
HRP	Horse radish peroxidase
HSC	Hematopoietic stem cell
IC	Inhibitory concentration
ICL	Intracellular loop
IHC	Immunohistochemistry
IPTG	Isopropyl- β -D-thiogalactopyranosid
kDa	Kilo Dalton
GC-MS	Gas chromatography-mass spectrometry
LB	Luria Bertani
LF	Lethal Factor
M	Molar
<i>m/z</i>	mass-charge relation
MAPK	mitogen-activated protein kinase
MD	Molecular Dynamics
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
min	minute
MS	Mass spectrometry
MW	Molecular weight
nat	Natural
NF κ B	Nuclear Factor “Kappa-light-chain-enhancer” of activated B cells

Ni-NTA	Nickel nitrilotriacetic acid
NME	<i>N</i> -methyl
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NOTA	1,4,7-triazacyclononane- <i>N,N',N''</i> -triacetic acid
OD	Optical density
PA	Protective antigen
Pen	Penicillamine
PET	Positron emission tomography
PI3K	Phosphatidylinositol 3-kinase
PKA	Protein kinase A
POPC	1- palmitoyl-2-oleoylphosphatidylcholine
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PyAOP	7-Azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
RLL	Rabbit Reticulocyte Lysate
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
RT	Room temperature
SAR	Structure–activity relationship
SD	Standard deviation
SDF-1 α	Stromal cell-derived factor-1 α
SDS	Sodium dodecylsulfate
SDS-PAGE	Sodiumdodecylsulfatepolyacrylamidegelelectrophoresis
SPPS	Solid phase peptide synthesis
SRM	Selected-reaction monitoring
SrtA	Sortase A
SUMO	Small Ubiquitin-related Modifier
TBS	Tris buffered saline
TBST	TBS buffer containing Tween 20
<i>t</i> -Bu	Tertiary butyl

TFA	Trifluoroacetic acid
TIS	Triisopropylsilane
TM	Transmembrane
TLC	Thin layer chromatography
TRIS	Tris(hydroxymethyl)aminomethane
TOCSY	Total correlation spectroscopy
VEGF	Vascular endothelial growth factor
vCCL	viral chemokine/cytokine ligand (or vMIP)
vMIP	viral macrophage inflammatory protein (or vCCL)
WT	Wild type

Table of Contents

LIST OF ABBREVIATIONS.....	V
TABLE OF CONTENTS	1
PART I.....	7
ABSTRACT.....	8
CHAPTER 1 INTRODUCTION.....	11
1.1 THE CHEMOKINE NETWORK.....	11
1.2 CXCR4-CXCL12 SIGNALING.....	14
1.3 TWO SIDE MODEL	15
1.4 ESSENTIAL PHYSIOLOGICAL ROLES OF CXCL12/CXCR4 AXIS	17
1.5 PATHOLOGICAL ROLES OF CXCL12/CXCR4 SIGNALING.....	19
1.6 REFERENCES	22
CHAPTER 2 EXPLORING THE N-TERMINAL REGION OF C-X-C MOTIF CHEMOKINE 12 (CXCL12): IDENTIFICATION OF PLASMA-STABLE CYCLIC PEPTIDES AS NOVEL, POTENT C-X-C CHEMOKINE RECEPTOR TYPE 4 (CXCR4) ANTAGONISTS	27
2.1 INTRODUCTION.....	27
2.2 RESULTS AND DISCUSSION.....	29
2.2.1 Design.....	29
2.2.2 Chemistry	29
2.2.3 CXCR4 Binding Assay	31
2.2.3 CXCR3 and CXCR7 Binding Assays	31
2.2.4 Cell Migration and p-ERK Induction Experiments.....	34
2.2.5 Evaluation of 2 and 10 on Human Colon Cancer Cells.....	34
2.2.6 CXCR4 Internalization Assay	34
2.2.7 Plasma Stability Assay	36
2.2.8 NMR Spectroscopy	37
2.2.9 Molecular Modeling.....	38

2.3 CONCLUSION	42
2.4 REFERENCES	44
2.5 EXPERIMENTAL SECTION	49
2.5.1 Chemistry	49
2.5.2 Biological Studies.....	53
CHAPTER 3 STRUCTURE-ACTIVITY RELATIONSHIP STUDY LED TO THE IDENTIFICATION OF A NEW POTENT AND PLASMA STABLE CXCR4 ANTAGONIST MIMICKING THE N-TERMINAL REGION OF CXCL12	57
3.1 INTRODUCTION.....	57
3.2 RESULTS AND DISCUSSION	60
3.2.1 Design.....	60
3.2.2 Chemistry	62
3.2.3 CXCR4 Binding Assay	62
3.2.4 CXCR3 and CXCR7 Binding Assays	63
3.2.5 Plasma Stability Assay	65
3.3 CONCLUSION	65
3.4 REFERENCES	67
3.5 EXPERIMENTAL SECTION	70
3.5.1 Chemistry	70
3.5.2 Biological Studies.....	73
CHAPTER 4 PRE-CLINICAL DEVELOPMENT OF A NOVEL CYCLIC PEPTIDE ⁶⁸ GA PET TRACER FOR COLON CANCER	75
4.1 INTRODUCTION.....	75
4.2 RESULTS AND DISCUSSION	84
4.2.1 Design.....	84
4.2.2 Chemistry	84
4.2.3 CXCR4 Binding Assay	86
4.3 CONCLUSION	87
4.4 REFERENCES	88
4.5 EXPERIMENTAL SECTION	92

4.5.1 Chemistry	92
4.5.2 Biological studies	95
PART II	99
ABSTRACT.....	100
CHAPTER 1 INTRODUCTION.....	103
1.1 SURVIVAL THROUGH DESTRUCTION	103
1.2 UBIQUITINATION MACHINERY	104
1.3 THE PROTEASOME: A COMPLEX OF TWO HALVES	107
1.3.1 The proteolytic core particle.....	107
1.3.2 The 19S Regulatory Particle.....	110
1.4 DEFINING THE GEOMETRY OF THE TWO-COMPONENT PROTEASOME DEGRON	117
1.5 N-DEGRONS AND THE N-END RULE PATHWAY	117
1.6 A SECOND COMPONENT OF THE PROTEASOME TARGETING CODE	118
1.7 DEFINING THE GEOMETRY OF THE TWO COMPONENT PROTEASOME DEGRON.....	120
1.8 REFERENCES	123
CHAPTER 2 RESCUE OF C-TERMINAL DESTABILIZED PROTEINS BY USE OF D-AMINO ACIDS	129
2.1 INTRODUCTION.....	129
2.1.1 Delivery of therapeutic proteins: challenges and strategies	129
2.1.2 Break on through to the other side: the bacterial toxin mediated delivery	131
2.1.3 AT- a paradigm for toxin-mediated protein translocation.....	132
2.1.4 Fusion and conjugation approaches.....	136
2.1.5 SrtA-catalyzed ligation.....	136
2.2 RESULTS AND DISCUSSION.....	138
2.2.1 Design.....	138
2.2.2 Chemistry	138
2.2.2.1 Peptide synthesis.....	138
2.2.2.2 Sortase A attaches peptide sequences onto the C-terminus of LF _N and LF _N -DTA.....	139

2.2.3 Biological Studies.....	139
2.2.3.1 Translocation of LF _N -DTA-X and LF _N -X Constructs.....	139
2.2.3.2 Optimization of western blot workflow.....	142
2.2.3.3 Western Blot Analysis with LF _N -DTA-X and LF _N -X Constructs	146
2.2.3.4 Protein Synthesis Inhibition Assay with LF _N -DTA-X Constructs	147
2.3 CONCLUSION.....	151
2.4 REFERENCES	152
2.5 EXPERIMENTAL SECTION.....	156
2.5.1 Chemistry	156
2.5.2 Biological Studies.....	160

PART I

--

UNIVERSITY OF NAPLES FEDERICO II

DEPARTMENT OF PHARMACY



Design, Synthesis and Functionalization of Cyclic Peptides Targeting Chemokine Receptor 4

ABSTRACT

OBJECTIVES

In the chemokine receptor subfamily, CXCR4 dictates directional cell migration during routine immune surveillance and activates leukocytes for the inflammatory response, towards gradients of the CXCL12 ligand. Although CXCL12/CXCR4 axis evolved to benefit the host, inappropriate regulation or expression of these proteins is involved in pathological conditions, including infection, chronic inflammatory diseases, atherosclerosis, and more than 20 different human cancers. Notably, overexpression of CXCR4 has been associated with increased metastatic potential and poor prognosis in many solid tumors. CXCR4/CXCL12 pairing plays a central role in Paget's hypothesis of "seed-and-soil", and thus targeting the pathway is not only a logic therapeutic approach but also a strategy in blocking cancer progression or malignant transformation. There is, therefore, an impetus for non-invasive means to sensitively identify tumours prone to progression. As such, the aim of this work was to develop a number of novel CXCR4 antagonists to assess their suitability as imaging radiotracers for positron emission tomography (PET).

METHODS AND RESULTS

As part of our ongoing efforts in identifying new CXCR4 antagonists, a CXCL12-derived small cyclic peptide, which selectively interacted with CXCR4, was identified. Although the sequence possessed an interesting *in vitro* and *in vivo* pharmacological profile, it suffered from degradation in biological fluids and its IC₅₀ was in the low micromolar range.

Herein, to enhance both the CXCR4 affinity and metabolic stability, I embarked on a lead optimization campaign. Efforts to elucidate the necessary residues for the interaction with the receptor and the critical stereocenters of mine sequences systematically included tools such as alanine scanning and D-aminoacid scanning. Next, the overall, dynamic and topographical requirements and electronic properties of the cyclic region of the obtained compounds were investigated. Inspired by the promising results, the newly generated potent, selective and functional antagonist has been tailored as a target-specific CXCR4 imaging probe. Exploring the flexibility of the receptor binding pocket, a study of the optimal site of attachment and an optimization of suitable linker units was carried on.

CONCLUSIONS

To meet the clinical need for a highly specific and sensitive tool for the chemokine receptor 4 assessment and quantification *in vivo*, a CXCR4 targeted nuclear probe has been developed. Despite the additional moieties, the final compound showed high affinity to the chemokine CXCR4 receptor and the collected results suggest that the ^{nat}Ga is a promising agent for preclinical non-invasive PET imaging.

Chapter 1

Introduction

1.1 The chemokine network

Chemokines are small globular signaling cytokines (8-12kDa) secreted by various stromal and epithelial cells.¹ In a process known as chemotaxis, the interaction with their respective receptors induces cytoskeletal rearrangements and concentration gradient-driven chemotactic migration of nearby responsive cells.²

Structures of many chemokines have been solved by NMR spectroscopy and X-ray crystallography.³ Based on these protein structure insights, virtually all chemokines, regardless of family and despite low degree of sequence homology, adopt the same monomeric architecture. They exhibit a common folding motif in which a flexible N-terminal region (N-loop) precedes the short single-turn 3_{10} helix of the highly structured core domain, then the polypeptide chain is forced into three antiparallel β -strands in a Greek key like arrangement and overlaid by an α -helix (Figure 1.1). Intramolecular disulfide bonds stabilize the overall topology. To date, approximately a chemokine superfamily of 50 human ligands and 20 different transmembrane receptors^{4,5} has been discovered and traditionally divided into four different categories (CXC, CC, C, and CX3C) on the basis of the pattern of conserved -disulfide-bonded cysteine residues in proximity to the amino terminus of the mature proteins (here C represents cysteine and X/X3 represents one or three random non cysteine amino acids; in particular, ligands are referred to by subfamily L and an identifying number, and similarly receptors are named according to the type of chemokine that they bind and are recognized by subgroup R).⁶⁻⁸

Chemokines act in a coordinated fashion with G-protein-coupled receptors (GPCRs) present on the surface of a variety of mononuclear cell types.⁹⁻¹² The chain length of 340 to 370 amino acids includes a short acidic N-termini that faces the extracellular milieu and seven helical membrane-spanning regions, connected by three extracellular loops linked by disulfide bonds and other three intracellular portions (the second loop exhibits a DRY motif which is mainly responsible for

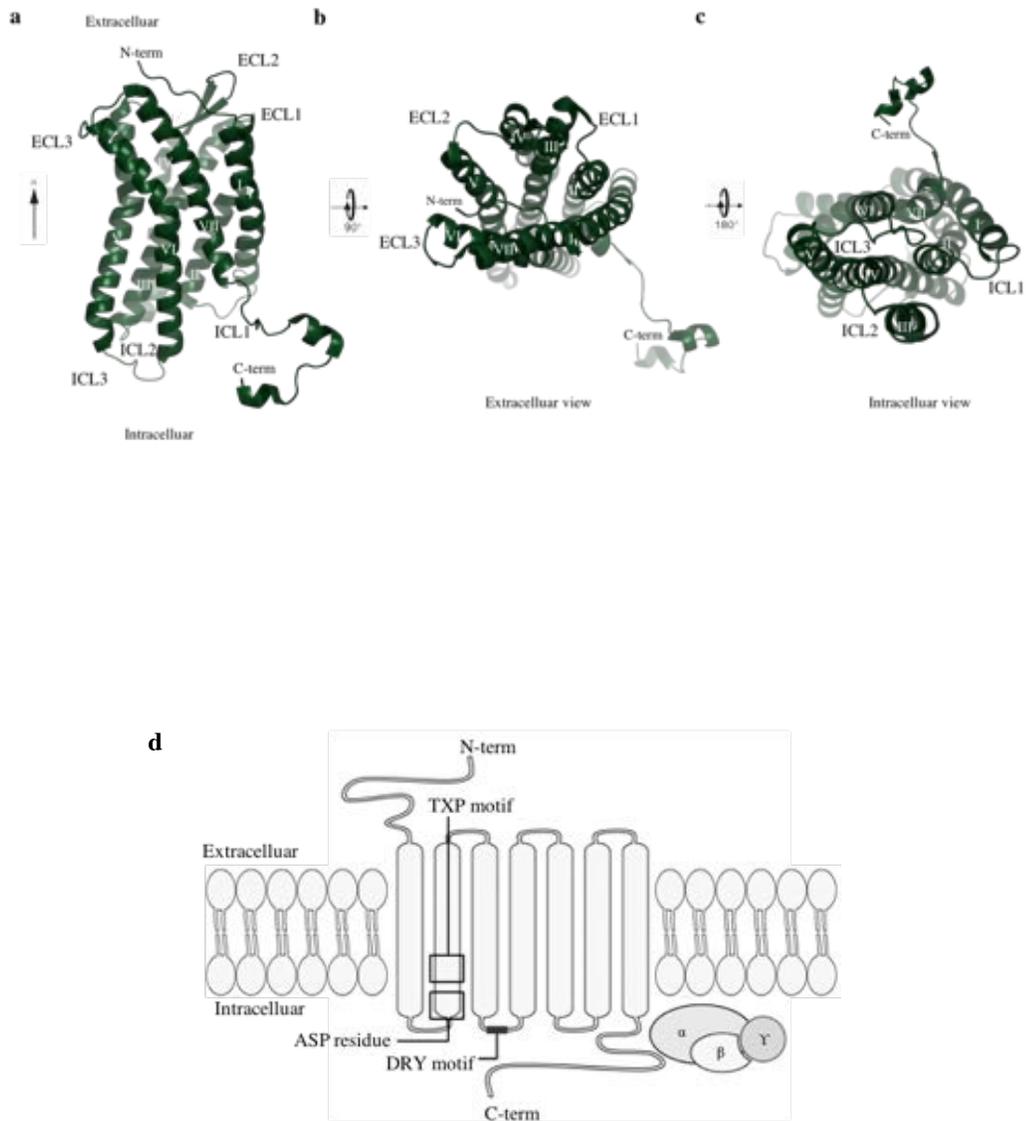


Figure 1.2 a) Backbone representation of CXCR4 showing helices (TM1-TM7), extracellular loops (ECL1-ECL3) and intracellular loops (ICL1-ICL3) b) View from the extracellular side c) View from the intracellular side d) Chemokine receptors are G-protein-coupled receptors (GPCRs) with seven transmembrane domains. Conserved motifs that are critical for signalling include: an aspartic acid residue, a Thr-X-Pro (TXP; X denotes any amino acid) motif in the second transmembrane domain, and an Asp-Arg-Tyr (DRY) motif at the boundary of the third transmembrane domain.

signaling); a cytoplasmic C-terminus tail is enriched for serine and threonine residues, that, when phosphorylated, contribute to desensitization of the receptor. In addition, chemokine ligands also bind to atypical non-G protein-signaling (so-called decoy) chemokine receptors (i.e. DARC and D6) thought to be primarily involved in scavenging a wide variety of ligands due to their lack of ability to activate transduction events.¹³(Figure 1.2)

The downstream physiological effects of chemokine receptor binding are numerous as a result of engagement in many different physiological processes including inflammation, organogenesis, embryogenesis, immune system development and cancer metastasis. Besides the structural criteria, chemokines can be grouped into functional subfamilies.^{14,15}

Inflammatory chemokines are up regulated under inflammatory conditions and control the recruitment of effector leukocytes to injured tissues. Many of the inflammatory chemokines have broad target-cell selectivity and act on cells of the innate, as well as the adaptive, immune system. Homeostatic chemokines, by contrast, are expressed constitutively and normally navigate leukocytes during hematopoiesis.¹⁶ Recent findings, however, indicate that several chemokines cannot be assigned unambiguously to either one of the two functional categories. These ‘dual-function’ chemokines participate in immune defense functions and also target non-effector leukocytes at sites of leukocyte development and immune surveillance.

Moreover, the majority of chemokine receptors display a great deal of redundancy within the superfamily,¹⁷ with different receptors having overlapping activities. Also, multiple chemokines can activate the same receptor. This functional promiscuity is considered a strategy to maintain normal cellular processes if a particular chemokine is defective, although the cellular responses to ligand binding can also be fine-tuned to exert minor differences in biological activity.

1.2 CXCR4-CXCL12 Signaling

The most abundant splice variant of SDF-1 (stromal cell-derived factor-1 α , SDF-1 α), a 68-residue polypeptide, also known as CXCL12, belongs to the CXC chemokines subfamily, which feature two conserved cysteine residues at specific positions in the N-terminal region.¹⁸⁻²⁰ In general, the CXC chemokine family can be divided into two subfamilies, depending on the presence or absence of the highly conserved tripeptide motif Glu-Leu-Arg (ELR) situated at the NH₂ terminal. The presence or

absence of this sequence is important for receptor selectivity and/or downstream signaling. Members of the CXC chemokine family containing the ELR motif attract primarily neutrophils and are potent inducers of angiogenic activity, while chemokines that lack the ELR motif are rather angiostatic and attract lymphocytes.²¹ CXCL12 is an exception to this generalization: while being an ELR– CXC chemokine, it exhibits angiogenic activity, which was initially observed in mice lacking the chemokine gene that have defective formation of large vessels.

Unlike most chemokine receptors which are able to bind several ligands, CXCL12 is the sole chemokine specific for CXCR4. Disruption of the murine gene CXCR4 results in haematopoietic and cardiac defects, and often in fetal lethality in homozygous mutant animals, which have been similarly observed in CXCL12-deficient mice. However, this relationship is near-monogamous. In recent years, the ligands macrophage migration inhibitory factor (MIF1) and trefoil factor family 2 (TFF2), have been shown to be capable of inducing CXCR4-dependent responses in cells. Chemokine mimicry is also involved in the interaction of CXCR4 with vMIP-II. This chemokine-like protein is encoded by the Kaposi's sarcoma-associated herpesvirus 8, sharing 40% similarity with mammalian chemokines. vMIP-II working as an antagonist is employed by the virus to escape the host immune system. In addition, SDF1 has been shown to also bind with the CXCR7 receptor.²² However, unlike CXCR4, CXCR7 fails to induce classical chemokine responses, such as cell migration and intracellular signaling. CXCR7 has been shown to act as a scavenger for CXCL12, thereby regulating CXCL12 mediated CXCR4 activation and function.

Once CXCL12 binds to CXCR4, the intracellular G-protein coupled receptor change its tertiary structure and undergoes dissociation of GTP-bound α and $\beta\gamma$ subunits leading to a cascade of further signaling events. The $G\beta\gamma$ complex activates phospholipase C (PLC), leading to accumulation of inositol (1,4,5)-triphosphate (IP3) and diacylglycerol (DAG). IP3 stimulates the release from intracellular stores of Ca^{2+} , which together with DAG activates protein kinase C leading to the phosphorylation of target proteins, thus contributing to cell migration. On the other hand, the $G\alpha$ monomer activates phosphatidylinositol-3-kinase,²³⁻²⁵ which induces adenylyl cyclase (AC) activation and subsequent modulation of the protein kinase A (PKA) signaling pathway.

1.3 Two side model

As a typical chemokine, the structure of the mature CXCL12 protein is characterized by a core of a C-terminal α -helix packed against the three-stranded β -sheet and a flexible N-terminus.

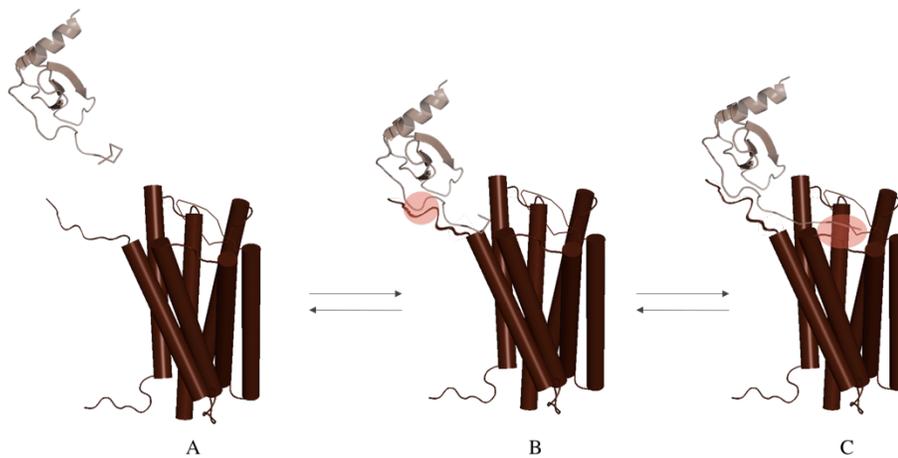


Figure 1.3. A model for interaction of CXCL12 with CXCR4. a) the receptor and ligand separately prior to any interaction between the two; b) Interaction of the CXCL12 RFFESH loop (site 1) with the N-terminal segment of the receptor; c) The N-terminal region (site 2) of CXCL12 bound in groove at the top of the helices.

In a functional formulation, as reported by Crump et al. in a model which dates back to the mid-1990s,²⁶ the affine and specific chemokine globular nucleus first interacts with the CKR N-terminus (chemokine recognition site 1, CRS1), then the contact between residues in the receptor transmembrane domain (CRS2) and the unstructured chemokine N-terminus elicit receptor activation. Thus, the so-called “two-site” model describes chemokine-chemokine receptor interactions functionally (two-step) and spatially (two-site).

The initial interaction involves residues 12–17 of CXCL12 and 2–36 of CXCR4. Data showed that the RFFESH sequence working like a key that permits access to more buried receptor site, provides an initial SDF-1 docking. However, the loop is important for optimal binding, but is not sufficient for receptor activation. Subsequently, the N-terminal region, which becomes structured during the binding, has to penetrate the receptor TM helical bundle: the following change in the receptor conformation allows intracellular G-protein binding and signaling.

Mature CXCL12 is intrinsically unstable and truncated variants of mature CXCL12 missing 2,3,5 or 7 N-terminal aa have been identified in human blood; however, only nontruncated CXCL12 able to induce intracellular calcium flux and chemotaxis of stem cells *in vitro*.²⁷ Moreover, it has been proven that the activity is not affected by the addition of a Gly as well as the modification of few residues within the N-terminus is tolerated, suggesting that the bound region is not completely

hidden. Instead, variants marked by modifications to Lys-1 and Pro-2 are not longer able to induce the required conformational change, resulting in antagonists. Seems in fact that only Lys and Pro residues are directly involved in receptor activation, while changes to other residues result in functional molecules of variable potency.²⁸

The NMR structure of CXCL12 in complex with a 38- residue, sulfotyrosine-containing peptide has been determined (PDB ID: 2K05).²⁹ This structure derived from the CXCR4 N-terminus is considered to represent a part of the “site one” complex proving an insight into interactions between the natural ligand and residues that are absent from the CXCR4 receptor structure. More specifically, the posttranslational sulfation of Tyr-21, Tyr-12, Tyr-7 increases the binding affinity of CXCR4 for CXCL12 through electrostatic interactions between acidic sulfated tyrosines within CXCR4 and basic residues of CXCL12.

Further, two competitive inhibitors of CXCL12 complexed to the receptor allowed the identification of the “site-two” of the chemokine signaling trigger.^{30,31} The CVX15 peptide, rich in basic residues, may trace to some extent the path of the N-terminal signaling peptide of CXCL12 (KPVLSYSR), and the binding site of IT1t may point to the major anchor region for this domain. Both of them block ligand binding by acting as orthosteric competitors of the CXCL12 N-terminal signaling trigger: CVX15 occupies the complete binding cavity and extends out towards the extracellular side of the protein, while the IT1t ligand shows a unique binding mode and it is the first to portray a ligand binding within “the minor ligand pocket”.

1.4 Essential physiological roles of CXCL12/CXCR4 axis

Chemokines were recognized originally for their ability to dictate the retention and homing of hematopoietic stem cells (HSC) in the bone marrow microenvironment as well as the modulation of lymphocyte trafficking. Indeed, studies of CXCR4 and CXCL12 gene knockout mice have shown that CXCL12-CXCR4 signaling is intimately involved in many functions within and outside the immune system.³²

The lifelong production of blood cells depends on HSC_S and their ability to self-renew and to differentiate into all blood lineages. The original pool of HSCs is detected starting in the aorta-gonads-mesonephros (AGM) region in the midgestation of the mouse embryo and then moving to the fetal liver.

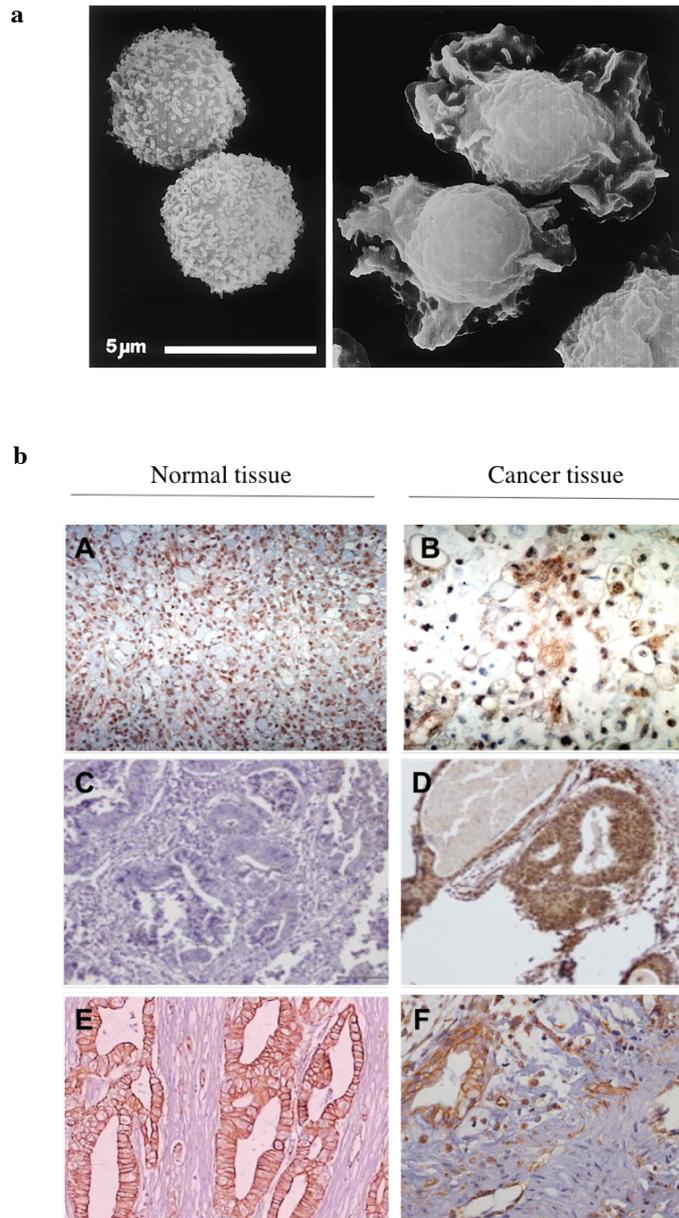


Figure 1.4. a) Shape change of human neutrophil leukocytes that is observed by scanning electron microscopy within seconds after the stimulation of the chemoattractant CXCL12 in cells. (Adapted from Baggiolini M., *Nature*. 1998) b) Expression of CXCR4 in various cancers. Representative IHC staining on gastric cancer tissue; CXCR4 negative (A) and high (B) expression; representative IHC staining on rectal cancer tissue; CXCR4 negative (C) and high (D) expression; representative IHC staining on pancreas cancer tissue; CXCR4 negative (E) and high (F) expression.

Later, HSCs migrate to colonize the bone marrow establishing a stockpile of multipotent cells in specialized niches, where the progeny will come from. Results obtained with genetically engineered mice lacking CXCL12 or CXCR4 have highlighted that ligand-receptor pair is essential for the settlement of HSPCs in the bone marrow during ontogeny. The conditional deletion of CXCL12 raised the depletion of HSC_S cells, resulting in perinatal lethality.³³

Moreover, mutant mice with CXCL12 or CXCR4 gene disruption have been demonstrated other multiple functions of the chemokine and its receptor signaling pathway, that leads to the promotion of the proliferation of B cells, NK cells and plasmacytoid dendritic cells (pDCs), also known as type I interferon (IFN)-producing cells. In contrast to B cells, the generation of the earliest T cell precursors in the adult thymus does not depend on CXCL12-CXCR4 signaling. However, mice deficient of three chemokine receptors, including CXCR4, CCR7, and CCR9 in the embryonic thymus, showed a severely reduced number of T cell precursors.³⁴⁻³⁶

CXCL12/CXCR4 interaction is also required for the normal migration through the hindgut mesentery and the genital ridges colonization genital ridges of Primordial Germ Cells (PGC_S), the founders of sperm or oocytes.³⁷⁻⁴⁰

Studies in mutant zebrafish revealed that PGCs are able to activate the migratory program but fail to undergo directed migration towards the gonad and instead randomly disperse throughout the embryo. Nagasawa's group also described that CXCL12- or CXCR4-deficient mice have irregular formation of the membranous portion of the cardiac ventricular septum and incomplete vascularization of the gastrointestinal tract during ontogeny. Moreover, the CXCL12/CXCR4 axis is essential for the migration of granule cells in appropriate position in the cerebellum network during neurogenesis, reducing also the number of dividing cells in the migratory stream in the mutant embryos. These data therefore underline the role of CXCL12 as a key regulator in the temporal and spatial program of maintenance and proper development of particular types of tissue cells, including stem and progenitor cells.

1.5 Pathological roles of CXCL12/CXCR4 signaling

Although CXCL12-CXCR4 axis evolved to benefit the host, inappropriate regulation or expression of these proteins is involved in pathological conditions, including infection, chronic inflammatory diseases, atherosclerosis, cancer.

In the early 1990s, it was discovered that the receptor CD4 alone is not sufficient for HIV-1 fusion.^{41,42} The entry of HIV into host cells is instead mediated by CXCR4 and CCR5 which acting

as coreceptors, allow the formation of the heterotrimeric complex with the gp120 envelope glycoprotein and trigger fusion between viral and host membranes.

WHIM Syndrome (warts, hypogammaglobulinemia, infections, and myelokathexis syndrome) is a severe clinical condition that has genetic base.^{43,44} As a result of an autosomal dominant mutation, the C-terminus of CXCR4 is truncated by 10 to 19 residues. This causes inability to down-regulate the receptor after it has been activated and thus enhances myeloid cells retention, an outcome known as myelokathexis, the confinement and apoptosis of mature neutrophils in the bone marrow.^{45,46} This gain-of-function mutation increases the susceptibility to bacterial and viral infections.

Increased leukocyte recruitment, activation and accumulation are hallmarks of a variety of autoimmune pathologies that involve chemokines and their receptors.

Sorrentino et al. showed that reduced levels of circulating EPC_s in patients with type 2 diabetes caused by aberrant expression of CXCL12 are correlated with disease severity.^{47,48} On the other hand, CXCL12 mediating chemotaxis of hemangioblasts in the eye destroys the retinal architecture, leading to proliferative retinopathy and blindness.

Moreover, CXCR4-expressing T cells are abundantly detected in the synovial tissues of RA patients, where they increase the inflammation level of the joints leading to the destruction of cartilage and bone.^{49,50}

There is now overwhelming evidence that CXCL12-CXCR4 axis is also involved in the progression of cancer, where it functions in several capacities.^{51,52} (Figure 1.4; Figure 1.5) Primarily, the specific chemokine–receptor pair is involved in tumor metastasis. This is not surprising, in view of its role as chemoattractant in cell migration. And due to their nature of constitutively expressed ligand, these proteins are crucial for the CXCL12 provides a physical address for the secondary destination of the tumour cells, directing these receptor-bearing altered cells to sites of metastasis where the ligand is expressed.

The general mechanisms involved in normal lymphocyte trafficking and tumor migration are similar. In general, chemokines cause cell movement by inducing changes in cytoskeletal structure and dynamics. Cells lose the cell-to-cell contact, polymerization and breakdown of actin leads to formation of protrusions (pseudopods),⁵³ while integrins allow the focal adhesions to the endothelial cells of the vessel wall to help propel the cell forward.⁵⁴⁻⁵⁶

Moreover, CXCR4 is strongly expressed on malignant cells that thereby became able to migrated, unlike their normal counterparts. It has been demonstrated that antibodies against CXCR4 blocked

metastasis in a mouse model of breast cancer, while normally breast epithelial tissue does not express any CXCR4.⁵⁷ Since that seminal publication, the expression of CXCR4 has been observed as a common finding in about 23 different human cancers and correlated with increasing malignancy and aggressive growth. In addition, tumor invasion is supported by the production by the tumor and stromal cells of matrix metalloproteinases (MMPs), which have an interesting interdependent relationship with chemokines since they control the function of one another.⁵⁸⁻⁶⁰ Among them, MMP9 plays a pivotal role in remodeling of the extracellular matrix and its expression can be up-regulated by the activation of CXCL12-CXCR4 signaling through the ERK1/2 pathway.⁶¹

The molecular strategies for survival and growth are often the result of utilizing, and often reprogramming, existing physiological pathways. The expression of the chemokine receptor CXCR4 is regulated both at the transcriptional level and post-transcriptionally through translation, receptor desensitization and internalization. For instance, enhanced CXCR4 translation has been shown to be associated with the oncogene HER2,⁶² which protect the chemokine receptor from ligand-mediated ubiquitination and degradation. Beside, tumor cells response to their environmental hypoxia inducing pathophysiological regulation of CXCR4 transcription via HIF-1 (hypoxia-inducible factor 1).⁶³

Hypoxia or low oxygen tension is the primary factor in the induction of angiogenesis. Like leucocyte migration, blood vessel formation is directional and oriented towards the tumour to increase vascularization and enhance growth. Chemoattractants contribute to this process both by recruiting precursor ECs and by inducing their proliferation. As mentioned above, CXCL12 is an exception to the ELR rule showing angiogenic characteristics. It has been demonstrated that CXCL12 binding to CXCR4 shapes the tumor microenvironment inducing Akt phosphorylation and increasing the production of VEGF in the human breast cancer cell line, MDA-MB-23.⁶⁸

Chemokine signalling sustains cancer cell survival in the primary tumour, but it is likely to be of particular importance to metastasized cells. CXCL12 allows the metastasized cells to “feel at home” in a foreign environment, helping them to grow and resist apoptotic stimuli. Addition of CXCL12 to *in vivo* cultures of numerous types of CXCR4-expressing cancer cells results in their prolonged survival and proliferation.

Administration of antagonists of CXCR4 triggers in cell killing and tumour regression of glioblastoma multiforme-derived tumour cells⁶⁵ and in a B16 murine model of melanoma.⁶⁶ Moreover, it has been demonstrated that stimulation of numerous cancer cells with CXCL12 and other chemokines activates the PI3K/Akt [protein kinase B (PKB)] pathway which is well known to promote survival effects.⁶⁷

1.6 References

1. Laing, K. J.; Secombes, C. J. Chemokines. *Dev. Comp. Immunol.* **2004**, 28, 443-460.
2. Hogaboam, C.M.; Carpenter, K.J.; Schuh, J.M.; Proudfoot, A.A.; Bridger, G.; Buckland, K.F. The therapeutic potential in targeting CCR5 and CXCR4 receptors in infectious and allergic pulmonary disease. *Pharmacol. Ther.* **2005**, 107, 314-28.
3. Koenen, R.R.; Weber, C. Therapeutic targeting of chemokine interactions in atherosclerosis. *Nat Rev Drug Discov.* **2010**, 9(2), 141-153.
4. Christopherson, K.; Hromas, R. Chemokine regulation of normal and pathologic immune responses. *Stem Cells* **2001**, 19, 388-396.
5. Rot, A.; von Andrian, U. H. Chemokines in innate and adaptive host defense: basic chemokine grammar for immune cells. *Annu. Rev. Immunol.* **2004**, 22, 891-928.
6. Murphy, P.M.; Baggiolini, M.; Charo, I.F.; Hebert, C.A.; Horuk, R.; Matsushima, K.; Miller, L.H.; Oppenheim, J.J.; Power, C.A. International union of pharmacology. nomenclature for chemokine receptors. *Pharmacol Rev.* **2000**, 52, 145-176.
7. Zlotnik, A.; Yoshie, O. Chemokines: a new classification system and their role in immunity. *Immunity* **2000**, 12, 121-127.
8. Bazan, J.F.; Bacon, K.B.; Hardiman, G.; Wang, W.; Soo, K.; Rossi, D.; Greaves, D.R.; Zlotnik, A.; Schall, T.J. A new class of membrane-bound chemokine with a CX3C motif. *Nature* **1997**, 385, 640-644.
9. Lodowski, D.T.; Palczewski, K. Chemokine receptors and other G protein-coupled receptors. *Curr Opin HIV AIDS.* **2009**, 4, 88-95
10. Bockaert, J.; Pin, J. P. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J* **1999**, 18, 1723.
11. Murphy, P. M. International Union of Pharmacology. XXX. Update on chemokine receptor nomenclature. *Pharmacol. Rev.* **2002**, 54, 227.
12. Baldwin, J.M. The probable arrangement of the helices in G protein-coupled receptors. *EMBO J* **1993**, 12, 1693-1703.
13. Baggiolini, M. Chemokines and leukocyte traffic. *Nature* **1998**, 392, 565-568.
14. Zlotnik, A.; Burkhardt, A.M.; Homey B. Homeostatic chemokine receptors and organ-specific metastasis. *Nat Rev Immunol.* **2011**, 11, 597-606.
15. Turner, M.D.; Nedjai, B.; Hurst, T.; Pennington, D.J. Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease. *Biochim Biophys Acta - Mol Cell Res.* **2014**, 1843(11), 2563-2582.
16. Zlotnik, A.; Yoshie, O. The Chemokine Superfamily Revisited. *Immunity* **2012**, 36(5), 705-716.
17. Devalaraja, M. N.; Richmond, A. Multiple chemotactic factors: fine control or redundancy? *Trends Pharmacol Sci* **1999**, 20, 151-156.
18. Sun, X.; Cheng, G.; Hao, M.; Zheng, J.; Zhou, X. CXCL12/CXCR4/CXCR7 chemokine axis and cancer progression. *Cancer Metastasis Rev.* **2010**, 29, 709-722.
19. Nagasawa, T. CXC chemokine ligand 12 (CXCL12) and its receptor CXCR4. *J Mol Med.* **2014**, 92(5), 433-439.
20. Hattermann, K.; Mentlein, R. An infernal trio: the chemokine CXCL12 and its receptors CXCR4 and CXCR7 in tumor biology. *Ann Anat.* **2013**, 195(2), 103-110.
21. Strieter, R. M.; Poverini, P. J.; Kunkel, S. L.; Arenberg, D. A.; Burdick, M. D.; Kasper, J.; Dzuiba, J.; Van Damme, J.; Walz, A.; Marriott, D. The functional role of the ELR motif in CXC chemokine-mediated angiogenesis. *J Biol Chem* **1995**, 270, 27348-57.
22. Balabanian, K.; Lagane, B.; Infantino, S.; Chow, K. Y.; Harriague, J.; Moepps, B.; Arenzana-Seisdedos, F.; Thelen, M.; Bachelier, F. The chemokine SDF-1/CXCL12 binds

- to and signals through the orphan receptor RDC1 in T lymphocytes. *J Biol Chem* **2005**, 280, 35760-35766.
23. Bajetto, A.; Bonavia, R.; Barbero, S. Glial and neuronal cells express functional chemokine receptor CXCR4 and its natural ligand stromal cell-derived factor *J Neurochem.* **1999**, 73, 2349-2357
 24. Zheng, J.; Ghorpade, A.; Niemann, D. Lymphotropic virions affect chemokine receptor mediated neural signaling and apoptosis; implications for human immunodeficiency virus type 1-associated dementia. *J Virol.* **1999**, 73, 8256-67
 25. Thelen, M. Dancing to the tune of chemokines. *Nature Immunol.* **2001**; 2, 129-134.
 26. Crump, M.P.; Gong, J.; Loetscher, P. Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 activation from binding and inhibition of HIV-1. *EMBO J.* **1997**, 16(23), 6996-7007.
 27. Richter, R.; Jochheim-Richter, A.; Ciuculescu, F.; Kollar, K.; Seifried, E.; Forssmann, U. Identification and characterization of circulating variants of CXCL12 from human plasma: effects on chemotaxis and mobilization of hematopoietic stem and progenitor cells. *Stem Cells Dev* **2014**, 23, 1959-1974.
 28. Lee, P.E.; Hsiung, G.R.; Seitz, D.; Gill, S.S.; Rochon, P. Cholinesterase inhibitors. *BC Med J.* **2011**, 53(8), 404-408.
 29. Veldkamp, C.T.; Seibert, C. Structural Basis of CXCR4 Sulfotyrosine Recognition by the Chemokine SDF-1/CXCL12 *Sci. Signal* 2008, 1(37), 1-18.
 30. Kleist, A.B.; Getschman, A.E.; Ziarek, J.J. New paradigms in chemokine receptor signal transduction: Moving beyond the two-site model. *Biochem Pharmacol.* **2016**, 114, 53-68.
 31. Wu, B.; Chien, E.Y.T.; Mol, C.D. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science.* **2011**, 330, 1066-1071.
 32. Nagasawa T. CXCL12/SDF-1 and CXCR4. *Front Immunol.* **2015**, 6, 10-12.
 33. Zou, Y.R.; Kottmann, H.; Kuroda, M.; Taniuchi, I.; Littman, D.R. Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development. *Nature.* **1998**, 393, 595-599.
 34. Nagasawa, T. The chemokine CXCL12 and regulation of HSC and B lymphocyte development in the bone marrow niche. *Adv Exp Med Biol.* **2007**, 602, 69-75.
 35. Nagasawa, T.; Nakajima, T.; Tachibana, K.; Iizasa, H.; Bleul, C.C.; Yoshie, O.; Matsushima, K.; Yoshida, N.; Springer, T.A.; Kishimoto, T. Molecular cloning and characterization of a murine pre-B-cell growth- stimulating factor/stromal cell-derived factor 1 receptor, a murine homolog of the human immunodeficiency virus 1 entry coreceptor fusin. *Proc Natl Acad Sci U S A.* **1996**, 93, 14726- 14729.
 36. Nie, Y.; Waite, J.; Brewer, F.; Sunshine, M.J.; Littman, D.R.; Zou, Y.R. The role of CXCR4 in maintaining peripheral B cell compartments and humoral immunity. *J Exp Med.* **2004**, 200, 1145-1156.
 37. Zuccarello, D.; Ferlin, A.; Garolla, A.; Menegazzo, M.; Perilli, L.; Ambrosini, G.; Foresta, C. How the human spermatozoa sense the oocyte: a new role of SDF1- CXCR4 signalling. *Int J Androl* **2011**, 34, e554-65.
 38. Dumstrei, K.; Mennecke, R.; Raz, E. Signaling pathways controlling primordial germ cell migration in zebrafish. *J Cell Sci.* **2004**, 117, 4787-4795.
 39. Knaut, H.; Blader, P.; Strahle, U.; Schier, A.F. Assembly of trigeminal sensory ganglia by chemokine signaling. *Neuron.* **2005**, 47, 653-666.
 40. Knaut, H.; Werz, C.; Geisler, R.; Nusslein-Volhard, C. A zebrafish homologue of the chemokine receptor Cxcr4 is a germ-cell guidance receptor. *Nature* **2003**, 421, 279-282.
 41. Deng, H.; Liu, R.; Ellmeier, W. Identification of a major co-receptor for primary iso- lates of HIV-1. *Nature* **1996**, 381, 661-666.

42. Feng, Y.; Broder, C.C.; Kennedy, P.E. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **1996**, *272*, 872–877.
43. McDermott, D.H.; Gao, J.L.; Liu, Q. Chromothriptic Cure of WHIM Syndrome. *Cell* **2015**, *160*, 686–699.
44. Balabanian, K.; Lagane, B.; Pablos, J.L. WHIM syndromes with different genetic anomalies are accounted for by impaired CXCR4 desensitization to CXCL12. *Blood* **2005**, *105*, 2449–2457.
45. Balabanian, K.; Brotin, E.; Biajoux, V. Proper desensitization of CXCR4 is required for lymphocyte development and peripheral compartmentalization in mice. *Blood* **2012**, *119*, 5722–5730.
46. Kawai, T.; Choi, U.; Cardwell, L. WHIM syndrome myelokathexis reproduced in the NOD/SCID mouse xenotransplant model engrafted with healthy human stem cells transduced with C-terminus-truncated CXCR4. *Blood* **2007**, *109*, 78–84.
47. Egan, C.G.; Lavery, R.; Caporali, F. Generalised reduction of putative endothelial progenitors and CXCR4-positive peripheral blood cells in type 2 diabetes. *Diabetologia*. **2008**, *51*(7), 1296–1305.
48. Egan, C.G.; Caporali, F.; Huqui, A.F.; Zito, M.C.; Focardi, M.; Mondillo, S.; Pierli, C.; Marzilli, M.; Sorrentino, V. Reduced levels of putative endothelial progenitor and CXCR4+ cells in coronary artery disease: kinetics following percutaneous coronary intervention and association with clinical characteristics. *Thromb Haemost.* **2009**, *101*(6), 1138–46.
49. Chung, S-H; Seki, K.; Choi, B-I. CXC chemokine receptor 4 expressed in T cells plays an important role in the development of collagen-induced arthritis. *Arthritis Res Ther.* **2010**, *12*(5), R188.
50. Nanki, T.; Hayashida, K.; El-Gabalawy, H.S. Stromal Cell-Derived Factor-1-CXC Chemokine Receptor 4 Interactions Play a Central Role in CD4+ T Cell Accumulation in Rheumatoid Arthritis Synovium. *J Immunol.* **2000**, *165*(11), 6590–6598.
51. O’Hayre, M.; Salanga, C.L.; Handel, T.M.; Allen, S.J. Chemokines and cancer: migration, intracellular signalling and intercellular communication in the microenvironment. *Biochem J.* **2008**, *409*(3), 635–649.
52. Allavena, P.; Marchesi, F.; Mantovani, A. Current Cancer Therapy Reviews, 2005, 1, 81.
53. Friedl, P.; Wolf, K. Tumour-cell invasion and migration: diversity and escape mechanisms. *Nat. Rev. Cancer* **2003**, *3*, 362–374.
54. Tan, W.; Martin, D.; Gutkind, J. S. The Ga13-Rho signalling axis is required for SDF-1-induced migration through CXCR4. *J. Biol. Chem.* 2006, *281*, 39542–39549.
55. Tanaka, T.; Bai, Z.; Srinoulprasert, Y.; Yang, B. G.; Hayasaka, H.; Miyasaka, M. Chemokines in tumour progression and metastasis. *Cancer Sci.* **2005**, *96*, 317–322.
56. Li, S.; Guan, J. L.; Chien, S. Biochemistry and biomechanics of cell motility. *Annu. Rev. Biomed. Eng.* **2005**, *7*, 105–150.
57. Muller, A.; Homey, B.; Soto, H.; Ge, N.; Catron, D.; Buchanan, M. E.; McClanahan, T.; Murphy, E.; Yuan, W.; Wagner, S. N. Involvement of chemokine receptors in breast cancer metastasis. *Nature* **2001**, *410*, 50–56.
58. Jodele, S.; Blavier, L.; Yoon, J. M; DeClerck, Y. A. Modifying the soil to affect the seed: role of stromal-derived matrix metalloproteinases in cancer progression. *Cancer Metastasis Rev.* **2006**, *25*, 35–43.
59. Giraudo, E.; Inoue, M.; Hanahan, D. An amino-bisphosphonate targets MMP-9-expressing macrophages and angiogenesis to impair cervical carcinogenesis. *J. Clin. Invest.* **2004**, *114*, 623–633.
60. Comerford, I.; Nibbs, R. J. Post-translational control of chemokines: a role for decoy receptors? *Immunol. Lett.* **2005**, *96*, 163–174.

61. Brand, S.; Dambacher, J.; Beigel, F.; Olszak, T.; Diebold, J.; Otte, J.-M.; Goeke, B.; Eichhorst, S. T. CXCR4 and CXCL12 are inversely expressed in colorectal cancer cells and modulate cancer cell migration, invasion and MMP-9 activation. *Exp. Cell. Res.* **2005**, *310*, 117–130
62. Li, Y.M.; Pan, Y.; Wei, Y. Upregulation of CXCR4 is essential for HER2-mediated tumor metastasis. *Cancer Cell.* **2004**, *6*(5), 459-469.
63. Staller, P.; Sulitkova, J.; Lisztwan, J.; Moch, H.; Oakeley, E. J.; Krek, W. Chemokine receptor CXCR4 downregulated by von Hippel–Lindau tumour suppressor pVHL. *Nature* **2003**, *425*, 307–311.
64. Liang, Y.; Bollen, A. W.; Gupta, N. CC chemokine receptor-2A is frequently overexpressed in glioblastoma. *J. Neurooncol.* **2007**, *86*, 153–163
65. Redjal, N.; Chan, J. A.; Segal, R. A.; Kung, A. L. CXCR4 inhibition synergizes with cytotoxic chemotherapy in gliomas. *Clin. Cancer Res.* **2006**, *12*, 6765–6771.
66. Lee, C. H.; Kakinuma, T.; Wang, J.; Zhang, H.; Palmer, D. C.; Restifo, N. P.; Hwang, S. T. Sensitization of B16 tumour cells with a CXCR4 antagonist increases the efficacy of immunotherapy for established lung metastases. *Mol. Cancer Ther.* **2006**, *5*, 2592–2599.
67. Vivanco, I.; Sawyers, C. L. The phosphatidylinositol 3-Kinase AKT pathway in human cancer. *Nat. Rev. Cancer* **2002**, *2*, 489–501.

Chapter 2

Exploring the N-terminal Region of C-X-C Motif Chemokine 12 (CXCL12): Identification of Plasma-Stable Cyclic Peptides as Novel, Potent C-X-C Chemokine Receptor Type 4 (CXCR4) Antagonists

2.1 Introduction

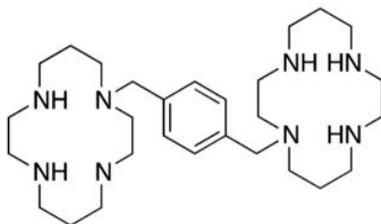
CXCL12/ CXCR4 axis has been implicated in a wide range of diseases such as infections,¹ inflammations,² and cancer.³ Thus, the inhibition of the CXCL12/CXCR4 pathway has emerged as a novel anticancer strategy, widely supported by a large number of *in vitro* and *in vivo* studies.⁴ In particular, CXCR4 has been shown to play a prognostic role in primary colorectal cancer, alone or in association with the vascular endothelial growth factor (VEGF).⁵ In fact, CXCR4 can promote angiogenesis *in vitro* by modulating the secretion of VEGF.^{6,7}

The pursuit of CXCR4 antagonists suitable for anticancer intervention culminated in the identification of plerixafor, previously known as AMD3100 (1,1'-[1,4-phenylenebis-(methylene)]bis[1,4,8,11-tetraazacyclotetradecane]) (Chart 2.1).

In preclinical models, plerixafor (**1**) has shown considerable antimetastatic properties, thus, offering the first proof of concept of CXCR4's importance as anticancer target. **1** has been approved from FDA for hematopoietic stem cell mobilization in patients with non-Hodgkin's lymphoma and multiple myeloma undergoing autologous bone marrow transplantation and resistant to conventional protocol for mobilization.⁸ Recently, new alternative CXCR4 antagonists have been developed and subjected to preclinical and clinical studies.⁹ Among them, the (S)-N1-((1H-benzo[d]imidazol-2-yl)methyl)-N1-(5,6,7,8-tetrahydroquinolin-8-yl)butane-1,4-dia- mine (AMD11070), which is endowed with oral bioavailability, had a trial suspended due to nonclinical reports of hepatotoxicity and histologic findings.¹⁰ Alternatively, some peptide antagonists, such as the peptides 4F-benzoyl-Arg-Arg-Nal-Cys-Tyr-Cit-Lys-D-Lys-Pro-Tyr-Arg-Cit-Cys-Arg-NH₂ (BKT140)¹¹ and H-Lys-Gly-Val-Ser-Leu-Ser-Tyr-Arg-Lys(Lys- Gly-Val-Ser-Leu-Ser-Tyr-Arg*)-NH₂ (CTCE-9908),¹²

also moved to clinical studies, but their application appeared to be partially hampered by *in vivo* instability and other pharmacokinetic issues.^{11,12} Cyclo[Phe-Tyr-Lys(iPr)-D-Arg-2-Nal-Gly-D-Glu]-Lys(iPr)-NH₂ (LY2510924),¹³ a CXCR4 cyclic antagonist peptide currently engaged in two phase 2 studies in small cell lung cancer and renal cancer, shows CD34⁺ cell mobilization at doses ≥ 2.5 mg/day and a tolerable safety profile up to 20 mg/day according to a previous phase 1 study. As part of our ongoing efforts in identifying new CXCR4 antagonists, a CXCL12-derived small cyclic peptide (**2**), which selectively interacts with CXCR4, was identified.¹⁴ Although **2** possesses an interesting *in vitro* and *in vivo* pharmacological profile, it suffers from degradation in biological fluids and its IC₅₀ is in the low micromolar range. Herein, to enhance both the CXCR4 affinity and metabolic stability of **2**, we embarked on a lead optimization campaign. By modifying both the N- and C-termini and through a D-amino acid scan, a new, potent, selective, and plasma stable CXCR4 antagonist was identified (**10**). The novel cyclic peptides were assayed against CXCR4 by measuring their ability to inhibit receptor binding of PE-conjugated-12G5 anti-CXCR4 antibodies and CXCL12-dependent migration in CEM human lymphoblastoid cells, highly expressing CXCR4. Further biological investigations on two different human colon cancer cell lines (HT29 and HCT116) overexpressing CXCR4 confirmed the *in vitro* efficacy and thus the capability to inhibit the CXCR4 receptor as demonstrated through impairment of CXCL12-mediated cell migration and ERK phosphorylation. A favorable selectivity profile toward CXCR4 over CXCR3 and CXCR7 was also found. Finally, the molecular bases for the improved CXCR4 affinity of **10** were unraveled by NMR and molecular modeling. The improved properties of **10**, with respect to **2**, might pave the way for the future preclinical development of our peptide in CXCR4 overexpressing colon cancer.

Chart 2.1. Chemical Structure of 1 (Plerixafor)



2.2 Results and discussion

2.2.1 Design

The peptide (H-Arg-Ala-[Cys-Arg-Phe-Phe-Cys]- COOH) (**2**), previously identified by some of us, inhibits the association between CXCR4 and its ligand CXCL12 at a low micromolar level ($IC_{50} = 6.2 \mu\text{M}$) and in turn impairs CXCR4 function using *in vitro* and *in vivo* systems.¹⁴ However, as already stated in the Introduction, it displays low stability in biological fluids, as in human serum, **2** starts to be metabolized by 30% after only 30 min and is fully converted into the cyclic pentapeptide cyclo[Cys-Arg-Phe-Phe-Cys] within 2 h.

Interestingly, the inactivation of **2** proceeds through the sequential cleavage of the exocyclic Arg1-Ala2 dipeptide at the N-terminal region. Additionally, the IC_{50} of **2** (Table 2.1) needs to be improved to achieve an optimal *in vivo* activity. Interestingly, from a structural point of view, **2** shares common pharmacophoric features such as two aromatic and two positively charged residues, with FC131, a cyclic pentapeptide previously described as a potent CXCR4 antagonist.¹⁵ Thus, we envisaged that the micromolar activity of **1** may be ascribable to a wrong orientation of those pharmacophoric amino acids within the receptor binding site.

To enhance both the peptide CXCR4 affinity and metabolic stability, a systematic modification scheme was designed. Specifically, to increase the stability of **2**, the N-terminus was acetylated (**3** and **4**, Table 2.1) while the C-terminus was amidated (**4** and **5**, Table 2.1). The obtained compounds were validated for inhibiting CXCR4 function through *in vitro* assays evaluating the anti-CXCR4 antibody (12G5) competition and the CXCL12 dependent migration in CEM human lymphoblastoid cells, highly expressing CXCR4. Next, the most promising peptide emerging from this first set of modifications (**3**) was subjected to a D- amino acid scan (Table 2.1, **6–12**) to investigate the role of each residue's chirality in the peptide–receptor interaction. A new, stable, and potent CXCR4 analogue (**10**) was discovered and the specific contribution of each putative pharmacophoric side chains was then investigated by performing an L-Ala scan (Table 2.1, **13–16**).

2.2.2 Chemistry

All the peptides were synthesized adopting a classical Fmoc/Boc solid phase peptide strategy. AM-PS resin functionalized with a RINK-amide linker was used for the CONH2-terminal peptides (**4,5**, Table 2.1), while 2-Cl chlorotrityl resin was selected as solid support for the COOH-terminal peptides (**2, 3, 6–16**, Table 2.1) in order to minimize the racemization of the first cysteine residue

loaded onto resin.¹⁶ The linear eptapeptides were built using standard N-Fmoc/tBu solid-phase peptide synthesis strategy. After the cleavage of the linear eptapeptides from the resin, the resulting free thiol groups were oxidized using Iodine to form a disulfide bridge. The synthetic approach adopted for the generation of the peptides enabled the rapid and efficient synthesis of a small library of peptides.

Table 2.1. Half-Maximal Inhibitory Concentration (IC₅₀, μM, Mean ± SD) of CXCR4 Antagonist Peptides Necessary to Reduce by 50% The Binding to CXCR4 of CXCR4-Specific mAb 12G5 in CCRF-CEM Cells^a

compd	sequence	IC ₅₀ (μM)
2	H-Arg-Ala-[Cys-Arg-Phe-Phe-Cys]-COOH	6.2 ± 1.3
3	Ac-Arg-Ala-[Cys-Arg-Phe-Phe-Cys]-COOH	0.131 ± 0.0014
4	Ac-Arg-Ala-[Cys-Arg-Phe-Phe-Cys]-CONH ₂	1.23 ± 0.28
5	H-Arg-Ala-[Cys-Arg-Phe-Phe-Cys]-CONH ₂	5.9 ± 0.64
6	Ac-Arg-Ala-[Cys-Arg-Phe-Phe-DCys]-COOH	5.9 ± 0.42
7	Ac-Arg-Ala-[Cys-Arg-Phe-DPhe-Cys]-COOH	> 10
8	Ac-Arg-Ala-[Cys-Arg-DPhe-Phe-Cys]-COOH	> 10
9	Ac-Arg-Ala-[Cys-DArg-Phe-Phe-Cys]-COOH	> 10
10	Ac-Arg-Ala-[DCys-Arg-Phe-Phe-Cys]-COOH	0.053 ± 0.004
11	Ac-Arg-DAla-[Cys-Arg-Phe-Phe-Cys]-COOH	> 10
12	Ac-DArg-Ala-[Cys-Arg-Phe-Phe-Cys]-COOH	> 10
13	Ac-Ala-Ala-[DCys-Arg-Phe-Phe-Cys]-COOH	> 10
14	Ac-Arg-Ala-[DCys-Ala-Phe-Phe-Cys]-COOH	> 10
15	Ac-Arg-Ala-[DCys-Arg-Ala-Phe-Cys]-COOH	> 10
16	Ac-Arg-Ala-[DCys-Arg-Phe-Ala-DCys]-COOH	> 10
1		0.006 ± 0.004

^aEach experiment was performed in triplicate

2.2.3 CXCR4 Binding Assay

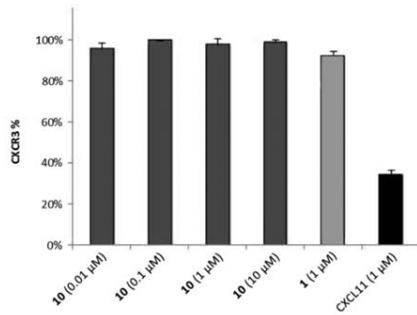
The novel cyclic peptides were assayed against CXCR4 measuring their ability to inhibit receptor binding of PE-conjugated-12G5 anti-CXCR4 antibodies in CEM-CCRF human T leukemia CXCR4 expressing cells.¹⁷ As shown in Table 2.1, the conversion of COOH- into CONH₂-termini, in association or not with acetylation of the N-terminus, minimally affected the peptide efficacy. Indeed, **4** and **5** showed IC₅₀ of 1.23 and 5.9 μM, respectively. Conversely, **3**, which is the acetyl-derivative of **2**, reduced the association of the 12G5 CXCR4-antibody to the receptor with an IC₅₀ of 0.131 μM, thus being about 50-fold more potent than the lead (IC₅₀ = 6.2 μM). Thus, the Nα-acetylation of the Arg1 residue represented a successful modification only in the presence of the COOH-terminus, which appeared to be a critical group for receptor binding. Next, the overall conformation of **3** was modified by performing a D-amino acid scan study. The sequential replacement of each residue with its corresponding D-isomer resulted in analogues (**6–12**), which mostly had significantly weaker binding affinity than **3**, with the lone exception of **10**.

The latter bears a D-Cys instead of L-Cys at the position 3 and showed an IC₅₀ of 0.053 μM that was 120 times lower than that of **2**. In line with our hypothesis, these data clearly pointed out that the conformational changes induced in **10** through the use of a D-amino acid were able to better orient the pharmacophoric side chains to increase the binding. An alanine scan study, restricted to the phenylalanine and arginine residues, was then carried out on **10** to assess their role in the peptide binding. As expected, no or very low CXCR4 binding was observed up to 10 μM concentration of the alanine substituted peptides **13–16**.

2.2.3 CXCR3 and CXCR7 Binding Assays

The exclusive binding of a drug to its target is required to avoid potential side effects resulting from the interactions with other receptors. Here, we assessed the selectivity profile of our most potent peptide (**10**) toward other chemokine receptors, specifically CXCR3 and CXCR7. In fact, we evaluated the anti-CXCR3 and anti-CXCR7 efficacy of **10** on a colon cancer cell line (COLO205) overexpressing CXCR3¹⁸ and a breast cancer cell line (MCF-7) overexpressing CXCR7.¹⁹ As shown in Figure 2.1, **10** did not bind either CXCR7 or CXCR3 up to 10 μM concentration. Interestingly, similar results had been previously obtained for **2**,¹⁴ indicating that the chemical modifications here introduced have not affected the selectivity profile of the lead.

a



b

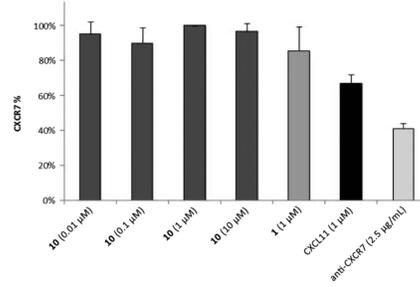
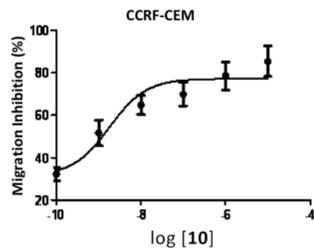


Figure 2.1. Binding assays on CXCR3 and CXCR7. Indirect binding experiments with the anti-CXCR7, clone 11G8, and anti-CXCR3 clone 49801 were conducted in MCF-7 and COLO205, CXCR7, and CXCR3 overexpressing cells, respectively. Data are presented as bar graph showing mean ± SD.

a



b

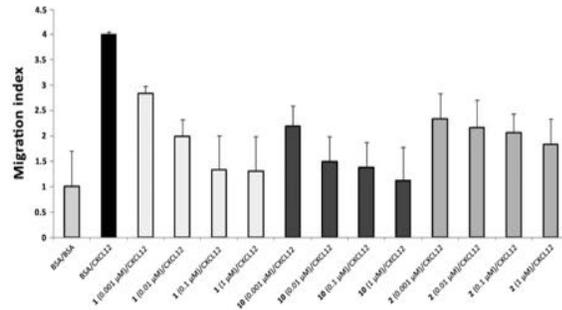


Figure 2.2. Inhibition of CEM cells migration experiments. a) Migrated cells on the lower surface were fixed, stained with H&E, and counted microscopically. The cells were counted in 10 different consecutive high power fields (magnification 200×). The 0.70% of total input cells migrated in buffer alone containing 1% BSA. The results are expressed as migration index (CXCL12 dependent migration/BSA dependent migration). 2 and 10 were compared to the CXCR4 reference antagonist, 1. Data are presented as bar graph showing mean ± SD. b) Migration inhibition curve of 10 in CCRF-CEM cells.

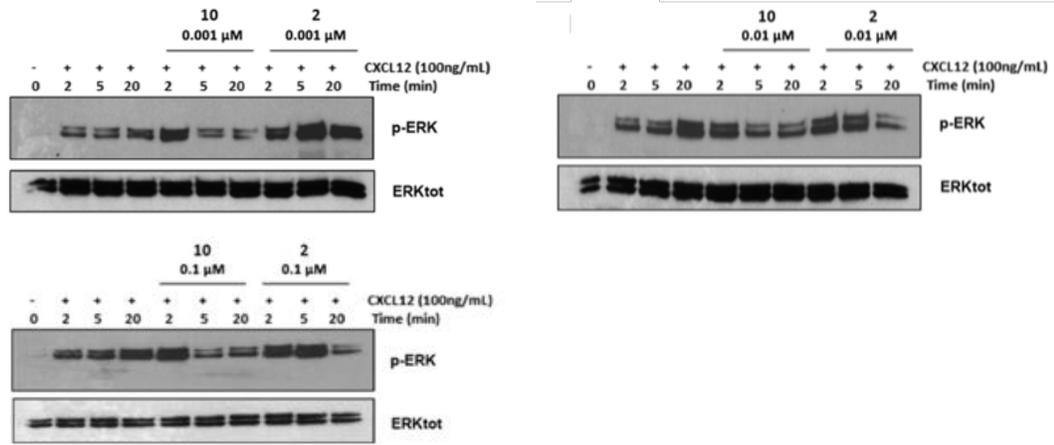


Figure 2.3. CXCL12-mediated p-ERK induction assay. CEM cells were serum-starved and treated with CXCL12 (100 ng/mL) in the presence of 2 and 10.

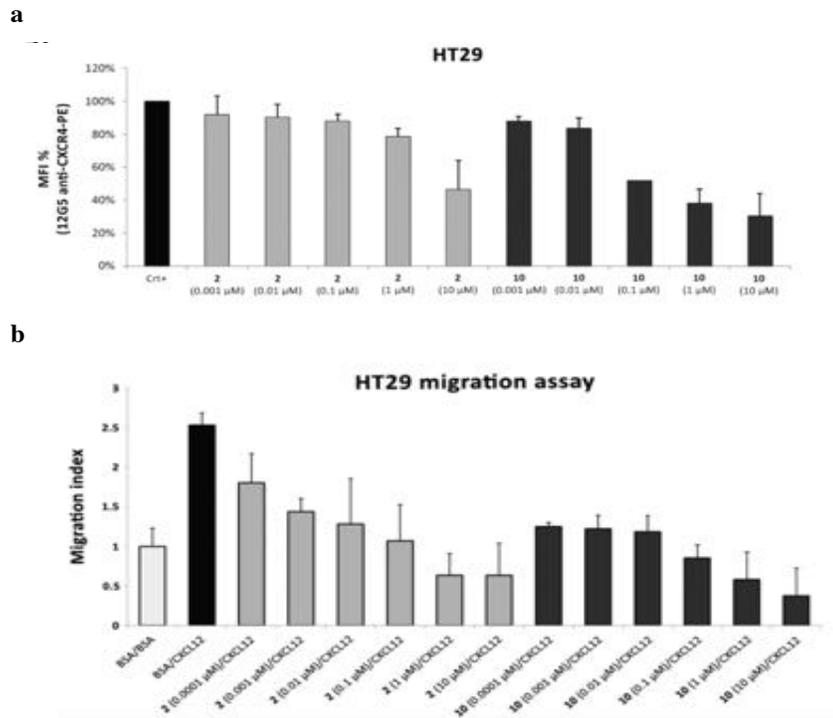


Figure 2.4. a) CXCR4 binding experiments on 2 and 10 in human colon cancer cells. CXCR4 binding was indirectly evaluated through inhibition of 12G5 anti-CXCR4 antibody binding in HT29 human colon cancer CXCR4 expressing cells. b) CXCL12-induced migration assays on human colon cancer cells HT29. 0.81% of total input cells migrated in buffer alone containing 1% BSA.

2.2.4 Cell Migration and p-ERK Induction Experiments

CXCL12 dependent migration and p-ERK induction represent major activities elicited by the CXCL12-induced CXCR4 signaling. For these reasons, we evaluated the ability of the new analogue **10** to inhibit these two CXCR4 functions in comparison with the parent peptide **1**. CXCL12 dependent migration of CEM cells was evaluated in the presence of **10** or **2**. CEM cells were allowed to migrate toward CXCL12 (100 ng/mL) for 18 h. As shown in Figure 2.2a, **10** inhibited CEM migration more efficiently than **2** and in a dose-dependent manner (Figure 2.2b). Subsequently, the effect of **2** and **10** on the CXCL12-dependent p-ERK induction (Figure 2.3) was evaluated, showing that **10** can efficiently inhibit the CXCL12- induced p-ERK more efficiently than **2** (Figure 2.3a–c).

2.2.5 Evaluation of **2** and **10** on Human Colon Cancer Cells

To evaluate the CXCR4 antagonistic activity of **10** in a model suitable for future clinical development, we assessed its efficacy on the CXCR4 overexpressing HT29 human colon cancer cell line. As shown in Figure 2.4, in HT29 cells, **10** binds CXCR4 in a concentration-dependent manner, showing higher potency than **2** (IC_{50} decreased from $5.2 \pm 2.6 \mu\text{M}$ (**2**) to $0.064 \pm 0.024 \mu\text{M}$ (**10**)). Subsequently, the ability of our peptides to inhibit CXCL12 dependent HT29 cells migration was evaluated. Results of these experiments indicated that both **2** and **10** can affect cells migration toward CXCL12, with the latter being effective even at nanomolar concentration (Figure 2.4b).

2.2.6 CXCR4 Internalization Assay

To better understand the effect of **2** and **10** on CXCR4 expression and activation, a receptor internalization assay was carried out on the colon cancer cell line HCT116 engineered to express human GFP- tagged CXCR4 receptor. HCT116-GFP-CXCR4 cells were treated with **2** and **10**, then stimulated with 50 ng/mL CXCL12 to induce receptor internalization and observed by confocal microscopy. As showed in Figure 2.5, in the absence of CXCL12, CXCR4 localizes at both the membrane and cytoplasmic level, with some accumulation at the perinuclear region. After 45 min of stimulation with CXCL12, GFP-CXCR4 is extensively internalized as indicated by the reduction of membrane fluorescence intensity and formation of CXCR4 containing intracellular vesicle. In the presence of **2** or **10**, receptor internalization is impaired as suggested by intracellular vesicles reduction and consistent membrane fluorescence. A slight further reduction in the vesicle size is detected when the cells are treated with **10**.

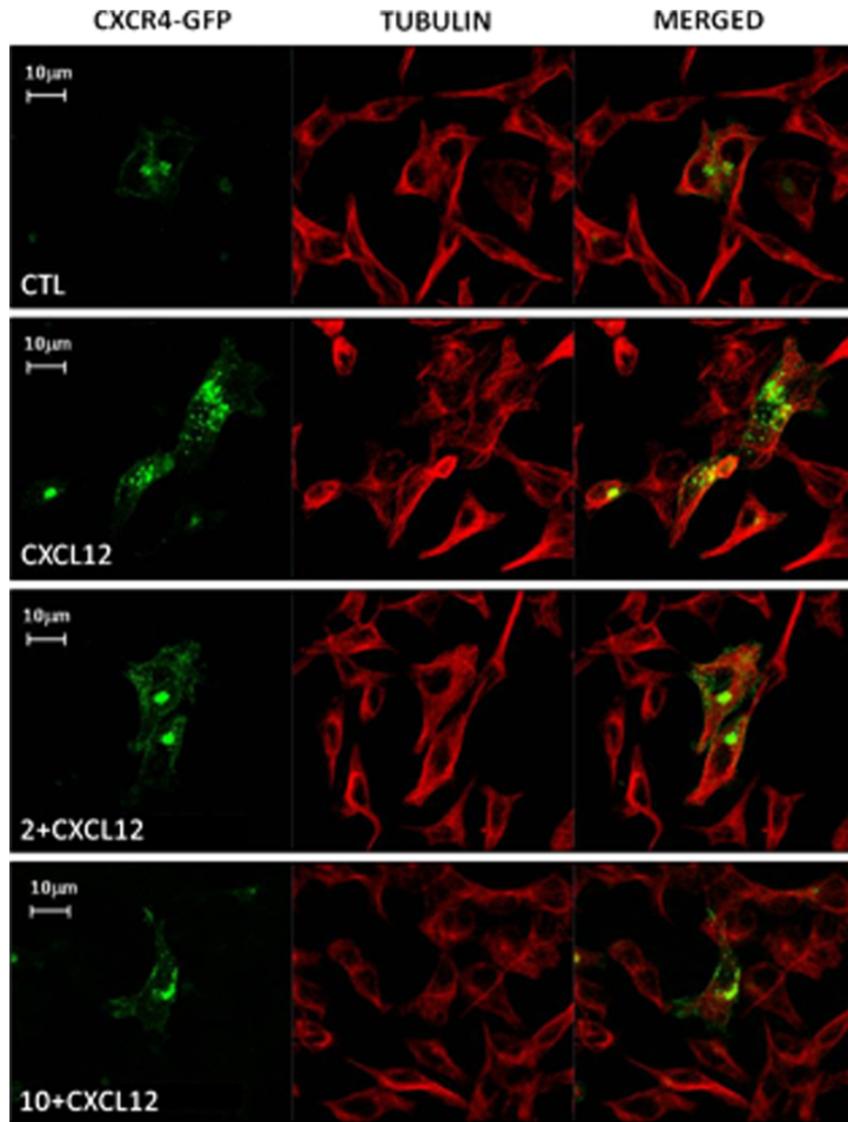
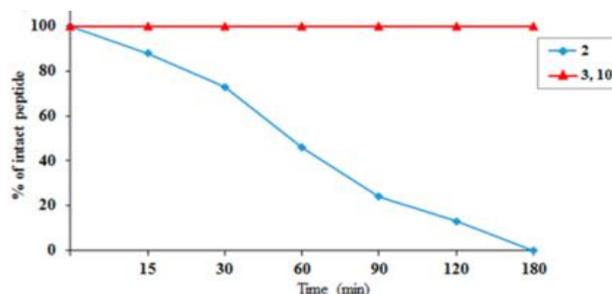


Figure 2.5. CXCR4 internalization assay: HCT116 colon cancer cells were engineered to express GFP-tagged CXCR4; α -tubulin staining was used to define cells cytoplasm and shape. Cells were treated for 45 min with 100 nM of either 2 or 10 and then for 45 min with CXCL12.

2.2.7 Plasma Stability Assay

The effects of **2** modifications on metabolic stability were assessed by incubating **2**, **3**, and **10** in diluted human plasma at 37 °C adopting a protocol described in the literature.²⁰ Time course aliquots were collected, treated with acetonitrile to precipitate the plasma proteins, and analyzed by ESI-RP-HPLC. As shown in Figure 2.6, after 60 min, the concentration of **2** was reduced by 50%, resulting in a cleaved form of the peptide lacking the Arg1 residue and the Arg1–Ala2 dipeptide at the N-terminal region (Figure 2.6b). After 120 min, **2** was fully converted into the cyclic pentapeptide metabolite. Conversely, **3** and **10**, both featuring an N-terminal acetyl cap, were stable up to 180 min. Taken together, these results clearly pointed out that the acetyl group at the N-terminal region was sufficient to protect the peptide against human plasma proteolytic degradation.

a



b

Compound	aa sequence	m/z ^[a]					
		0 min	15 min	30 min	60 min	120 min	180 min
2	RA[CRFFC]	900.4	900.4 ^[b]	900.4	900.4	900.4	
			788.3 ^[c]	788.3	788.3	788.3	788.3
			673.3 ^[d]	673.3	673.3	673.3	673.3
3	Ac-RA[CRFFC]	942.5	942.5	942.5	942.5	942.5	942.5
10	Ac-RA[cRFFC]	942.5	942.5	942.5	942.5	942.5	942.5

[a] Found ESI-MS, [M+H]⁺.

[b] **2**

[c] **2-R**

[d] **2-RA**

Figure 2.6. a) Human plasma stability profiles of **2** and **10** after different intervals of incubation with human plasma. Relative concentrations of peptides were determined by integration of the A230 peaks from RP-HPLC b) ESI-MS characterizations of **2** and **10** at different intervals of incubation.

2.2.8 NMR Spectroscopy

NMR experiments were performed on **2** and **10** in order to understand the conformational differences induced in our peptides by the diverse configuration of the Cys3 residue and by the introduction of an acetyl N-cap. Complete ¹H NMR chemical shift assignments (Tables 2.2 and 2.3) were performed for **2** and **10** according to the Wüthrich²¹ procedure. DQF-COSY,²² TOCSY,²³ and NOESY²⁴ experiments, with the support of the XEASY software package,²⁵ were carried out in 200 mM SDS micellar solution. The employment of SDS micelles to investigate the conformational properties is justified on the basis of their interaction with a membrane receptor. For peptides that bind membrane receptors, such as GPCR, the use of membrane mimetic solution is suggested, assuming a membrane-assisted mechanism of interactions between the peptides and their receptors,²⁶ hence micelle solutions have been extensively used for peptide hormone conformational studies.²⁷ For both peptides, NMR parameters indicated the presence of β -turn structures. In particular, NOE contacts between H α -NH_{i+2} of Arg1 and D-Cys3 and Arg4 and Phe6 suggested turn structures encompassing residues 1–3 and 3–6, respectively. The presence of turn structures is confirmed by low values of temperature coefficients of amide protons ($|\Delta\delta/\Delta T| < 4$ ppb/ K) of Phe5 and Phe6 for both peptides and D-Cys3 for **10** (Tables 2.3 and 2.4). Upfield shifts of side chain protons of the Arg4 and Phe5 residues and NOE contacts between Arg4 and Phe5 and between Phe5 and Phe6 point to a spatial proximity of these couple of side chains in both peptides. For both peptides, NMR constraints from SDS micelle solution were used as the input data for a simulated annealing structure calculation. Two ensembles of well-defined structures were obtained. In fact, the 10 lowest energy structures for **2** (Figure 2.7a) and **10** (Figure 2.7b) showed a backbone RMSD of 0.19 and 0.17 Å, respectively, that satisfied the NMR-derived constraints (violations smaller than 0.40 Å). Predicted turn structures were found in both peptides: a γ -turn along residues 1–3 and a type IV β -turn along residues 3–6. Interestingly, in **2** calculated structures, Cys3 shows a positive value of the dihedral angle ϕ ($\sim 60^\circ$), which could explain the observed constructive replacement with a D-Cys residue. The main difference between the two peptide structures consists in the different spatial disposition of the Arg1 side chain, which is close to Phe5 and Phe6 only in **10** as confirmed by various NOE interactions between Arg1 with Phe5 and acetyl protons with Phe6 observable only in the NOESY spectrum of **10** while missing in that of **2**. Geometry of D-Cys residue and lipophilicity of the acetyl moiety in **10** must be responsible for the observed difference. Interestingly, Ala- and D-scan results clearly demonstrated that both functionality and spatial arrangement of Arg1 side chain matter for the activity of our derivatives. Hence, Arg1 side chain

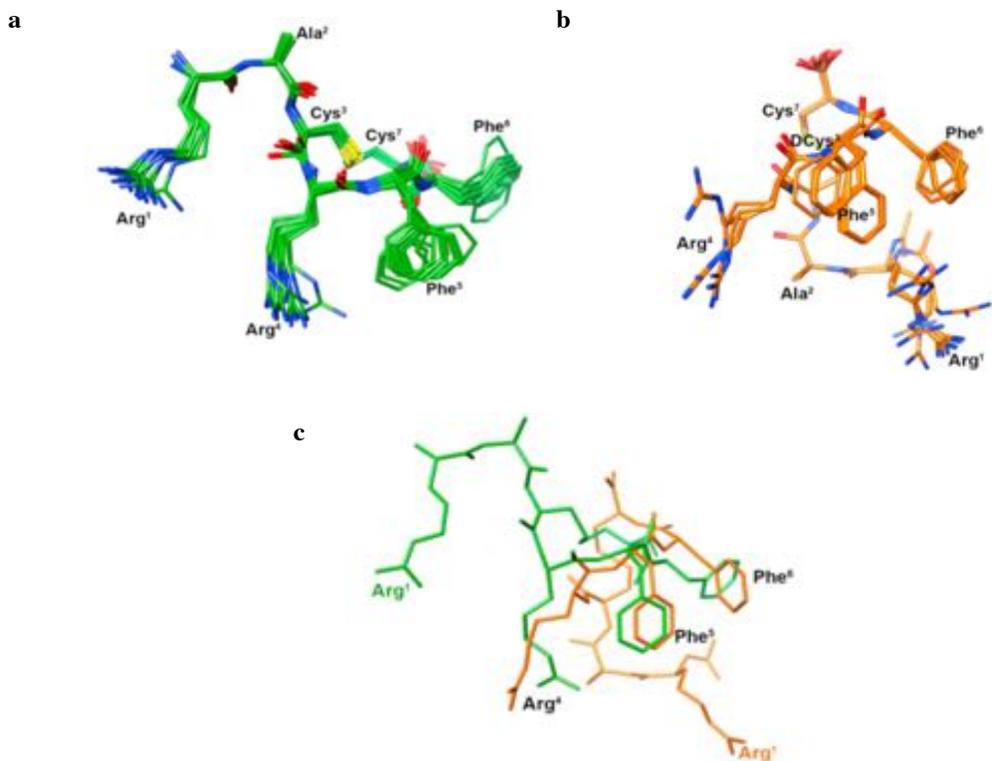


Figure 2.7. Superposition of the 10 lowest energy conformers of 2 (a) and 10 (b). Structure models were superimposed using the backbone heavy atoms. Heavy atoms have different colors (carbon, orange/ green; nitrogen, blue; oxygen, red; sulfur, yellow). Hydrogen atoms are hidden for clarity reasons. (c) Superposition of 2 (green) with 10 (orange) using the three main pharmacophoric points: the guanidino group of the Arg⁴ and the centroids of both the Phe⁵ and Phe⁶ phenyl rings.

reorientation in 10 is likely to contribute to the increased potency of the peptide. This issue has been further investigated by molecular modeling studies.

2.2.9 Molecular Modeling

To investigate at an atomic level the binding to CXCR4 of the most potent peptide of the series, namely 10, extensive computational studies were performed. In the present case, straightforward docking methods did not represent the best option, as they cannot thoroughly sample the large conformational space of mid–high molecular weight peptides upon binding. For this reason, and to take into account the receptor flexibility and the biological environment, we performed an over 100

ns molecular dynamics (MD) simulation on the 10/CXCR4 complex in explicit solvent and membrane. The starting binding conformation was obtained through a manual docking of 10 in the crystal structure of CXCR4 in complex with the 16-mer cyclic peptide CVX15 (PDB code 3OE0). Specifically, a superposition between 10 and CVX15 was accomplished based on previous structure– activity relationship (SAR) studies on potent CXCR4 peptide antagonists, which share one positively charged and two aromatic side chains as common pharmacophoric features.²⁸ (see Experimental Section for details).

At the very beginning of the MD calculations, 10 slightly rearranges at the CXCR4 binding site, assuming a conformation that is almost conserved for the rest of the simulation (Supporting Information, Figure S3). In this pose (Figure 8), the peptide Arg1 and Arg4 side chains make tight salt bridges with the D187 and D97 carboxylate groups, respectively, whereas the Phe5 and Phe6 side chains are buried into a semiaromatic pocket where they can establish favorable interactions with the side chains of Y116, L120, R188, F199, H203, Y255, and I259. Additionally, water-mediated hydrogen bonds are formed between the Arg1 and N37 side chains and between the Phe6 backbone amide and the E288 carboxylate moiety. Interestingly, a number of intramolecular interactions can be also observed, which are likely to stabilize the peptide bioactive conformation. In particular, the Arg1 side chain is hydrogen bonded to the Cys7 terminal carboxylate and to the N-acetyl carbonyl group, while an additional H-bond is established between the D-Cys3 CO and the Cys7 NH groups.

Overall, this interaction pattern is consistent with the low nanomolar IC₅₀ of 10 and in agreement with mutagenesis data reporting residues such D97, H113, Y116, D187, and E288 as critical for ligand binding to CXCR4.^{28,29} Furthermore, comparative studies on 2 (see the Supporting Information for details) helped us to elucidate the molecular bases for the improved CXCR4 binding affinity of 10 with respect to the lead peptide. In fact, some differences can be observed in the binding mode of the two peptides, which are ascribable to the inverted chirality of the Cys3 residue and the absence of the N- acetyl cap in 2 (Supporting Information, Figure S4). For instance, differently from 10, the Arg1 side chain of 2 cannot contact D187 side chain but extends out of the binding site interacting with the lipids polar heads. Moreover, the intramolecular H-bonds formerly described for 10 (i.e., between the Arg1 side chain and the Cys7 terminal carboxylate) are not observed in the case of 2, turning out in a lower conformation stability that might negatively impact the binding.

Table 2.2. NMR Resonance Assignments^a of 2 in SDS-d₂₅ Solution at 298 K.

residue	NH (³ J _{αN} , -Δδ/ΔT) ^b	C ^α H	C ^β H	Others
Arg ¹		4.17	1.98, 2.05	1.58, 1.70(γ); 3.22(δ); 7.24(ε);
Ala ²	8.62 (6.1, 6.5)	4.29	1.38	
Cyss ³	8.32(7.3, 6.7)	4.50	3.03, 3.30	
Arg ⁴	8.28 (8.2, 5.3)	4.06	1.20, 1.40	0.99, 1.12(γ); 2.90(δ); 6.99(ε)
Phe ⁵	7.91 (9.0, 2.6)	4.54	2.55	7.09(δ); 7.18(ε)
Phe ⁶	8.02 (8.2, 3.3)	4.64	3.03, 3.33	7.35(δ); 7.24(ε)
Cyss ⁷	7.99 (7.8, 5.0)	4.72	3.07, 3.31	

^a Obtained at pH = 5.5, with TSP (δ 0.00 ppm) as reference shift. Chemical shifts are accurate to ±0.02 ppm.

^b ³J_{αN} coupling constants in Hz. -Δδ/ΔT = temperature coefficients (ppb/K) calculated in the range 298K – 313 K.

Table 2.3. NMR Resonance Assignments^a of 2 in SDS-d₂₅ Solution at 298 K.

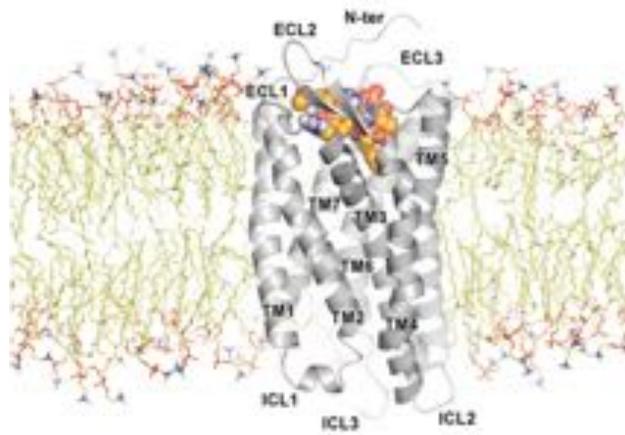
residue	NH (³ J _{αN} , -Δδ/ΔT) ^b	C ^α H	C ^β H	Others
Arg ¹	8.14 (9.7, 5.3)	4.22	1.84, 1.97	1.74(γ); 3.25(δ); 7.14(ε);
Ala ²	8.45 (7.4, 6.0)	4.27	1.45	
DCyss ³	7.95 (6.4, 2.9)	4.41	3.18, 3.26	
Arg ⁴	8.13 (7.5, 4.0)	4.02	1.21, 1.47	1.05(γ); 2.93(δ); 6.97(ε)
Phe ⁵	7.99 (9.4, 3.4)	4.54	2.65	7.04(δ); 7.16(ε); 7.11(ζ)
Phe ⁶	8.23 (8.2, 3.8)	4.52	3.05, 3.35	7.32(δ); 7.25(ε);
Cyss ⁷	7.92 (b.s, 3.4)	4.73	3.12, 3.33	

^a Obtained at pH = 5.5, with TSP (δ 0.00 ppm) as reference shift. Chemical shifts are accurate to ±0.02 ppm.

^b ³J_{αN} coupling constants in Hz. Δδ/ΔT = temperature coefficients (ppb/K) calculated in the range 298K – 313 K.

Further signals: CH₃CO, 2.12 ppm.

a



b

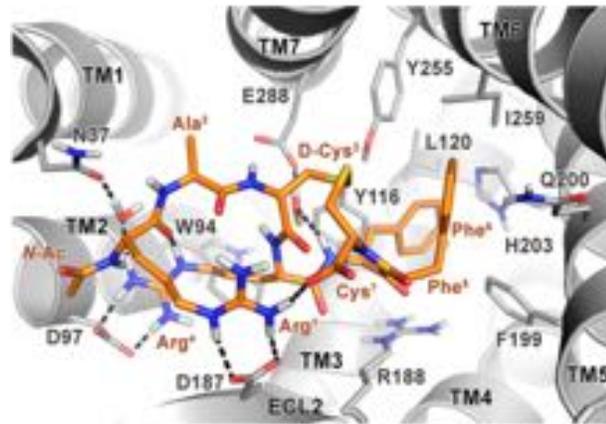


Figure 2.8. (a) Side view of the predicted CXCR4/10 complex embedded in the POPC bilayer. CXCR4 is represented as gray cartoons, while the peptide is shown as orange spheres. Lipids are shown as yellow sticks. (b) Binding mode of 10 (orange sticks) at the CXCR4 receptor (gray cartoons) obtained through over 100 ns long MD simulations. Receptor amino acids and waters important for peptide binding are shown as sticks. Hydrogen bonds are displayed as dashed black lines. Nonpolar hydrogens are omitted for clarity reasons.

The interaction models presented here also allow rationalizing the effects of all the introduced chemical modifications on our peptides potency. In particular: (i) mutation into Ala of Arg1, Arg4, Phe5, and Phe6 (**13–16**) is detrimental for peptide affinity toward CXCR4 because the side chains of these residues represent the major anchor points for receptor binding; (ii) the chiralities of the amino acids occupying positions from 3 to 7 (**6–12**), particularly D-Cys3 and Cys7, are crucial to stabilize the binding of the peptide pharmacophoric groups at the CXCR4 receptor; (iii) amidation of the Cys7 terminal carboxylate and/or removal of the N-Ac cap (**2–5**) lower the peptide affinity toward CXCR4 because such groups participate in intramolecular interactions that contribute to stabilize the peptide binding conformation.

Finally, our model suggests that at least one of the two Phe residues in **10** might be substituted with bulkier aromatic residues, such as 1- or 2- Nal, providing some hints for a further optimization of our peptides potency.

2.3 Conclusion

In recent years, the CXCL12/CXCR4 pathway has emerged as a major target for the therapy of multiple diseases including various types of blood and solid tumors. For this reason, a number of CXCR4 antagonists are now used or are entering as diagnostic and therapeutic tools for cancer treatment. However, in spite of the medicinal chemists' efforts, the majority of the available CXCR4 antagonists suffer pharmacokinetic or toxicity issues, prompting the design of new anti-CXCR4 ligands with improved pharmacological properties. In this scenario, some of us recently developed a CXCL12-derived cyclic peptide as selective CXCR4 antagonist (**2**). Herein, through a rational lead optimization strategy, we improved both the binding potency which was in the low micromolar range and the metabolic stability of **2**, thus obtaining a selective and plasma stable CXCR4 antagonist (**10**). Ala-amino acid scan, extensive NMR, and computational studies elucidated the binding mode of **10** to CXCR4, providing reasons for the increased binding with respect to **2**. In parallel, cell-based assays revealed that **10** inhibits both the CXCL12-induced migration and ERK phosphorylation, which are hallmarks for monitoring CXCR4 functions. Furthermore, we were able to demonstrate that **10** impairs CXCR4 internalization in human colon cancer cells overexpressing CXCR4, which might lay the foundations for a future preclinical investigation of our novel peptide. In this perspective, in cell NMR studies are now in progress in our laboratories to ascertain the competition mechanism between **10** and **1** in order to get more insights into the pharmacodynamics

properties of our peptides. Nonetheless, the modeling data herein reported pave the way for a further optimization of the **10** IC₅₀ and another round of synthesis is ongoing accordingly. Finally, the high stability of 10 in human plasma is surely an encouraging starting point for more extensive in vivo pharmacokinetic studies. In this light, a GPCR gene expression profiling could be particularly important in the evaluation of possible toxicities beyond those arising from CXCR4 blockade.

2.4 References

1. (a) Feng, Y.; Broder, C. C.; Kennedy, P. E.; Berger, E. A. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **1996**, *272*, 872–877. (b) Vicenzi, E.; Liò, P.; Poli, G. The Puzzling Role of CXCR4 in Human Immunodeficiency Virus Infection. *Theranostics* **2013**, *3*, 18–25. (c) Demmer, O.; Frank, A. O.; Hagn, F.; Schottelius, M.; Marinelli, L.; Cosconati, S.; Brack-Werner, R.; Kremb, S.; Wester, H. J.; Kessler, H. A Conformationally Frozen Peptoid Boosts CXCR4 Affinity and Anti- HIV Activity. *Angew. Chem., Int. Ed.* **2012**, *51*, 8110–8113.
2. Yamada, M.; Kubo, H.; Kobayashi, S.; Ishizawa, K.; He, M.; Suzuki, T.; Fujino, N.; Kunishima, H.; Hatta, M.; Nishimaki, K.; Aoyagi, T.; Tokuda, K.; Kitagawa, M.; Yano, H.; Tamamura, H.; Fujii, N.; Kaku, M. The increase in surface CXCR4 expression on lung extravascular neutrophils and its effects on neutrophils during endotoxin-induced lung injury. *Cell. Mol. Immunol.* **2011**, *8*, 305–314.
3. (a) Domanska, U. M.; Kruizinga, R. C.; Nagengast, W. B.; Timmer-Bosscha, H.; Huls, G.; de Vries, E. G.; Walenkamp, A. M. A review on CXCR4/CXCL12 axis in oncology: no place to hide. *Eur. J. Cancer* **2013**, *49*, 219–230. (b) Müller, A.; Homey, B.; Soto, H.; Ge, N.; Catron, D.; Buchanan, M. E.; Mcclanahan, T.; Murphy, E.; Yuan, W.; Wagner, S.M.; Barrera, J.L.; Mohar, A.; Veraástegui, E.; Zlotnik, A. Involvement of chemokine receptors in breast cancer metastasis. *Nature* **2001**, *410*, 50–56. (c) Demmer, O.; Dijkgraaf, I.; Schumacher, U.; Marinelli, L.; Cosconati, S.; Gourni, E.; Wester, H. J.; Kessler, H. Design, synthesis, and functionalization of dimeric peptides targeting chemokine receptor CXCR4. *J. Med. Chem.* **2011**, *54*, 7648–7662.
4. Burger, J. A.; Peled, A. CXCR4 antagonists: targeting the microenvironment in leukemia and other cancers. *Leukemia* **2009**, *23*, 43–52.
5. (a) Ottaiano, A.; Franco, R.; Aiello Talamanca, A.; Liguori, G.; Tatangelo, F.; Delrio, P.; Nasti, G.; Barletta, E.; Facchini, G.; Daniele, B.; Di Blasi, A.; Napolitano, M.; Ierano, C.; Calemma, R.; Leonardi, E.; Albino, V.; De Angelis, V.; Falanga, M.; Boccia, V.; Capuozzo, M.; Parisi, V.; Botti, G.; Castello, G.; Iaffaioli, R. V.; Scala, S. Overexpression of both CXC chemokine receptor 4 and vascular endothelial growth factor proteins predicts early distant relapse in stage II-III colorectal cancer patients. *Clin. Cancer Res.* **2006**, *12*, 2795–2803. (b) D’Alterio, C.; Avallone, A.; Tatangelo, F.; Delrio, P.; Pecori, B.; Cella, L.; Pelella, A.; D’Armiento, F. P.; Carlomagno, C.; Bianco, F.; Silvestro, L.; Pacelli, R.; Napolitano, M.; Iaffaioli, R. V.; Scala, S. A prognostic model comprising pT stage, N status, and the chemokine receptors CXCR4 and CXCR7 powerfully predicts outcome in neoadjuvant resistant rectal cancer patients. *Int. J. Cancer* **2014**, *135*, 379–390. (c) Ingold, B.; Schulz, S.; Budczies, J.; Neumann, U.; Ebert, M. P.; Weichert, W.; Röcken, C. The role of vascular CXCR4 expression in colorectal carcinoma. *Histopathology* **2009**, *55*, 576–586. (d) Wang, X.; Ding, X.; Nan, L.; Wang, Y.; Wang, J.; Yan, Z.; Zhang, W.; Sun, J.; Zhu, W.; Ni, B.; Dong, S.; Yu, L. Investigation of the roles of exosomes in colorectal cancer liver metastasis. *Oncol. Rep.* **2015**, *33*, 2445–2453.

6. (a) Zlotnik, A. New insights on the role of CXCR4 in cancer metastasis. *J. Pathol.* **2008**, 215, 211–213. (b) Fulton, A. M. The chemokine receptors CXCR4 and CXCR3 in cancer. *Curr. Oncol. Rep.* **2009**, 11, 125–131. (c) Teicher, B. A.; Fricker, S. P. CXCL12 (SDF-1)/CXCR4 Pathway in Cancer. *Clin. Cancer Res.* **2010**, 16, 2927–2931. (d) Mukherjee, D.; Zhao, J. The Role of chemokine receptor CXCR4 in breast cancer metastasis. *Am. J. Cancer Res.* **2013**, 3, 46–57.
7. (a) De Clercq, E.; Yamamoto, N.; Pauwels, R.; Balzarini, J.; Witvrouw, M.; De Vreese, K.; Debyser, Z.; Rosenwirth, B.; Peichl, P.; Datema, R. Highly potent and selective inhibition of human immunodeficiency virus by the bicyclam derivative JM3100. *Antimicrob. Agents Chemother.* **1994**, 38, 668–674. (b) Donzella, G. A.; Schols, D.; Lin, S. W.; Este, J. A.; Nagashima, K. A.; Maddon, P. J.; Allaway, G. P.; Sakmar, T. P.; Henson, G.; De Clercq, E.; Moore, J. P. AMD3100, a small molecule inhibitor of HIV-1 entry via the CXCR4 co-receptor. *Nat. Med.* **1998**, 4, 72–77. (c) De Clercq, E. The bicyclam AMD3100 story. *Nat. Rev. Drug Discovery* **2003**, 2, 581–587.
8. (a) Vose, J. M.; Ho, A. D.; Coiffier, B.; Corradini, P.; Khouri, I.; Sureda, A.; Van Besien, K.; Dipersio, J. Advances in mobilization for the optimization of autologous stem cell transplantation. *Leuk. Lymphoma* **2009**, 50, 1412–1421. (b) Dugan, M. J.; Maziarz, R. T.; Bensinger, W. I.; Nademane, A.; Liesveld, J.; Badel, K.; Dehner, C.; Gibney, C.; Bridger, G.; Calandra, G. Safety and preliminary efficacy of plerixafor (Mozobil) in combination with chemotherapy and G-CSF: an open-label, multicenter, exploratory trial in patients with multiple myeloma and non-Hodgkin’s lymphoma undergoing stem cell mobilization. *Bone Marrow Transplant.* **2010**, 45, 39–47.
9. (a) Hendrix, C. W.; Collier, A. C.; Lederman, M. M.; Schols, D.; Pollard, R. B.; Brown, S.; Jackson, J. B.; Coombs, R. W.; Glesby, M. J.; Flexner, C. W.; Bridger, G. J.; Badel, K.; MacFarland, R. T.; Henson, G. W.; Calandra, G. Safety, pharmacokinetics, and antiviral activity of AMD3100, a selective CXCR4 receptor inhibitor, in HIV-1 infection. *JAIDS, J. Acquired Immune Defic. Syndr.* **2004**, 37, 1253–1262. (b) Micallef, I. N.; Stiff, P. J.; DiPersio, J. F.; Maziarz, R. T.; McCarty, J. M.; Bridger, G.; Calandra, G. Successful stem cell remobilization using plerixafor (mozobil) plus granulocyte colony-stimulating factor in patients with non-Hodgkin lymphoma: results from the plerixafor NHL phase 3 study Rescue Protocol. *Biol. Blood Marrow Transplant.* **2009**, 15, 1578–1586.
10. Stone, N. D.; Dunaway, S. B.; Flexner, C.; Tierney, C.; Calandra, G. B.; Becker, S.; Cao, Y. J.; Wiggins, I. P.; Conley, J.; MacFarland, R. T.; Park, J. G.; Lalama, C.; Snyder, S.; Kallungal, B.; Klingman, K. L.; Hendrix, C. W. Multiple-dose escalation study of the safety, pharmacokinetics, and biologic activity of oral AMD070, as a selective CXCR4 receptor inhibitor, in human subjects. *Antimicrob. Agents Chemother.* **2007**, 51, 2351–2358.
11. Peled, A.; Abraham, M.; Avivi, I.; Rowe, J. M.; Beider, K.; Wald, H.; Tiomkin, L.; Ribakovskiy, L.; Riback, Y.; Ramati, Y.; Aviel, S.; Galun, E.; Shaw, H. L.; Eizenberg, O.; Hardan, I.; Shimoni, A.; Nagler, A. The high affinity CXCR4 antagonist BKT140 is safe and induces a robust mobilization of human CD34+ cells in patient with multiple myeloma.

- Clin. Cancer Res.* **2014**, *20*, 469–479.
12. Kwong, J.; Kulbe, H.; Wong, D.; Chakravarty, P.; Balkwill, F. An antagonist of chemokine receptor CXCR4 induces mitotic catastrophe in ovarian cancer cell. *Mol. Cancer Ther.* **2009**, *8*, 1893–1905.
 13. (a) Peng, S. B.; Zhang, X.; Paul, D.; Kays, L. M.; Gough, W.; Stewart, J.; Uhlik, M. T.; Chen, Q.; Hui, Y. H.; Zamek-Gliszczynski, M. J.; Wijsman, J. A.; Credille, K. M.; Yan, L. Z. Identification of LY2510924, a novel cyclic peptide CXCR4 antagonist that exhibits antitumor activities in solid tumor and breast cancer metastatic models. *Mol. Cancer Ther.* **2015**, *14*, 480–490. (b) Galsky, M. D.; Vogelzang, N. J.; Conkling, P.; Raddad, E.; Polzer, J.; Roberson, S.; Stille, J. R.; Saleh, M.; Thornton, D. A phase I trial of LY2510924, a CXCR4 peptide antagonist, in patients with advanced cancer. *Clin. Cancer Res.* **2014**, *20*, 3581–3588.
 14. Portella, L.; Vitale, R.; De Luca, S.; D’Alterio, C.; Ierano, C.; Napolitano, N.; Riccio, A.; Polimeno, M. N.; Monfregola, L.; Barbieri, A.; Luciano, A.; Ciarmiello, A.; Arra, C.; Castello, G.; Amodeo, P.; Scala, S. Preclinical Development of a Novel Class of CXCR4 Antagonist Impairing Solid Tumors Growth and Metastases. *PLoS One* **2013**, *8*, e74548.
 15. Fujii, N.; Oishi, S.; Hiramatsu, K.; Araki, T.; Ueda, S.; Tamamura, H.; Otaka, A.; Kusano, S.; Terakubo, S.; Nakashima, H.; Broach, J. A.; Trent, J. O.; Wang, Z.; Peiper, S. C. Molecular-size reduction of a potent CXCR4-chemokine antagonist using orthogonal combination of conformation- and sequence-based libraries. *Angew. Chem., Int. Ed.* **2003**, *42*, 3251–3253.
 16. Siedler, F.; Weyher, E.; Moroder, L. Cysteine racemization in peptide synthesis: a new and easy detection method. *J. Pept. Sci.* **1996**, *2*, 271–275.
 17. Fricker, S. P.; Anastassov, V.; Cox, J.; Darkes, M. C.; Grujic, O.; Idzan, S. R.; Labrecque, J.; Lau, G.; Mosi, R. M.; Nelson, K. L.; Qin, L.; Santucci, Z.; Wong, R. S. Y. Characterization of the molecular pharmacology of AMD3100: a specific antagonist of the G-protein coupled chemokine receptor, CXCR4. *Biochem. Pharmacol.* **2006**, *72*, 588–596.
 18. Kawada, K.; Hosogi, H.; Sonoshita, M.; Sakashita, H.; Manabe, T.; Shimahara, Y.; Sakai, Y.; Takabayashi, A.; Oshima, M.; Taketo, M. M. Chemokine receptor CXCR3 promotes colon cancer metastasis to lymph nodes. *Oncogene* **2007**, *26*, 4679–4688.
 19. Ribas, R.; Ghazoui, Z.; Gao, Q.; Pancholi, S.; Rani, A.; Dunbier, A.; Dowsett, M.; Martin, L. A. Identification of chemokine receptors as potential modulators of endocrine resistance in oestrogen receptor- positive breast cancers. *Breast Cancer Res.* **2014**, *16*, 447.
 20. Mansfeld, F. M.; Toth, I. Synthesis and Plasma Stability of Disulfide-Bridged Cyclic Endomorphin-1 Derivatives. *Int. J. Org. Chem.* **2012**, *2*, 1–6.
 21. Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons, Inc: New York, 1986.
 22. (a) Piantini, U.; Sorensen, O. W.; Ernst, R. R. Multiple Quantum Filters for Elucidating NMR Coupling Network. *J. Am. Chem. Soc.* **1982**, *104*, 6800–6801. (b) Marion, D.; Wüthrich, K. Application of Phase Sensitive Two-Dimensional Correlated Spectroscopy (COSY) for Measurements of ¹H-¹H Spin-Spin Coupling Constants in Proteins. *Biochem.*

- Biophys. Res. Commun.* **1983**, 113, 967–974.
23. Braunschweiler, L.; Ernst, R. R. Coherence transfer by isotropic mixing: Application to Proton Correlation Spectroscopy. *J. Magn. Reson.* **1983**, 53, 521–528.
 24. Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **1979**, 71, 4546–4553.
 25. Bartels, C.; Xia, T.; Billeter, M.; Guentert, P.; Wüthrich, K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* **1995**, 6, 1–10.
 26. (a) Moroder, L.; Romano, R.; Guba, W.; Mierke, D. F.; Kessler, H.; Delporte, C.; Winand, J.; Christophe, J. New evidence for a membrane bound pathway in hormone receptor binding. *Biochemistry* **1993**, 32, 13551–13559. (b) Sargent, D. F.; Schwyzer, R. Membrane lipid phase as catalyst for peptide-receptor interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, 83, 5774–5778.
 27. (a) Di Cianni, A.; Carotenuto, A.; Brancaccio, D.; Novellino, E.; Reubi, J. C.; Beetschen, K.; Papini, A. M.; Ginanneschi, M. Novel octreotide dicarba-analogues with high affinity and different selectivity for somatostatin receptors. *J. Med. Chem.* **2010**, 53, 6188–6197. (b) Grieco, P.; Carotenuto, A.; Campiglia, P.; Gomez-Monterrey, I.; Auriemma, L.; Sala, M.; Marcozzi, C.; d’Emmanuele di Villa Bianca, R.; Brancaccio, D.; Rovero, P.; Santicioli, P.; Meini, S.; Maggi, C. A.; Novellino, E. New insight into the binding mode of peptide ligands at urotensin-II receptor: structure–activity relationships study on P5U and urantide. *J. Med. Chem.* **2009**, 52, 3927–3940.
 28. Thiele, S.; Mungalpara, J.; Steen, A.; Rosenkilde, M. M.; Vabø, J. Determination of the binding mode for the cyclo-pentapeptide CXCR4 antagonist FC131 using a dual approach of ligand modifications and receptor mutagenesis. *Br. J. Pharmacol.* **2014**, 171, 5313–5329.
 29. Wu, B.; Chien, E. Y.; Mol, C. D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Brooun, A.; Wells, P.; Bi, F. C.; Hamel, D. J.; Kuhn, P.; Handel, T. M.; Cherezov, V.; Stevens, R. C. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **2010**, 330, 1066–1071.
 30. Kaiser, E.; Colecott, R. L.; Bossinger, C. D.; Cook, P. I. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal. Biochem.* **1970**, 34, 595–598.
 31. Hancock, W. S.; Battersby, J. E. A new micro-test for the detection of incomplete coupling reactions in solid-phase peptide synthesis using 2,4,6-trinitrobenzenesulphonic acid. *Anal. Biochem.* **1976**, 71, 260–264.
 32. Hwang, T. L.; Shaka, A. J. Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *J. Magn. Reson., Ser. A* **1995**, 112, 275–279.
 33. Güntert, P.; Mumenthaler, C.; Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **1997**, 273, 283–298.
 34. Prime, version 4.0; Schrödinger, LLC: New York, 2015.

35. Maestro, version 10.1; Schrödinger, LLC: New York, 2009.
36. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
37. Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossvary, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W. L.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *AMBER 14 Reference Manual*; University of California: San Francisco, 2014; pp 29–31.

2.5 Experimental Section

2.5.1 Chemistry

Materials. N^α -Fmoc-protected amino acids, 2-Cl-trtCl resin, Fmoc-Rink amide-Am resin, O-benzotriazole-N,N,N',N'-tetra-methyl-uroniumhexafluorophosphate (HBTU), N,N-diisopropylethyl-amine (DIEA), triisopropylsilane (TIS), trifluoroacetic acid (TFA), and piperidine were purchased from IRIS Biotech (Marktredwitz, Germany), N-hydroxybenzotriazole (HOBt), N,N-dimethylformamide (DMF), dichloromethane (DCM), and from Sigma-Aldrich (Milano, Italy). Peptide synthesis solvents, reagents, H_2O , and CH_3CN for HPLC were reagent grade and were acquired from commercial sources (Sigma-Aldrich, Milano, Italy) and used as received unless otherwise noted. Peptides were purified by preparative HPLC (Shimadzu HPLC system) equipped with a C18-bounded preparative RP-HPLC column (Phenomenex Kinetex 21.2 mm \times 150 mm 5 μ m). Peptides were analyzed by analytical HPLC (Shimadzu Prominance HPLC system) equipped with a C18-bounded analytical RP-HPLC column (Phenomenex Kinetex, 4.6 mm \times 250 mm 5 μ m) using a gradient elution (10–90% acetonitrile in water (0.1% TFA) over 15 min; flow rate = 1.0 mL/min; diode array UV detector). Molecular weights of compounds were confirmed by ESI-mass spectrometry using an Agilent 6110 quadrupole LC/MS system.

General Procedure for COOH-Terminal Peptides. 2-Cl-trtCl resin (62.0 mg, 1.60 mmol/g) was swollen in DCM over 0.5 h, and a solution of Fmoc-L-Cys(trt)-OH (87.9 mg, 0.15 mmol, 1.5 equiv) and DIPEA (26 μ L, 1.5 equiv) in DCM (2 mL) was added. The mixture was stirred for 10 min. An additional amount of DIPEA (34.8 μ L, 0.2 mmol, 2 equiv) was added, and the mixture was then shaken for 1 h. The residual chloride groups contained in the resin were capped by adding MeOH (200 μ L) in DCM (2 mL) and stirring for 15 min to avoid eventually parallel synthesis of side products. Fmoc group removal was performed using 20% piperidine in DMF (1 \times 5 min and 1 \times 25 min). The peptide resin was then washed with DCM (3 \times 0.5 min) and DMF (3 \times 0.5 min), and positive Kaiser ninhydrin³⁵ and TNBS³⁶ tests were observed. Fmoc-L-Phe-OH (155.0 mg, 0.4 mmol, 4 equiv) or Fmoc-D-Phe-OH (155.0 mg, 0.4 mmol, 4 equiv) or Fmoc-L-Ala-OH (124.5 mg, 0.4 mmol, 4 equiv), Fmoc-L-Arg(Pbf)-OH (259.5 mg, 0.4 mmol, 4 equiv) or Fmoc-D-Arg(Pbf)-OH (259.5 mg, 0.4 mmol, 4 equiv) or Fmoc-L-Ala-OH (124.5 mg, 0.4 mmol, 4 equiv), Fmoc-L-Cys(Trt)-OH (234.3 mg, 0.4 mmol, 4 equiv) or Fmoc-D-Cys(Trt)-OH (234.3 mg, 0.4 mmol, 4 equiv), Fmoc-L-Ala-OH (124.5 mg, 0.4 mmol, 4 equiv) or Fmoc-D-Ala-OH (124.5 mg, 0.4 mmol, 4 equiv) were sequentially added sequentially to the resin bound H-L-Cys(Trt). Each coupling reaction was achieved using a 4-fold excess of amino acid with HBTU (151.7 mg, 0.4 mmol, 4

equiv) and HOBt (61.2 mg, 0.4 mmol, 4 equiv) in the presence of DIPEA (140 μ L, 0.8 mmol, 8 equiv) in DMF. Fmoc deprotections were accomplished with 20% piperidine in DMF solution (1 \times 5 min, 1 \times 25 min). Washings with DMF (3 \times 0.5 min) and DCM (3 \times 0.5 min) were performed through every coupling/deprotection step. Kaiser ninhydrin and TNBS tests were employed for monitoring the progress of peptide synthesis. For the peptides bearing an acetyl group at the N-terminus, after removing the last Fmoc group, the resin bound peptide was treated with Ac₂O (19 μ L, 0.2 mmol, 2 equiv) and DIPEA (35 μ L, 0.2 mmol, 2 equiv) in DCM (2 mL) and the mixture was shaken for 1 h. A negative Kaiser ninhydrin and TNBS tests were observed.

General Procedure for CONH₂-Terminal Peptides. Rink amide resin (208 mg, 0.48 mmol/g) was swollen in DMF over 0.5 h, and the Fmoc protecting group of the linker was removed with 20% piperidine in DMF solution (1 \times 5 min and 1 \times 25 min). The resin was washed with DMF (3 \times 0.5 min) and DCM (3 \times 0.5 min) and positive Kaiser ninhydrin and TNBS tests were observed. The linear sequences were assembled sequentially by adding Fmoc-L-Cys(trt)-OH (234.3 mg, 0.4 mmol, 4 equiv), Fmoc-L-Phe-OH (154.98 mg, 0.4 mmol, 4 equiv), Fmoc-L-Arg(Pbf)-OH (259.51 mg, 0.4 mmol, 4 equiv), Fmoc-L-Cys(trt)-OH (234.3 mg, 0.4 mmol, 4 equiv), and Fmoc-L-Ala-OH (124.54 mg, 0.4 mmol, 4 equiv).

Each coupling reaction was achieved using a 4-fold excess of amino acid in the presence of HBTU (151.7 mg, 0.4 mmol, 4 equiv), HOBt (61.2 mg, 0.4 mmol, 4 equiv), and DIPEA (140 μ L, 0.8 mmol, 8 equiv) in DMF (3 mL). Fmoc deprotections were accomplished with 20% piperidine solution in DMF (1 \times 5 min, 1 \times 25 min). Washings with DMF (3 \times 0.5 min) and DCM (3 \times 0.5 min) were performed through every coupling/deprotection step. Kaiser ninhydrin and TNBS tests were employed for monitoring the progress of peptide synthesis. For the peptides bearing an acetyl group at the N-terminus, after removing the last Fmoc group, the resin bound peptide was treated with Ac₂O (19 μ L, 0.2 mmol, 2 equiv) and DIPEA (35 μ L, 0.2 mmol, 2 equiv) in DCM (2 mL) and the mixture was shaken for 1 h. Negative Kaiser ninhydrin and TNBS tests were observed.

General Procedure for Peptide Oxidation and Purification. The peptide was released from the solid support and all the protecting groups cleaved, treating the resin with TFA/DCM/TIS (80/15/5, v/v/v) (3 mL solvent/0.1 mmol) for 2 h. The resin was then filtered off, and the crude linear peptide was recovered by precipitation with chilled ether to give a powder. The crude peptide (0.1 mmol) was dissolved in 30 mL of glacial AcOH, and 200 μ L of HCl 1N were added to improve the solubility of the peptide. Next, a solution of iodine (40 mg, mmol) in glacial AcOH (5 mL) was

added and the mixture was mechanical stirring for 12 h at room temperature. The resulting mixture was concentrated in vacuo by rotatory evaporator (1/4 of the original volume) and transferred to a 50 mL conical centrifuge tube, where the oxidated crude peptide was precipitated by adding 40 mL of chilled ether, recovered by filtration, and dried overnight. Final peptide purification was achieved by preparative RP- HPLC in 0.1% TFA with an ACN gradient (10–90% ACN in H₂O over 15 min, flow rate of 15 mL/min) on a Phenomenex Kinetex C18 column (21.2 mm × 150 mm 5 μm). Analytical RP-HPLC were performed in 0.1% TFA with an ACN gradient (10–90% ACN in H₂O over 20 min, flow rate of 1.0 mL/min) on a Phenomenex Kinetex C18 column (0.46 mm × 150 mm 5 μm).

2: Purity >95%, tR 12.2 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₃₉H₅₇N₁₃O₈S₂; calculated mass, 899.4; found, 900.3 (M + H⁺), 450.7 (M + 2H⁺)/2.

3: Purity >95%, tR 12.5 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₄₁H₅₉N₁₃O₉S₂; calculated mass, 941.4; found, 942.6 (M + H⁺), 471.4 (M + 2H⁺)/2.

4: Purity >95%, tR 12.0 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₃₉H₅₇N₁₃O₈S₂; calculated mass, 898.4; found, 899.6 (M + H⁺), 450.0 (M + 2H⁺)/2.

5: Purity >95%, tR 12.2 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₄₁H₆₀N₁₄O₈S₂; calculated mass, 940.4; found, 941.6 (M + H⁺), 471.0 (M + 2H⁺)/2.

6: Purity >95%, tR 12.6 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₄₁H₅₉N₁₃O₉S₂; calculated mass, 941.4; found, 942.5 (M + H⁺), 471.8 (M + 2H⁺)/2.

7: Purity >95%, tR 12.6 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₄₁H₅₉N₁₃O₉S₂; calculated mass, 941.4; found, 942.4 (M + H⁺), 471.8 (M + 2H⁺)/2, 964.5 (M + Na⁺).

8: Purity >95%, tR 12.4 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₄₁H₅₉N₁₃O₉S₂; calculated mass, 941.4; found,

942.4 (M + H⁺), 471.8 (M + 2H⁺)/2.

9: Purity >95%, tR 12.4 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₄₁H₅₉N₁₃O₉S₂; calculated mass, 941.4; found, 942.4 (M + H⁺), 471.8 (M + 2H⁺)/2.

10: Purity >95%, tR 12.6 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₄₁H₅₉N₁₃O₉S₂; calculated mass, 941.4; found, 942.4 (M + H⁺), 471.8 (M + 2H⁺)/2, 964.4 (M + Na⁺).

11: Purity >95%, tR 12.5 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₄₁H₅₉N₁₃O₉S₂; calculated mass, 941.4; found, 942.4 (M + H⁺), 471.8 (M + 2H⁺)/2.

12: Purity >95%, tR 12.5 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₄₁H₅₉N₁₃O₉S₂; calculated mass, 941.4; found, 942.6 (M + H⁺), 471.8 (M + 2H⁺)/2, 964.5 (M + Na⁺).

13: Purity >95%, tR 15.1 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₃₈H₅₂N₁₀O₉S₂; calculated mass, 856.3; found, 857.4 (M + H⁺).

14: Purity >95%, tR 15.1 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₃₈H₅₂N₁₀O₉S₂; calculated mass, 856.3; found, 857.4 (M + H⁺).

15: Purity >95%, tR 13.6 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₃₅H₅₅N₁₃O₉S₂; calculated mass, 865.4; found, 866.4 (M + H⁺), 433.7 (M + 2H⁺).

16: Purity >95%, tR 13.8 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula, C₃₅H₅₅N₁₃O₉S₂; calculated mass, 865.4; found, 866.4 (M + H⁺), 433.7 (M + 2H⁺).

Human Plasma Stability Assay. Human plasma, H₂O, ACN, and TFA were obtained from Sigma-Aldrich (Milano, Italy) and used without further purification. Human plasma was diluted to

50% in pure water. Analytical HPLC-ESI-MS was performed on an Agilent Technologies 1200 series equipped with an Agilent Technologies 6110 quadrupole LC/MS using a Phenomenex Luna C18 column (5 μ m, 4.6 mm \times 150 mm) and H₂O (0.1% v/v TFA)/ACN (0.1% v/v TFA) as eluents.

A solution of 1 mg/mL of each peptide was prepared in water, and 150 μ L aliquots were mixed with 150 μ L of prewarmed (37 $^{\circ}$ C) plasma. At selected time points (0, 30, 60, 120, and 180 min), samples (50 μ L) were collected and mixed with 1% TFA in acetonitrile (75 μ L) to precipitate plasma proteins which were deleted by centrifugation at 13000 rpm for 10 min. The supernatant was analyzed by HPLC-ESI-MS using a Phenomenex Luna C18 column (5 μ m, 4.6 mm \times 150 mm) and an elution gradient of 10–90% solvent B over 20 min (solvent A, 0.1% TFA in water; solvent B, 0.1% TFA ACN) at a flow rate of 1 mL/min.

2.5.2 Biological Studies

Binding Assay. CXCR4 binding was evaluated as previously described.²² Briefly, 5×10^5 CCRF–CEM, HT29 cells were preincubated with increasing peptide concentrations (0.01 μ M, 0.1 μ M, 1 μ M, 10 μ M) in the binding buffer (PBS 1 \times plus 0.2% BSA and 0.1% NaN₃) for 30 min at 37 $^{\circ}$ C, 5% CO₂ and then labeled for 30 min using anti-CXCR4 PE-antibody (FAB170P, clone 12G5, R&D Systems, Minneapolis, MN, USA). To evaluate the specific peptide binding to CXCR4, the experiments were also conducted in COLO205, human colon cancer cells, overexpressing CXCR3, using anti-CXCR3 FITC-antibody (R&D FAB160F clone 49801) and in MCF-7, human breast cancer cell line, overexpressing CXCR7, using anti-CXCR7 APC-antibody (R&D FAB4227A clone 11G8). The cells were counted through a FACS Canto II cytofluorometer (Becton Dickinson Immunocytometry Systems, Mountain View, CA, USA).

Migration Assay. Migration was assayed in 24-well Transwell chambers (Corning Inc., Corning, NY) using inserts with 5 and 8 μ m (optimal for lymphocytes and epithelial cells, respectively) pore membranes. Membranes were precoated with collagen (human collagen type I/III) and fibronectin (20 μ g/mL each). Cells were placed in the upper chamber (2×10^5 cells/well) in culture medium containing 1% BSA (migration media) in the presence of the peptide; 100 ng/mL CXCL12 was added to the lower chamber. After 18 h of incubation, cells atop the filter were removed using a cotton wool swab; migration of cells in the lone medium (control) was compared with that observed in media containing CXCL12. The cells were counted in 10 different fields (original 400 \times magnification). The migration index was defined as the ratio between migrating cells in the experimental group and migrated cells in the control group.

Immunoblotting. Cells were homogenized in lysis buffer (40 mM Hepes pH 7.5, 120 mM NaCl, 5 mM MgCl₂, 1 mM EGTA, 0.5 mM EDTA, 1% Triton X-100) containing protease (Complete Tablets- EDTA free, Roche) and phosphatase inhibitors (20 mM α -glycerol-3- phosphate, 2.5 mM Na-pyrophosphate). The following primary antibodies were used: anti-p-ERK (sc7383, Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA) and anti-ERK2 (sc 154G, Santa Cruz Biotechnology CA, USA). P-ERK induction was plotted as the ratio between p-ERK in the presence of CXCL12 (100 ng/mL) and p-ERK in free serum.

Receptor Internalization Assay. HCT116-GFP-CXCR4 cell line was established in our lab by transfecting pEGFP-CXCR4 plasmid (Origene, USA) into the parental HCT116 cell line; GFP-CXCR4 expression in selected HCT116 clones was then confirmed by flow cytometry and immunofluorescence using anti-CXCR4 12G5 antibody (not shown). Subconfluent HCT116-GFP-CXCR4 cells were grown on 10 mm coverslips in DMEM plus 10% FBS-1% P/S-1% glutamine for 1 day, washed with PBS plus 0.5% BSA, and equilibrated in DMEM plus 0.5% BSA (Sigma-Aldrich). Cells were then exposed to 2 (0.1 μ M) or 10 (0.1 μ M) for 45 min at 37 °C. Cells were then treated with 50 mg/mL CXCL12 and incubated at 37 °C for 45 min, fixed with 4% formaldehyde for 10 min, washed three times in PBS, permeabilized with 0.1% Triton X-PBS for 10 min, and washed three times again in PBS. Cells were finally stained with α - tubulin/Alexa Fluor 594 and DAPI and inspected with a Zeiss LSM 510 confocal microscope equipped with Zeiss 2014 software.

NMR Spectroscopy. The samples in micelle solution (200 mM of SDS-d₂₅) for NMR spectroscopy were prepared by dissolving the peptides in 0.18 mL of 1H₂O (pH 5.5) and 0.02 mL of 2H₂O to obtain a concentration 2 mM of peptides. 2D DQF-COSY,²⁷ TOCSY,²⁸ and NOESY²⁹ spectra were recorded on a Varian INOVA 700 MHz spectrometer equipped with a z-gradient 5 mm triple-resonance probe head at a temperature of 298 K. The water signal was suppressed by gradient echo.³⁷ A mixing time of 80 and 200 ms was used for the TOCSY and NOESY experiments, respectively. The interactive program package XEASY³⁰ was used for the NMR analysis of DQF-COSY, TOCSY, and NOESY spectra. ³JHN-H α coupling constants were acquired from 1D 1H NMR and 2D DQF-COSY spectra. The temperature coefficients of the amide proton chemical shifts were calculated from 1D 1H NMR and 2D TOCSY experiments executed at different temperatures in the range 298–313 K by means of linear regression.

Structural Determinations. The NOE cross peaks were integrated with the XEASY program and were converted into upper distance bounds using the CALIBA software provided by the

program package DYANA.³⁸ Only NOE derived constraints (Supporting Information, Tables S3 and S4) were considered in the annealing procedures. An ensemble of 100 structures was generated with the simulated annealing calculations followed by successive steps of restrained and unrestrained energy minimization using the Discover algorithm (Accelrys, San Diego, CA). From the produced 100 conformations, 10 structures were chosen, whose interproton distances best fitted NOE derived distances.

Molecular Modeling. A manual docking of the 10 averaged NMR structure in the crystal structure of CXCR4 in complex with the 16-mer cyclic peptide antagonist CVX15 (PDB code: 3OE0) was first accomplished. Specifically, we superimposed the coordinates of the C α carbons of Arg4, Phe5, and Phe6 in 10 with the C α carbons of Arg1, Arg2, and NaI3 in CVX15, which was then removed from the complex. Subsequently, to limit steric clash within the CXCR4 binding cavity, the side chains of Arg1 and Arg4 in 10 were manually extended to interact with Asp181 and Asp97, respectively. Prior to MD simulations, the receptor first N-terminal (K25) and the last C-terminal (A301) residues were capped with ACE and NME, respectively, while missing intracellular loops (ICL1 and ICL3) were added and refined using Prime.³⁹ Histidine tautomeric/protonation states were assigned through the PROPKA module of the Maestro Protein Preparation Wizard.⁴⁰ The refined complex was embedded in a POPC (1-palmitoyl-2-oleoylphosphatidylcholine) phospholipids bilayer to mimic the physiological environment and then submitted to an over 100 ns MD simulation with NAMD 2.9.⁴¹ In detail, a 94 Å × 94 Å (in x and y axes) pre-equilibrated POPC phospholipid bilayer was first created with the aid of the membrane-builder tool of CHARMM-GUI.org (<http://www.charmm-gui.org>). With the purpose of placing the receptor into the bilayer, a hole was generated and all lipids in close contact (<1 Å distance from any protein atoms) were deleted. The complex was then solvated using the solvation module of VMD 1.9.2, using the TIP3 water model. The ff14SB and lipid14 Amber force fields⁴² were used to parametrize the protein and the peptide and the lipids, respectively. The addition of six Cl⁻ ions ensured neutrality. A 10 Å cutoff (switched at 8 Å) was used for atom pair interactions. The long-range electrostatic interactions were computed by means of the particle mesh Ewald (PME) method using a 1.0 Å grid spacing in periodic boundary conditions. The RATTLE algorithm was applied to constrain bonds involving hydrogen atoms; thus, a 2 fs integration time step interval could be used. The system was minimized and heated up to 300 K applying harmonic constraints, which were gradually released along the equilibration process. To prevent any distortion in the receptor secondary structural elements, these were constrained for further 10 ns. Production run was then performed in the NPT ensemble at 1 atm and 300 K.

The same protocol was followed for the 2/CXCR4 complex.

All of the pictures were rendered using PyMOL (www.pymol.org).

Chapter 3

Structure-Activity Relationship Study Led to the Identification of a New Potent and Plasma Stable CXCR4 Antagonist Mimicking the N-terminal Region of CXCL12

3.1 Introduction

Accumulating evidence indicates that human herpesvirus 8 (HHV-8) is the infectious agent responsible for Kaposi's sarcoma, both in patients with and without HIV infection, as well as it has been implicated in the pathogenesis of primary effusion lymphoma (PEL, a body cavity-based lymphoma), multicentric Castleman's disease, and recently multiple myeloma (MM).¹⁻⁵

HHV-8 encodes a variety of immunomodulatory proteins that are pirated from cellular genes by the virus: through this evolved sophisticated strategy of molecular mimicry HHV8 turns the host receptor network to its own advantage. DNA sequence analyses of HHV-8 suggest that three open reading frames encode for proteins with considerable identity to human CC chemokines (vCCL1/vMIP-I, vMIP-II/vMIP-II, and vCCL3/vMIP-III) and one for a G protein-coupled receptor (KSHV-GPCR, ORF74) that is constitutively active.^{6,7}

vMIP-I and vMIP-III are shown to be specific agonists of host CCR8 and CCR4,⁸ respectively, while vMIP-II, despite being a CC chemokine like protein, binds to a broad spectrum of both viral and human chemokine receptors across the 4 families (CCR, CXCR, XCR, and CX3CR) and, depending on the receptor, acts either as an antagonist or an agonist.

vMIP-II exerts its antagonist effect on CCR1, -2, -5,⁹⁻¹⁰ CXCR4¹¹ and CX3CR1¹² inhibiting endogenous ligand-induced calcium responses and cell migration (Figure 3.1a). Owing to the large spectrum of its partner receptors, vMIP-II is also able to block the recruitment of immune cells at different stages of their activation. Numerous studies have reported, for instance, that, despite the differences in the patterns of chemokine receptors expressed, the migration of both native and activated NK cells can be inhibited by the interaction of vMIP-II with CX3CR1 and CCR5, respectively.¹⁴ Moreover, vMIP-II may participate in controlling the 2 gene expression programs of HHV-8. vMIP-II acts as an inverse agonist toward the receptor ORF74 the HHV-8-encoded GPCR.

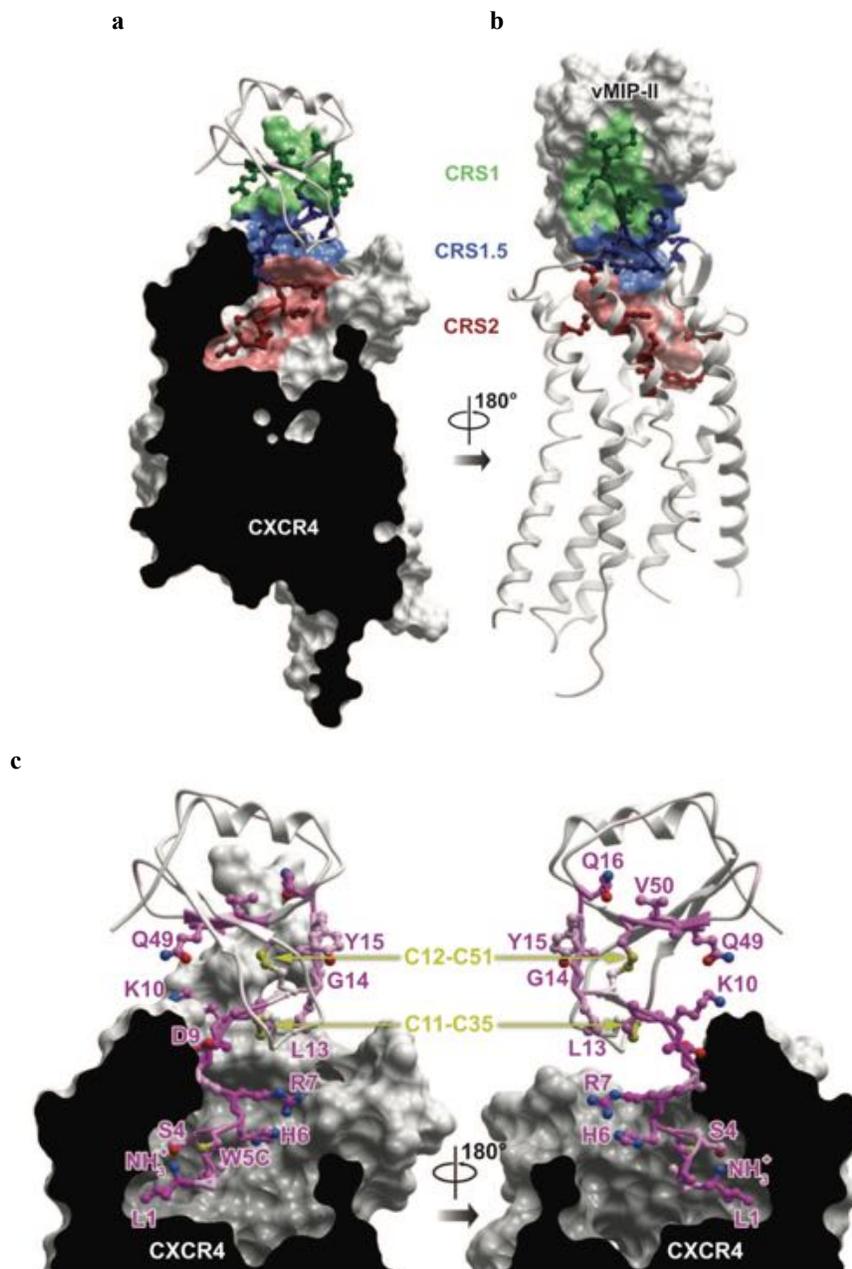


Figure 3.2. Interactions between CXCR4 and vMIP-II. a) and b) The interaction is mediated by a contiguous interface iproviding CRS1, CRS2 and CRS1.5. The vMIP-II residues making substantial contacts with the receptor are shown as sticks. c) Key residues from vMIP-II for the binding; noncarbon atoms are red (O), blue (N), and yellow (S); carbon stick color intensity is indicative of residue contact strength. (Modified from Qin et al., 2015)

ORF74, a viral homolog of the human CXCR2, is expressed during the early lytic stage of the virus conferring a highly proliferative potential to the cells. vMIP-II down-regulating the constitutive active receptor temporarily represses HHV-8 reactivation and, thereby helps helping to escape the host immune surveillance.^{15,16}

vMIP-II is produced as a 94-aa precursor with a 23-aa N-terminal signal peptide and a C-terminal arginine, which are cleaved to yield the mature 70-aa chemokine (7.9 kDa).^{17,18} The tridimensional structure of vMIP-II has been resolved by both x-ray¹⁹⁻²⁰ crystallography and NMR,²¹⁻²³ and it has been shown to adopt a folded architecture typical of human chemokines characterized by a flexible and disordered N terminus of 10 residues followed by a “core domain” consisting of an N-loop, three anti-parallel β -strands, and a C-terminal α -helix. Two disulfide bridges linking the cysteine residues 11–35 and 12–51 connect the N terminus and the β -sheet of vMIP-II (Figure 1b). However, in contrast to its closely related human CC chemokine homologs, in solution, vMIP-II appears to have no tendency to dimerize,^{21, 24,25} although an engineered dimeric variant (L13F) has been reported.²⁶

The recently resolved crystal structure of vMIP-II in complex with CXCR4⁸ revealed a 1:1 interaction stoichiometry with extensive contiguous interface between the two partners (Figure 3.2). Moreover, it has provided the molecular bases for the specificity of CC and CXC chemokines toward their respective receptors and for the cross-family interaction of the vMIP-II–CXCR4 pair. In particular, vMIP-II interacts via its globular core (N-loop, residues 13-LGYQ-16; β 3-strand, residues 49-QVC-51) with CXCR4 N-terminal residues 23-SMKEP-27, forming the common chemokine recognition site 1, CRS1. Moreover, in CSR2 the N-terminus of vMIP-II makes hydrogen bonds to receptor TM pocket residues D97, D262 and E288 and numerous van der Waals interactions. Therefore, vMIP-II bears CXC chemokine-like features that are involved in the interaction with CXC receptor-conserved residues of CXCR4. In addition, the receptor residues 27-PCFRE-31 are packed against the chemokine residues 6-HRPDKCC-12 in a peculiar CSR1.5 interaction. Thus, the three basic residues (H6, R7, K10) in the vMIP-II region proximal to the N-terminus and the CXCR4 basic aminoacid R30 may explain the unusual coupling between the viral chemokine and the receptor. Indeed, interestingly both vMIP-II and CXCR4 possess features that are atypical for their respective classes.

The N-terminus of vMIP-II shows little homology with either CC or CXC chemokines.²⁷ In accordance with the accepted chemokine-receptor 2 step binding mode, the chemokine N-terminus is crucial for biological function.²⁸⁻³¹ Thus, it is reasonable that the singular N-terminal sequence of

vMIP- II may confer unique activity distinct from the other chemokine ligands. To support the hypothesis of the dominant role of the vMIP-II N-terminus, a derived peptide containing the amino acid sequence residues 1-21 (LGASWHRPDKCCLGYQKRPLP) has been synthesized.³²

This peptide displayed antagonistic activity against CXCR4 blocking its signal transduction and coreceptor function in mediating HIV-1 entry, competing with the CXCR4 binding of ¹²⁵I-SDF- 1 α with an IC₅₀ of 190 nM (14.8 nM is the IC₅₀ of native vMIP-II in the same assay), but it did not recognize CCR5. Notably, the truncated peptide analogous containing the residues 6-18 showed a significant loss of affinity whereas the activity is preserved maintaining the first half of the sequence (1-10). Also, CXCR4 revealed unsuspected high permissivity to stereoisomer replacement at the receptor-peptide interface: the all-D amino acid analog of the first 21 residues of vMIP-II is still able to modulate inflammation and inhibit the replication of CXCR4-dependent HIV-1 strains. This allows the design of peptides with higher resistance to proteolytic degradation.^{27, 33-36}

3.2 Results and discussion

3.2.1 Design

The work presented here covers the results we obtained during a leading optimization study of our previous reported plasma stable CXCL12-derived CXCR4 peptide antagonist **10** (Ac-Arg-Ala-[DCys-Arg-Phe-Phe-Cys]-COOH), which was able to selectively bind the receptor with an IC₅₀ = 0.053 μ M (Table 1). In the early stage, the overall conformation, dynamic and topographical properties of the cyclic region of **1** were investigated. Specifically, the *L*-Cys (position 7) and the *D*-Cys (position 3) residues involved in the disulphide bridge were replaced with their geminal dimethyl analogous, *L*-Pen and *D*-Pen respectively (**17-19**). From this first round of peptide synthesis **17** emerged as the most potent analogue and was submitted to a second cycle of chemical modifications.

Our previous Ala-scan study performed on peptide **10** strongly indicated that the aromatic *L*-aa at residues 5 and 6 of were fundamental to interact with the receptor. Thus, in the attempt to improve the affinity of our lead peptide towards CXCR4 the phenylalanine aromatic moiety was revised. The effect of electron-donating (4-OH) (**20, 25**) and electron-withdrawing (4-NO₂ and 4-Cl) (**22, 27** and **21, 26**) groups on the phenyl rings of the core of the eptapeptide sequence were examined, as well

as the the addition of steric bulk by the introduction of non-natural amino acids (i.e. 1-Nal, 2-Nal) (**23-24**, **28-29**). The analogue (**29**) resulting from the replacement of *L*-Cys and *L*-Phe at the position 7 and 5, with a residue of *L*-Pen and *L*-2-Nal, respectively, was finally modified by introducing a residue of histidine instead of phenylalanine at the 6 position (**30**). The latter was inspired by the observation that the natural ligand CXCL12 and the viral protein vMIP-II share the common interacting tripeptide motif X-Y-Arg, where X and Y are both Phenylalanine in CXCL12, while X=tryptophan and Y=histidine in vMIP-II.^{9,28,37,39-41} Thus, considering that *L*-2-Nal is widely employed as *L*-Trp mimetic, we introduced a residue of L-His residue at the position 6 in order to recapitulate the tripeptide-binding moiety of vMIP-II.

Table 3.1. Half-Maximal Inhibitory Concentration (IC₅₀, μM, Mean ± SD) of CXCR4 Antagonist Peptides Necessary to Reduce by 50% The Binding to CXCR4 of CXCR4-Specific mAb 12G5 in CCRF-CEM Cells^a

compd	sequence	IC ₅₀ (μM)
10	Ac-Arg-Ala-[DCys-Arg-Phe-Phe-Cys]-COOH	0.053 ± 0.004
17	Ac-Arg-Ala-[DCys-Arg-Phe-Phe-Pen]-COOH	0.040 ± 0.01
18	Ac-Arg-Ala-[DPen-Arg-Phe-Phe-Cys]-COOH	>10
19	Ac-Arg-Ala-[DPen-Arg-Phe-Phe-Pen]-COOH	1.530 ± 0.040
20	Ac-Arg-Ala-[DCys-Arg-Phe-Tyr-Pen]-COOH	0.36 ± 0.060
21	Ac-Arg-Ala-[DCys-Arg-Phe-Phe(4-Cl)-Pen]-COOH	1.250 ± 0.063
22	Ac-Arg-Ala-[DCys-Arg-Phe-Phe(4-NO ₂)-Pen]-COOH	> 10
23	Ac-Arg-Ala-[DCys-Arg-Phe-1Nal-Pen]-COOH	0.18 ± 0.040
24	Ac-Arg-Ala-[DCys-Arg-Phe-2Nal-Pen]-COOH	1.8 ± 0.0560
25	Ac-Arg-Ala-[DCys-Arg-Tyr-Phe-Pen]-COOH	2.14 ± 0.030
26	Ac-Arg-Ala-[DCys-Arg-Phe(4-Cl)-Phe-Pen]-COOH	0.070 ± 0.020
27	Ac-Arg-Ala-[DCys-Arg-Phe(4-NO ₂)-Phe-Pen]-COOH	0.21 ± 0.070
28	Ac-Arg-Ala-[DCys-Arg-1Nal-Phe-Pen]-COOH	3.69 ± 0.081
29	Ac-Arg-Ala-[DCys-Arg-2Nal-Phe-Pen]-COOH	0.022 ± 0.002
30	Ac-Arg-Ala-[DCys-Arg-2Nal-His-Pen]-COOH	0.0015 ± 0.0005

^aEach experiment was performed in triplicate.

3.2.2 Chemistry

All the linear eptapeptides were assembled by stepwise orthogonal Fmoc/tBu-SPPS. Being cysteine derivatives are especially prone to racemization, we used 2-Cl chlorotrityl resin as support. Unfortunately, we observed that the classical reported protocol for loading the first amino led to lower yield when the *L*-Cys was replaced with a residue of *L*-Pen (15 % *L*-Pen Vs 60 % *L*-Cys). In our opinion, the observed lack of efficiency was probably due to the more hindered side chain of penicillamine compared with cysteine.

Thus, in order to decrease the loading in this sterically demanding circumstance, the coupling of the C-terminal residue was successfully performed using an excess of resin (2 equivalent) compared to the amino acid. Indeed, the alternative approach increased the yield of the ultimate linear compounds from 15 % to ca 55 %. Moreover, the peptides were N-terminally acetylated with acetic anhydride. Consequently, the assembled resin-bounded peptides were removed from the solid support, leading to the concomitant side chain deprotection, and the backbone cyclization was carried out using N-Chloro succinimide to form the desired disulphide bridge.

3.2.3 CXCR4 Binding Assay

CEM-CCRF human T leukemia CXCR4 expressing cells were incubated with PE-conjugated-12G5 anti-CXCR4 in the presence of the various peptides (Table 3.1). As a result of the replacement of the *L*-Cys⁷ with a topologically obligated residue, namely *L*-Pen, **17** showed a sensible enhancement in terms of affinity compared to **10**, exhibiting an IC₅₀ of 0.040 μM. Diametrically opposed results were achieved when *D*-Pen was introduced instead of the *D*-Cys residue, leading to the peptide **18**, and when both *D*- and *L*-Cys were replaced with the respective Pen analogues, yielding the analog **19**. Either way a dramatic lack of affinity was displayed. Taken together, these data suggest that keeping constant the size of the ring (**10** Vs **17**, Table 3.1), little modification such as the presence of the two methyl groups on the β-carbon of Pen induces substantial modification on the orientation of the pharmacophoric amino acid side chains, those are favorable for the binding only in the case of *L*-Cys substitution. Indeed, the position 3 is critical for the binding and does not tolerate side chain modifications. Further, a structure activity relationship approach to probe the influence of the functionalization on the two Phe (Phe⁵ and Phe⁶) residues of the template peptide **17** was performed. We observed that most of the compounds did not display significant competitive activity towards the CXCR4 receptor with respect to **17** in the concentrations tested. The Tyr-substituted analogue **20** resulted in ca 9-fold decreased affinity than the **17**, along with the peptide **21**, endowed with a Cl

electron-withdrawing substituent on the aromatic ring, that showed a non-nanomolar affinity to CXCR4 (Table 3.1). Moreover, complete loss of activity was observed when a NO₂ group was introduced at the para position of the Phe⁶ ring. Similar results were observed assaying the substitutions of modified electronic character at position 5: the **25-27** (Table 3.1) mutants (Tyr, Phe(4-Cl) and Phe(4-NO₂) alternatively introduced at position 5) resulted in 2- to >10-fold decreased ligand potency at the chemokine receptor 4. Substitution of the Phe⁶ side chains by the bulkier residues 1-Nal and 2-Nal led to the less potent relative analogues **23** and **24** (**23**_{IC₅₀}=0.18 μM; **24**_{IC₅₀}=1.8 μM, Table 3.1). Conversely, the replacement of the Phe⁵ residue with its more voluminous mimetic moieties 1-Nal and 2-Nal resulted in interesting data. While a marked lack of affinity was observed by introducing 1-Nal (**28**), peptide **29**, resulting from the replacement of the Phe⁵ with 2-Nal, is endowed with a 2-fold higher potency for the receptor compared to the parent peptide **17** (IC₅₀= 0.022 μM for **29** vs IC₅₀=0.04 for **17**). Interestingly, preliminary molecular modeling investigations validate these data showing that the transverse steric bulk of the 1-Nal side chain cannot be accommodated for steric reasons, while the more elongated 2-Nal residue can penetrate deep in the aromatic cage of the binding site.

Finally, we investigated the activity profile of **30** (Table 3.1). This peptide features at 6 position a His residue that was initially introduced to mimic the Arg, Trp, His motif of vMIP-II, as already stated in paragraph 3.2.1. However, our design hypothesis has been contradicted by the recent publication of the CXCR4/vMIP-II crystallographic complex. In fact, a visual inspection of this structure would suggest a minor role for the vMIP-II histidine residue in the binding to the chemokine receptor. Furthermore, the superposition of the CXCR4/vMIP-II complex with the crystal structure of CXCR4 bound to the canonical ligand CXCL12 shows that the vMIP-II histidine amino acid does not overlap with any of the main pharmacophoric side chains of CXCL12. In spite of these evidences, our binding assays revealed a remarkable one-order of magnitude decrease in the IC₅₀ of **30** compared to the parent peptide **17**. This would indicate that His⁶ can establish not only aromatic contacts but also tight electrostatic interactions with specific binding site residues. These aspects are currently being further investigated by extensive molecular dynamics simulations.

3.2.4 CXCR3 and CXCR7 Binding Assays

The discriminative binding of a well-designed drug with its designed receptor is a required critical character to prevent adverse reactions related to off-target activity. Here, we assessed the selectivity profile of our most potent peptide (**30**, Table 3.1) in two other GPCR binding assays,

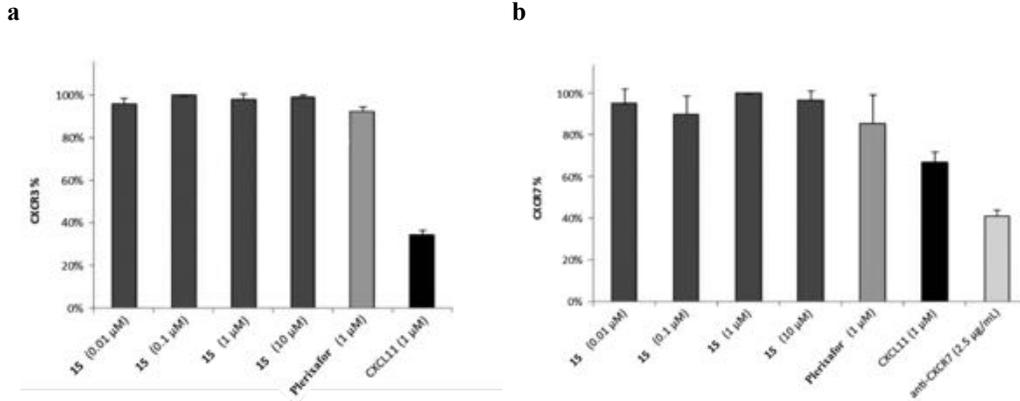


Figure 3.3. Binding assays on CXCR3 and CXCR7: indirect binding experiments with the anti-CXCR7, clone 11G8, and anti-CXCR3 clone 49801 were conducted in MCF-7 and COLO205, CXCR7, and CXCR3 overexpressing cells, respectively. Data are presented as bar graph showing mean \pm SD.

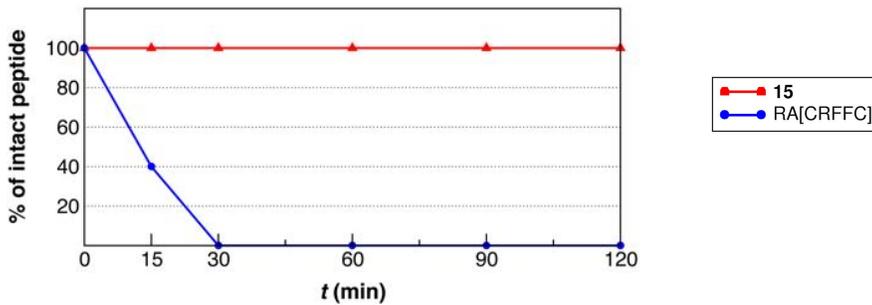


Figure 3.4. Human plasma stability profiles of 15 and RA[CRFFC] after different intervals of incubation with human plasma. Relative concentrations of peptides were determined by integration of the A230 peaks from RP-HPLC.

specifically CXCR3 and CXCR7. In fact, we evaluated the anti-CXCR3 and anti-CXCR7 efficacy of **30** on a colon cancer cell line (COLO205) overexpressing CXCR323 and a breast cancer cell line (MCF-7) overexpressing CXCR7. As shown in Figure 3.3, **30** did not bind either CXCR7 or CXCR3 up to 10 μ M concentration, showing a remarkable degree of functional selectivity along with

functional antagonism. Once again, as previous reported the chemical modifications here introduced have not affected the selectivity profile of the lead.

3.2.5 Plasma Stability Assay

The early metabolic liability was assessed by incubating **30** and the previous reported unstable sequence H-Arg-Ala-[Cys-Arg-Phe-Phe-Cys]-COOH in diluted human plasma at 37 °C. Time course aliquots were collected, treated with acetonitrile to precipitate the plasma proteins, and analyzed by ESI-RP-HPLC (Figure 3.4). After 15 min, the concentration of the hit peptide was reduced resulting in a cleaved form of the peptide lacking the Arg1 residue and the Arg1–Ala2 dipeptide at the N-terminal region. After 90 min, the sequence resulted fully converted into the corresponding cyclic pentapeptide metabolite. Conversely, the N-terminal acetylated **30** was stable up to 120 min. These data clearly highlighted the successfully introduction of the acetyl group at the N-terminal region towards the peptide against human plasma proteolytic degradation.

3.3 Conclusion

The core cyclic region of our lead peptide was examined in two phase structure-activity relationship (SAR) studies entailing the development of peptide analogous with different backbone geometry, including constrained building blocks to conformationally fix the sequence. Next, the side chain functionality was investigated, employing different groups to alter the electronic density of the Phe ring system. Thus, we were able to improve the binding affinity and selectively prevent CXCR4 signal transduction with our peptide **29**, which showed a 2-fold stronger affinity for the chemokine receptor when compared with the template peptide. Despite these preliminary good results, we did not restrict ourselves and we went further with the derivatization. At the time when we were involved in the peptide design step, very sparse structural information was available and structural basis for receptor recognition by its natural ligands (i.e. CXCL12 and VMIP-II) were not well understood. An interesting study suggested the overlapping conformational space of the side chains of endogenous agonist CXCL12 of the CXCR4 receptor and its viral antagonist vMIP-II. The three residue motif Trp-His-Arg (Ar1-Ar2-R), identified in the vMIP-II N-terminus was found to be present in the opposite orientation as R-Ar1-Ar2 (Arg-Phe-Phe) in the N-loop of CXCL12. Hence, such alignment of backbone atoms of residues supplied persuasive evidence that these side-chains

could form a CXCR4 receptor-binding motif. Accordingly, we decided to replace the Phe⁶ with a His residue (**30**) in order to recapitulate the tripeptide-binding moiety of vMIP-II. Later, the disclosure of the CXCR4/vMIP-II crystallographic complex helped to elucidate and rationalize the large body of cumulated data displaying, however, a minor role for the supposed binding-implicated vMIP-II His residue. Nonetheless, the introduction of the His moiety resulted in a remarkable effect on CXCR4 receptor binding affinity rendering this promising, and in human plasma stable, compound a good candidate for subsequent PET imaging studies.

3.4 References

1. Cesarman, E.; Moore, P. S.; Rao, P. H.; Inghirami, G.; Knowles, D. M.; Chang, Y. In vitro establishment and characterization of two acquired immunodeficiency syndrome-related lymphoma cell lines (BC-1 and BC-2) containing Kaposi's sarcoma-associated herpesvirus-like (KSHV) DNA sequences. *Blood* **2005**, *86*, 2708–2714.
2. Chadburn, A.; Cesarman, E.; Jagirdar, J.; Subar, M.; Mir, R. N.; Knowles, D. M. CD30 (Ki-1) positive anaplastic large cell lymphomas in individuals infected with the human immunodeficiency virus. *Cancer* **1993**, *72*: 3078–3090.
3. Gao, S. J., L. Kingsley, M. Li, W. Zheng, C. Parravicini, J. Ziegler, R. Newton, C. R. Rinaldo, A. Saah, J. Phair, R. Detels, Y. Chang, and P. S. Moore. KSHV antibodies among Americans, Italians and Ugandans with and without Kaposi's sarcoma. *Nat. Med.* **1996**, *2*, 925–928.
4. Miller, G.; Rigsby, M. O.; Heston, L.; Grogan, E.; Sun, R.; Metroka, C.; Levy, J. A.; Gao, S. J.; Chang, Y.; Moore, P. S. Antibodies to butyrate- inducible antigens of Kaposi's sarcoma-associated herpesvirus in patients with HIV-1 infection. *N. Engl. J. Med.* **1996**, *334*, 1292–1297.
5. Moore, P. S.; Gao, S. J.; Dominguez, G.; Cesarman, E.; Lungu, O.; Knowles, D. M.; Garber, R.; Pellett, P. E.; McGeoch, D. J.; Chang, Y. Primary characterization of a herpesvirus agent associated with Kaposi's sarcoma. *J. Virol.* **1996**, *70*, 549–558.
6. Boshoff, C.; Schulz, T.F.; Kennedy, M.M.; Graham, A.K.; Fisher, C.; Thomas, A.; McGee, J.O.; Weiss, R.A.; O'Leary, J.J. Kaposi's sarcoma- associated herpesvirus infects endothelial and spindle cells. *Nat Med* **1995**, *1*, 1274, 1995.
7. Arvanitakis, L.; GerasRaaka, E.; Varma, A.; Gershengorn, M.C.; Cesarman, E. Human herpesvirus KSHV encodes a constitutively active G-protein-coupled receptor linked to cell proliferation. *Nature* **1997**, *385*, 347.
8. Nicholas, J. Human herpesvirus 8-encoded cytokines. *Future Virology* **2010**, *5*(2), 197–206.
9. Kledal, T. N.; Rosenkilde, M. M.; Coulin, F.; Simmons, G.; Johnsen, A. H.; Alouani, S.; Power, C. A.; Lutichau, H. R.; Gerstoft, J.; Clapham, P. R.; Clark-Lewis, I.; Wells, T. N.; Schwartz, T. W. A broad-spectrum chemokine antagonist encoded by Kaposi's sarcoma-associated herpesvirus. *Science* **1997**, *277*, 1656–1659.
10. Lutichau, H. R.; Lewis, I. C.; Gerstoft, J.; Schwartz, T. W. The herpesvirus 8-encoded chemokine vMIP-II, but not the poxvirus-encoded chemokine MC148, inhibits the CCR10 receptor. *Eur. J. Immunol.* **2001**, *31*, 1217–1220.
11. Qin, L.; Kufareva, I.; Holden, L. G.; Wang, C.; Zheng, Y.; Zhao, C.; Fenalti, G.; Wu, H.; Han, G. W.; Cherezov, V.; Abagyan, R.; Stevens, R. C.; Handel, T. M. Structural biology. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science* **2015**, *347*, 1117–1122.
12. Chen, S.; Bacon, K. B.; Li, L.; Garcia, G. E.; Xia, Y.; Lo, D.; Thompson, D. A.; Siani, M. A.; Yamamoto, T.; Harrison, J. K.; Feng, L. In vivo inhibition of CC and CX3C chemokine-induced leukocyte infiltration and attenuation of glomerulonephritis in Wistar-Kyoto (WKY) rats by vMIP-II. *J. Exp. Med.* **1998**, *188*, 193–198.
13. Davis, C. N.; Zujovic, V.; Harrison, J. K. Viral macrophage inflammatory protein-II and fractalkine (CX3CL1) chimeras identify molecular determinants of affinity, efficacy, and selectivity at CX3CR1. *Mol. Pharmacol.* **2004**, *66*, 1431–1439.
14. Alexandrov, A.I. Microscale Fluorescent Thermal Stability Assay for Membrane Proteins. *Structure.* **2008**, *16*:351.

15. Stephens, B.; Handel, T.M. Chemokine Receptor Oligomerization and Allostery. In: Kenakin, T., editor. *Oligomerization and Allosteric Modulation in G-Protein Coupled Receptors*. Vol. 115. *Academic Press* **2013**. p. 375chap. 9
16. Navenot, J.M. Molecular anatomy of CCR5 engagement by physiologic and viral chemokines and HIV-1 envelope glycoproteins: differences in primary structural requirements for RANTES, MIP-1 α , and vMIP-II binding. *J Mol Biol.* **2001**, 313, 1181.
17. Moore, P. S.; Boshoff, C.; Weiss, R. A.; Chang, Y. Molecular mimicry of human cytokine and cytokine response pathway genes by KSHV. *Science* **1996**, 274, 1739–1744.
18. Li, Y.; Liu, D.; Cao, R.; Kumar, S.; Dong, C.; An, J.; Wilson, S.R.; Gao, Y.; Huang, Z. Crystal Structure of Chemically Synthesized vMIP-II. *Proteins: Structure, Function, and Bioinformatics* **2007**, 67, 243–246.
19. Fernandez, E.J.; Wilken, J.; Thompson, D.A.; Peiper, S.C.; Lolis, E. Comparison of the structure of vMIP- II with eotaxin-1, RANTES, and MCP-3 suggests a unique mechanism for CCR3 activation. *Biochemistry* **2000**, 39, 12837–12844.
20. Li, Y.; Liu, D.; Cao, R.; Kumar, S.; Dong, C.; An, J.; Wilson, S.R.; Gao, Y.G.; Huang, Z. Crystal structure of chemically synthesized vMIP-II. *Proteins* **2007**, 67, 243–246.
21. LiWang, A.C.; Wang, Z.X.; Sun, Y.; Peiper, S.C.; LiWang, P.J. The solution structure of the anti-HIV chemokine vMIP-II. *Protein Science* **1999**, 8, 2270–2279.
22. Shao, W.; Fernandez, E.; Wilken, J.; Thompson, D.A.; Siani, M.A.; West, J.; Lolis, E.; Schweitzer, B.I. Accessibility of selenomethionine proteins by total chemical synthesis: structural studies of human herpesvirus-8 MIP-II. *FEBS Letters* **1998**, 441, 77–82.
23. Shao, W.; Fernandez, E.; Sachpatzidis, A.; Wilken, J.; Thompson, D.A.; Schweitzer, B.I.; Lolis, E. CCR2 and CCR5 receptor-binding properties of herpesvirus-8 vMIP-II based on sequence analysis and its solution structure. *European Journal of Biochemistry* **2001**, 268, 2948–2959.
24. Zhao, B.; LiWang, P.J. Characterization of the interactions of vMIP-II, and a dimeric variant of vMIP- II, with glycosaminoglycans. *Biochemistry* **2010**, 49(33), 7012–7022.
25. LiWang, A. C.; Cao, J. J.; Zheng, H.; Lu, Z.; Peiper, S. C.; LiWang, P. J. Dynamics study on the anti-human immunodeficiency virus chemokine viral macrophage-inflammatory protein-II (VMIP-II) reveals a fully monomeric protein. *Biochemistry* **1999**, 38, 442–453.
26. Zhao, B.; Liwang, P. J. Characterization of the interactions of vMIP-II, and a dimeric variant of vMIP-II, with glycosaminoglycans. *Biochemistry* **2010**, 49, 7012–7022.
27. Zhou, N.; Luo, Z.; Luo, J.; Hall, J.W.; Huang, Z. A novel peptide antagonist of CXCR4 derived from the N-terminus of viral chemokine vMIP-II. *Biochemistry* **2000**, 39(13), 3782–3787.
28. Clark-Lewis, I.; Kim, K. S.; Rajarathnam, K.; Gong, J. H.; Dewald, B.; Moser, B.; Baggiolini, M.; Sykes, B. D. *J. Leukocyte Biol.* **1995**, 57, 703–711.
29. Heveker, N., Montes, M., Germeroth, L., Amara, A., Trautmann, A., Alizon, M., and Schneider-Mergener, J. (1998) *Curr. Biol.* 8, 369–376
30. Hebert, C. A.; Vitangcol, R. V.; and Baker, J. B. *J. Biol. Chem.* **1991**, 266, 18989–18994.
31. Crump, M. P.; Gong, J. H.; Loetscher, P.; Rajarathnam, K.; Amara, A.; Arenzana-Seisdedos, F.; Virelizier, J. L.; Baggiolini, M.; Sykes, B. D.; Clark-Lewis, I. *EMBO J.* **1997**, 16, 6996–7007.
32. Zhou, N.; Luo, Z.; Luo, J.; Fan, X.; Cayabyab, M.; Hiraoka, M.; Liu, D.; Han, X.; Pesavento, J.; Dong, C. Z.; Wang, Y.; An, J.; Kaji, H.; Sodroski, J. G.; Huang, Z. Exploring the stereochemistry of CXCR4-peptide recognition and inhibiting HIV-1 entry with D-peptides derived from chemokines. *J. Biol. Chem.* **2002**, 277, 17476–17485.

33. Dong, C. Z.; Kumar, S.; Choi, W. T.; Madani, N.; Tian, S.; An, J.; Sodroski, J. G.; Huang, Z. Different stereochemical requirements for CXCR4 binding and signaling functions as revealed by an anti-HIV, D-amino acid-containing SMM-chemokine ligand. *J. Med. Chem.* **2005**, 48, 7923–7924.
34. Choi, W. T.; Tian, S.; Dong, C. Z.; Kumar, S.; Liu, D.; Madani, N., An, J.; Sodroski, J. G.; Huang, Z. Unique ligand binding sites on CXCR4 probed by a chemical biology approach: implications for the design of selective human immunodeficiency virus type 1 inhibitors. *J. Virol.* **2005**, 79, 15398–15404.
35. Mori, M.; Liu, D.; Kumar, S.; Huang, Z. NMR structures of anti- HIV D-peptides derived from the N-terminus of viral chemokine vMIP-II. *Biochem. Biophys. Res. Commun.* **2005**, 335, 651–658.
36. Baggiolini, M.; Dewald, B.; Moser, B. Human chemokines: an update. *Annu. Rev. Immunol.* **1997**, 15, 675–705.
37. Crump, M.P.; Gong, J.H.; Loetscher, P.; Rajarathnam, K.; Amara, A. Solution structure and basis for functional activity of stromal cell-derived factor- 1; dissociation of CXCR4 activation from binding and inhibition of HIV-1. *EMBO J* **1997**, 16, 6996–7007.
38. Shan, L.; Qiao, X.; Oldham, E.; Catron, D.; Kaminski, H. Identification of viral macrophage inflammatory protein (vMIP)-II as a ligand for GPR5/ XCR1. *Biochem Biophys Res Commun* **2000**, 268: 938–41.
39. Crump, M.P.; Elisseeva, E.; Gong, J.; Clark-Lewis, I.; Sykes, B.D. Structure/ function of human herpesvirus-8 MIP-II (1–71) and the antagonist N-terminal segment (1–10). *FEBS Lett* 2001, 489, 171–175.
40. Booth, V.; Slupsky, C.M.; Clark-Lewis, I.; Sykes, B.D. Unmasking ligand binding motifs: identification of a chemokine receptor motif by NMR studies of antagonist peptides. *J Mol Biol* 2003, 327, 329–34.

3.5 Experimental Section

3.5.1 Chemistry

Materials. N^α-Fmoc-protected amino acids, 2-Cl-trtCl resin, Fmoc-Rink Amide-Am resin, O-benzotriazole-N,N,N',N'-tetramethyl-uroniumhexafluorophosphate (HBTU), N,N-diisopropylethylamine (DIEA), triisopropylsilane (TIS), trifluoroacetic acid (TFA), and piperidine were purchased from IRIS Biotech, N-hydroxybenzotriazole (HOBt), N,N-dimethylformamide (DMF), dichloromethane (DCM), and from Iris-Biotech GmbH (Marktredwitz, Germany). Peptide synthesis solvents, reagents, H₂O and CH₃CN for HPLC were reagent grade and were acquired from commercial sources (Sigma-Aldrich, Milano, Italy) and used as received unless otherwise noted. Peptides were purified by preparative HPLC (Shimadzu HPLC system) equipped with a C18-bounded preparative RP-HPLC column (Phenomenex Kinetex 21.2 mm × 150 mm 5 μm). Peptides were analyzed by analytical HPLC (Shimadzu Prominence HPLC system) equipped with a C18-bounded analytical RP-HPLC column (Phenomenex Kinetex, 4.6 mm × 250 mm 5 μm) using a gradient elution (10–90% acetonitrile in water (0.1% TFA) over 15 min; flow rate = 1.0 mL/min; diode array UV detector). Molecular weights of compounds were confirmed by ESI-mass spectrometry using an Agilent 6110 quadrupole LC/MS system.

General Procedure for the Synthesis of Peptides. 2-Cl-trtCl resin (62.0 mg, 1.60 mmol/g) was swollen in DMF dry over 0.5 h, and a solution of Fmoc-*L*-Pen(trt)-OH (92.1 mg, 0.15 mmol, 0.5 equiv) and DIPEA (26 μl, 1.5 eq) in DMF (2ml) was added. The mixture was stirred for 10 min. An additional amount of DIPEA (34.8 μl, 0.2 mmol, 2 equiv) was added and the mixture was then shaken for 1 h. The residual chloride groups contained in the resin were capped by adding MeOH (200 μl) in DCM (2 ml) and stirring for 1 h to avoid eventually parallel synthesis of side products. Fmoc group removal was performed using 20 % piperidine in DMF (1 x 5 min and 1 x 25 min). The peptide resin was then washed with DCM (3 x 0.5 min) and DMF (3 x 0.5 min) and positive Kaiser ninhydrine³⁵ and TNBS³⁶ tests were observed. Fmoc-*L*-Phe-OH (155.0 mg, 0.4 mmol, 4 equiv) or Fmoc-*L*-Tyr-OH (155.0 mg, 0.4 mmol, 4 equiv) or Fmoc-*L*-(4-Cl)Phe-OH (169 mg, 0.4 mmol, 4 equiv), Fmoc-*L*-(4-NO₂)Phe-OH (173 mg, 0.4mmol, 4 equiv) or Fmoc-*L*-1-Nal-OH (175 mg, 0.4 mmol, 4 equiv) or Fmoc-*L*-2-Nal-OH (175 mg, 0.4 mmol, 4 equiv) or Fmoc-*L*-Cys(Trt)-OH (234.3 mg, 0.4 mmol, 4 equiv) or Fmoc-*D*-Pen(Trt)-OH (245.5 mg, 0.4 mmol, 4 equiv), Fmoc-*L*-Ala-OH (124.5 mg, 0.4 mmoli, 4 equiv) or Fmoc-*L*-Arg-OH (260 mg, 0.4 mmol, 4 equiv) were sequentially added sequentially to the resin bound H-*L*-Pen(Trt). Each coupling reaction was achieved using a 4-fold excess of amino acid with HBTU (151.7 mg, 0.4 mmol, 4 equiv) and HOBt (61.2 mg, 0.4

mmol, 4 equiv) in the presence of DIPEA (140 μ l, 0.8 mmol, 8 equiv) in DMF. Fmoc deprotections were accomplished with 20% piperidine in DMF solution (1 x 5 min, 1 x 25 min). Washings with DMF (3 x 0.5 min) and DCM (3 x 0.5 min) were performed through every coupling/deprotection step. Kaiser ninhydrine and TNBS tests were employed for monitoring the progress of peptide synthesis. For the peptides bearing an acetyl group at the N-terminus, after removing the last Fmoc group, the resin bound peptide was treated with Ac₂O (19 μ l, 0.2 mmol, 2 equiv) and DIPEA (35 μ l, 0.2 mmol, 2 equiv) in DCM (2 ml) and the mixture was shaken for 1 h. A negative Kaiser ninhydrine and TNBS tests were observed.

General Procedure for Peptide Oxidation and Purification. The peptide was released from the solid support and all the protecting groups cleaved, treating the resin with TFA/DCM/TIS (80/15/5, v/v/v) (3 ml solvent/0.1 mmol) for 2 h. The resin was then filtered off and the crude linear peptide was recovered by precipitation with chilled ether to give a powder. The crude deprotected peptide (0.1 mmol) was dissolved in H₂O/acetonitrile (1:1) and a solution containing 1.5 equiv. of NCS was added. The mixture was then mechanical shaken for 15 min at room temperature, lyophilized and subsequently analysed. Oxidized cyclic eptapeptides were synthesized in high purity, and no linear precursors were present. Final peptide purification was achieved by preparative RP-HPLC in 0.1% TFA with an ACN gradient (10–90% ACN in H₂O over 15 min, flow rate of 15 mL/min) on a Phenomenex Kinetex C18 column (21.2 mm \times 150 mm 5 μ m). Analytical RP-HPLC were performed in 0.1% TFA with an ACN gradient (10–90% ACN in H₂O over 20 min, flow rate of 1.0 mL/min) on a Phenomenex Kinetex C18 column (0.46 mm \times 150 mm 5 μ m).

17: Purity > 95%, t_R 14.7 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 20 min, flow rate of 1.0 mL/min); molecular formula: C₄₃H₆₃N₁₃O₉S₂; calculated mass: 969.4; found: 970.5 (M+H⁺), 485.8 (M+2H⁺)/2, 987.4 (M + Na⁺).

18: Purity > 95%, t_R 14.7 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 20 min, flow rate of 1.0 mL/min); molecular formula: C₄₃H₆₃N₁₃O₉S₂; calculated mass: 969.5; found: 970.5 (M+H⁺), 485.9 (M+2H⁺)/2.

19: Purity > 95%, t_R 14.9 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 20 min, flow rate of 1.0 mL/min); molecular formula: C₄₅H₆₇N₁₃O₉S₂; calculated mass: 997.5; found: 998.6 (M+H⁺), 499.3 (M+2H⁺)/2.

20: Purity > 95%, t_R 15.5 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 20 min, flow rate of 1.0 mL/min); molecular formula: C₄₃H₆₃N₁₃O₁₀S₂; calculated mass: 985.4; found: 986.5 (M+H⁺), 493.4 (M+2H⁺)/2.

21: Purity > 95%, t_R 15.4 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 20 min, flow rate of 1.0 mL/min); molecular formula: C₄₃H₆₂N₁₃O₉S₂; calculated mass: 1003.4; found: 1004.5 (M+H⁺), 502.3 (M+2H⁺)/2.

22: Purity > 95%, t_R 14.9 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula: C₄₃H₆₂N₁₄O₁₁S₂; calculated mass: 1014.4; found: 1015.5 (M+H⁺), 507.8 (M+2H⁺)/2, 1038.5 (M+Na⁺).

23: Purity > 95%, t_R 15.7 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula: C₄₇H₆₅N₁₃O₉S₂; calculated mass: 1019.5; found: 1020.4 (M+H⁺), 510.8 (M+2H⁺)/2.

24: Purity > 95%, t_R 15.8 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula: C₄₇H₆₅N₁₃O₉S₂; calculated mass: 1019.5; found: 1020.4 (M+H⁺), 510.7 (M+2H⁺)/2.

25: Purity > 95%, t_R 14.0 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 20 min, flow rate of 1.0 mL/min); molecular formula: C₄₃H₆₃N₁₃O₁₀S₂; calculated mass: 985.4; found: 986.5 (M+H⁺), 493.9 (M+2H⁺)/2.

26: Purity > 95%, t_R 15.3 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 20 min, flow rate of 1.0 mL/min); molecular formula: C₄₃H₆₂N₁₃O₉S₂; calculated mass: 1003.4; found: 1004.5 (M+H⁺), 502.8 (M+2H⁺)/2.

27: Purity > 95%, t_R 14.9 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula: C₄₃H₆₂N₁₄O₁₁S₂; calculated mass: 1014.4; found: 1015.4 (M+H⁺), 507.8 (M+2H⁺)/2, 1038.5 (M+Na⁺).

28: Purity > 95%, t_R 15.6 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula: C₄₇H₆₅N₁₃O₉S₂; calculated mass: 1019.5; found: 1020.4 (M+H⁺), 510.6 (M+2H⁺)/2.

29: Purity > 95%, t_R 15.5 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula: C₄₇H₆₅N₁₃O₉S₂; calculated mass: 1019.5; found: 1020.3 (M+H⁺), 510.5. (M+2H⁺)/2.

30: Purity > 95%, t_R 12.5 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula: C₄₄H₆₃N₁₅O₉S₂; calculated mass: 1009.4; found: 1010.5 (M+H⁺).

Human Plasma Stability Assay. Human plasma, H₂O, ACN, and TFA were obtained from Sigma-Aldrich (Milano, Italy) and used without further purification. Human plasma was diluted to 90% in pure water. Analytical HPLC-ESI-MS was performed on an Agilent Technologies 1200 series equipped with an Agilent Technologies 6110 quadrupole LC/MS using a Phenomenex Luna C18 column (5 μm, 4.6 mm × 150 mm) and H₂O (0.1% v/v TFA)/ACN (0.1% v/v TFA) as eluents.

A solution of 1 mg/mL of each peptide was prepared in water, and 150 μL aliquots were mixed with 150 μL of prewarmed (37 °C) plasma. At selected time points (0, 30, 60, 120), samples (50 μL) were collected and mixed with 1% TFA in acetonitrile (75 μL) to precipitate plasma proteins which were deleted by centrifugation at 13000 rpm for 10 min. The supernatant was analyzed by HPLC-ESI-MS using a Phenomenex Luna C18 column (5 μm, 4.6 mm × 150 mm) and an elution gradient of 10–90% solvent B over 20 min (solvent A, 0.1% TFA in water; solvent B, 0.1% TFA ACN) at a flow rate of 1 mL/min.

3.5.2 Biological Studies

Binding Assay. CXCR4 binding was evaluated as previously described. Briefly, 5×10^5 CCRF–CEM, HT29 cells were preincubated with increasing peptide concentrations (0.01 μM, 0.1 μM, 1 μM, 10 μM) in the binding buffer (PBS 1× plus 0.2% BSA and 0.1% NaN₃) for 30 min at 37 °C, 5% CO₂ and then labeled for 30 min using anti-CXCR4 PE-antibody (FAB170P, clone 12G5, R&D Systems, Minneapolis, MN, USA). To evaluate the specific peptide binding to CXCR4, the experiments were also conducted in COLO205, human colon cancer cells, overexpressing CXCR3, using anti-CXCR3 FITC-antibody (R&D FAB160F clone 49801) and in MCF-7, human breast cancer cell line, overexpressing CXCR7, using anti-CXCR7 APC-antibody (R&D FAB4227A clone

11G8). The cells were counted through a FACS Canto II cytofluorometer (Becton Dickinson Immunocytometry Systems, Mountain View, CA, USA).

Chapter 4

Pre-Clinical Development of a Novel Cyclic Peptide ^{68}Ga PET Tracer for Colon Cancer

4.1 Introduction

Individual behavioral heterogeneity within a given cluster of diseases is a major challenge in clinical routine, especially in the field of oncology¹. Different sub-specialities of –omics approaches draw back the curtains on human vulnerability revealing peculiar traits among patients, but allowing at the same time the possibility to set-up a genetic, proteomic and metabolic profile. Thus, the “one site fits all” model has been converted in a thoroughly evidence-based tailored therapy, changing the overall landscape of medical paradigms, from diagnosis to skillful planning and monitoring treatment.² In this context, as an intersection of molecular biology and *in vivo* imaging, molecular imaging, through a whole-body read out in an intact living system, provides the assessment and management of specific information and functions, such as pathway activities and cell migration (Figure 4.1).^{3,4}

To date, several imaging tools have been developed. Among them, Positron Emission Tomography (PET) can measure biochemical and physiological aberrations that occur prior to macroscopic anatomical signs of disease, in contrast to the conventional imaging modalities which mainly provide detailed anatomical images.⁵ Following intravenous administration of a specific positron-emitting–bound biological molecules (tracer), PET enables uniform and dynamic scintigraphic snapshots at the molecular level in order to precisely track, visualize and quantify selected events in an unrivaled high sensitively manner⁶. Mechanistically, PET images in real time intact subjects by detecting pairs of high-energy γ -rays emitted indirectly by a tracer, which is injected in the test subject. As the radioisotopes decay the resulting positrons (β^+) eventually gets annihilated on contact with nearby electrons. The two photons produced each with energy of approximately 511 keV have trajectories 180 degrees opposite to each other.^{7,8} Images of the “coincidence events”, and so of the tracer concentration, in 3D space within the body are then reconstructed by computer analysis (Figure 4.2a).³

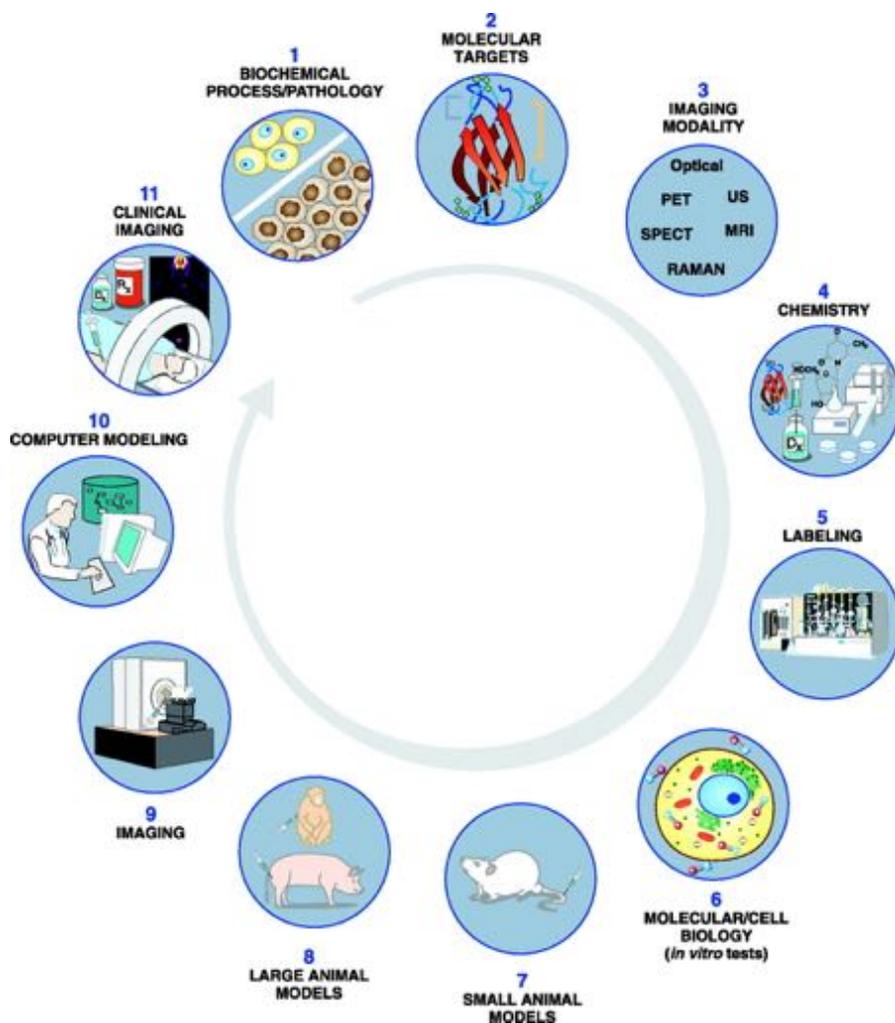


Figure 4.1. The overall process of molecular imaging research. The first step of this time consuming and technically demanding methodology is to identify a biochemical process or pathology of interest, and to assess the significance of visualizing this process/ pathology noninvasively via the tools of molecular imaging. The second step is to decide on a molecular target that will enable the visualization of the phenomena of interest. This is usually followed by selection of an appropriate imaging modality and a designed and synthesized imaging biomarkers. A number of *in vitro* (molecular/cell biology-based) and *in vivo* (animal model-based) tests are required to evaluate the target specific binding and favorable pharmacokinetics. If clinical studies are the end goal, FDA approval is required, and certain mathematical models/algorithms might need to be developed so that meaningful data can be obtained from images. (Adapted from James M. L. *et al.*, *Physiol Rev* 2012)

Conventional wisdom informs that transformed cancer cells exhibit elevated rate of glucose consumption, a phenomenon that has been clinically capitalized upon for metabolic nuclear imaging using ^{18}F -fluoro-deoxy- D-glucose (^{18}F -FDG).⁹ As an analog of glucose, this radiopharmaceutical is taken up into several tissue, phosphorylated and, then, trapped intracellularly emitting positrons: the further catabolism is, in fact, hindered by the lack of an oxygen atom at the C-2 position (Figure 2c). These areas of accumulation, used as metabolic traps, are detected by comparing the uptake with background activity.¹⁰ However, not all ^{18}F -FDG-mediated PET-positive lesions are cancer.¹¹⁻¹³ In general, the more metabolically active the cell, the more uptake of glucose. For instance, inflammatory cells as neutrophil and activated macrophages at the site of inflammation or infection also show high metabolic accumulation of the carbohydrate and, as a result, are FDG avid. Moreover, not all cancer cells use the same amount of glucose: the prostate would be poorly imaged by ^{18}F -FDG due to its low metabolic rate. Thus, a multitude of artifacts could be created confounding diagnosis.¹⁴

Therefore, there is the need of cancer-specific target that allows the study of a certain inner working of the human body.

Consequently, the development of optimal probes, which interrogate molecular targets, becomes the pivotal step in molecular imaging research. Translation of a biomarker from the bench to the animal models needs a deep optimization's study,¹⁵ regarding its design, synthesis, composition, reactivity and nuclear properties, that can be defined PET-chemistry.

A tracer typically comprises a signal agent (radionuclide), coupled by a linker/spacer to a targeting moiety, such as small molecule, peptide, protein, antibody and its fragments.¹⁶⁻¹⁹ To be successful, a desirable radioligand is expected to have some unique characteristics.^{20,21} The chemistry of the probe should preserve significant specificity for the process of interest and so high binding affinity to the target, achieving the ideal accumulation in the aimed tissue by preventing nonexclusive uptake.²² Elevated sensitivity and contrast ratio are also demanded: a minimal amount of probe is expected to provide images with high signal-to-noise in order to ensure statistically appropriate interpretation. High *in vivo* stability and favorable pharmacokinetics are necessary for crossing biological barriers, reaching the sufficient concentration at the specific target and retaining long enough time for imaging.^{23,24} Spacer molecules can influence the pharmacokinetic profile leading to an increase/decrease in lipophilicity of the whole construct.^{16,19,25,26} In terms of labeling,²⁷⁻²⁹ over the past several years in addition to the conventional positron-emitting radionuclides, such as ^{18}F , other metallic PET isotopes have been robustly developed,³⁰⁻³⁵ including ^{64}Cu , ^{68}Ga , ^{86}Y and ^{89}Zr . These innovative signaling agents are usually characterized by a longer half-life, allowing longitudinal

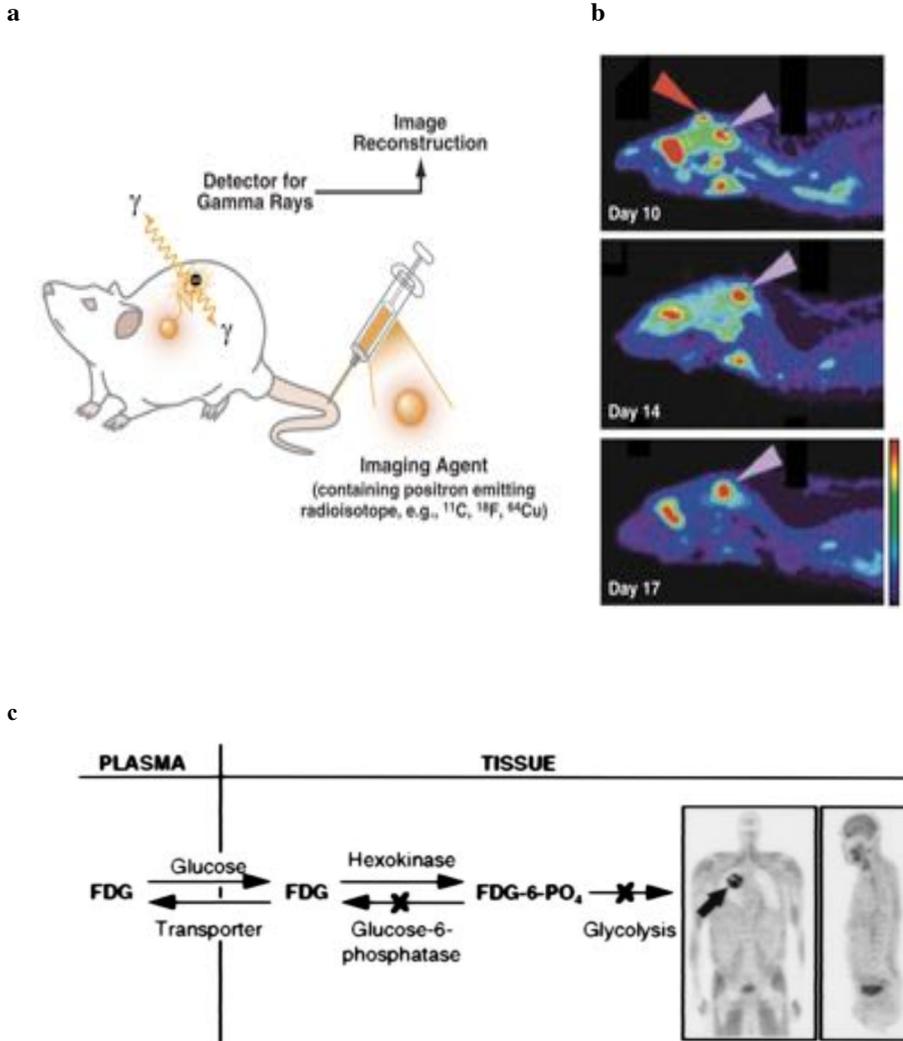


Figure 4.2. Small animal positron emission tomography (PET). a) schematic illustrating the basic principles of PET. First, a targeted imaging agent containing a positron emitting radioisotope is administered to the subject. Positrons are emitted from each imaging agent only once; these positrons travel short distances and collide with electrons in the surrounding tissues resulting in the production of two back-to-back 511 keV annihilation photons. Following the detection of gamma rays by the PET detector, the tomographic images are reconstructed from the electronic signals reflecting the distribution of the imaging agent in the subject, and hence providing information on the biochemical event the agent was targeting. b) images demonstrating the noninvasive visualization of an orthotopic brain tumor (2.5 mm in diameter) in a rat via the use of [^{18}F]-2-fluoro-2-deoxy-glucose ([^{18}F]FDG), as early as 10 days (D10) after implantation. Pink arrows show tumor, and red arrow shows wound due to intracerebral implantation of tumor cells. (Adapted from James M. L. et al., *Physiol Rev* 2012) c) Tracer kinetic models for FDG (Adapted from Phelps M. *Proc Natl Acad* 2000)

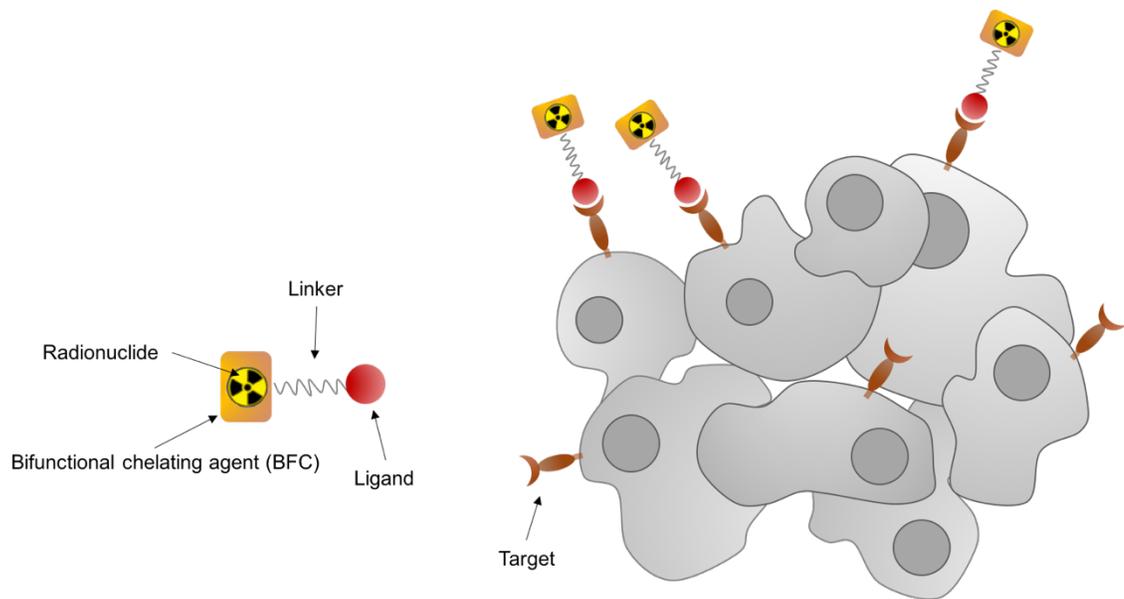


Figure 4.3. Schematic presentation of a peptide- conjugated radiotracer design.

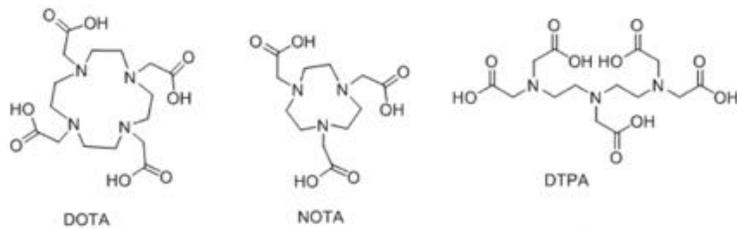


Chart 4.1. Examples of BFCAs used for radiolabeling.

studies on the same subject with a single injection of radiotracers. Then, the metal ion has to be matched with a suitable chelate system giving an effective combination of favorable coordination and stability properties. In addition to the metal binding moiety function, the chelator possesses a chemically reactive group for covalent attachment to the vector of interest, in a “bifunctional chelating” manner (Figure 4.3).³⁶ Hence the need of labeling standardized chemicals, quicker and easier to attach. Prominent examples are 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA),³⁷⁻⁴³ 1,4,7-triazacyclononane-1,4,7-trisacetic acid (NOTA)⁴⁴⁻⁴⁷ and the non-cyclic system diethylenetriamine-N,N,N',N',N'''-pentaacetic acid (DTPA).⁴⁸ Due to the open structure of the polyamino polycarboxy, DTPA complexes are rapidly formed and therefore the chelator is used for radiolabelling of sensitive proteins. DOTA's flexibility enables the accommodation of a broad range of radio-metals. The cage of NOTA, instead, fits perfectly the gallium III already at mild conditions (Chart 4.1).

The application of diagnostic probes with the ability to selectively target CXCR4 could have many potential advantages, including the ability to evaluate primary tumors for elevated levels of CXCR4 indicating the likelihood of recurrence and potential for metastasis, the ability to track metastatic spread throughout the body, facilitating rapid, targeted therapy, and the guidance of surgical procedures, thereby ensuring maximal tumor excision (Figure 4.4).^{49,50}

T140, a 14-mer peptide containing a single disulfide bridge and an amidated C-terminus, reported as a specific inverse agonist against CXCR4,⁵¹ has been used as a scaffold for the synthesis of receptor-targeted imaging probes (Chart 4.2). The successful introduction of a novel pharmacophore, a 4-fluorobenzoyl moiety, hooked up to the N-termini of the sequence, has resulted in the analog TN14003. This compound has been further labeled with the PET isotope fluorine-18 affording ¹⁸F-TN14003, which showed significant accumulation in CHO-CXCR4 positive tumors, but also high red blood cells (RBCs) uptake, demanding the co-injection of additional unlabeled mass of the peptide to render the probe available and visualize CXCR4 in vivo.⁵² Another downside was the time-consuming radiosynthesis of ¹⁸F-TN14003 and the low yield. To address this problem two DOTA molecules have been introduced, one on each of the free lysine residues of ¹⁸F-TN14003, and the following labeling was achieved with PET metal isotope copper-64 leading to [⁶⁴Cu]T140-2D. The conjugation did not disrupt the binding affinity to CXCR4, but the tracer exhibited similar behavior as ¹⁸F-TN14003 in terms of binding to mouse RBCs. RBCs express in fact scavenger receptors such as DARCs (Duffy antigen receptor of chemokine). These serpentine molecules intercept various chemokines and bind them “silently”.⁵³ However, the knockout of DARC receptors in mice does not eliminate the blood uptake, suggesting that the binding site for the CXCR4

antagonist on RBC is yet to be found. [⁶⁴Cu]T140-2D showed also elevated concentration in metabolic organs, such as liver, intestine, that remained high over time. In an effort to reduce the non-specific uptake, Chen et al. labeled the compound with another PET isotope, ⁶⁸Ga. Similar effects were obtained proving that the accumulation is a result of transchelation of the metal ion.⁵⁴ In a follow-up study, the same team resolved the issue of off-target interactions in their development of two new PET radiotracers, ⁶⁴Cu–NOTA–NFB and ⁶⁴Cu–DOTA–NFB: the substitution of the 4-fluor-benzoyl ring (NFB) with the chelator reduced the binding to RBC, but it also affected the affinity with CXCR4.⁵⁵

With the aim of preserving the four critical residues of T140 (Arg2, 2Nal3, Tyr5, Arg14) positioned across the disulfide bridge, a molecular-size reduction has been carried out. The downsized FC131 has been used as ⁶⁸Ga-labeled probe (⁶⁸Ga-CPCR4-2 or ⁶⁸Ga-Pentixafor);^{56,57} The cyclic penta-peptide showed a certain *in vitro* stability as well as a specific accumulation in CXCR4 positive tumors. However, the antagonistic activity is lower compared with the dimeric structure of FC131, which gave, by the way, undesired alteration of biodistribution due to its lipophilicity.

Another noteworthy research to image CXCR4 has focused on small molecule-based antagonist. The cyclam structure of AMD 3100⁵⁸⁻⁶⁰ and AMD3465⁶¹ offers the opportunity, without losing crucial pharmacophoric crucial interactions, for labeling with ⁶⁴Cu, which is usually incompatible with small molecule due to the substantial size of the chelator moieties. The serendipitously discovered AMD3100, originally founded as an active impurity during a research aimed at developing an anti-HIV agent, was recently approved for clinical use to mobilize hematopoietic CD34+ stem cells from the bone marrow into the circulation. ⁶⁴Cu-AMD3100 and the monocyclam analog ⁶⁴Cu-AMD3465 showed specifically accumulation in CXCR4 expressing tumors and rapid clearance from the blood, but suffer of high uptake into the liver ($\geq 40\%$ ID/6 and $\geq 32\%$ ID/6 respectively).

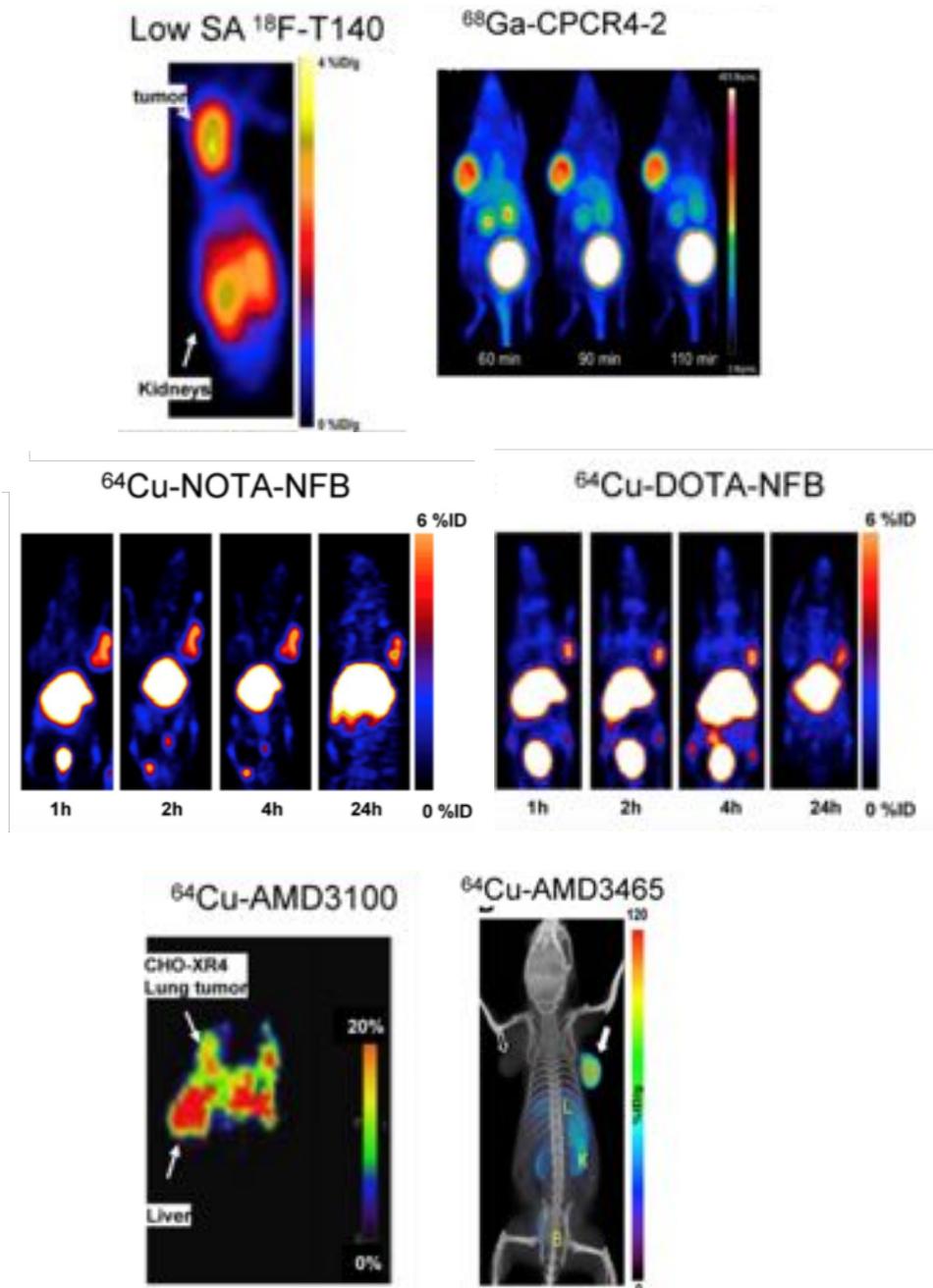


Figure 4.4. PET imaging of subcutaneous tumors using CXCR4 specific tracers. (Adapted from Weiss I.D., *Theranostics* 2013)

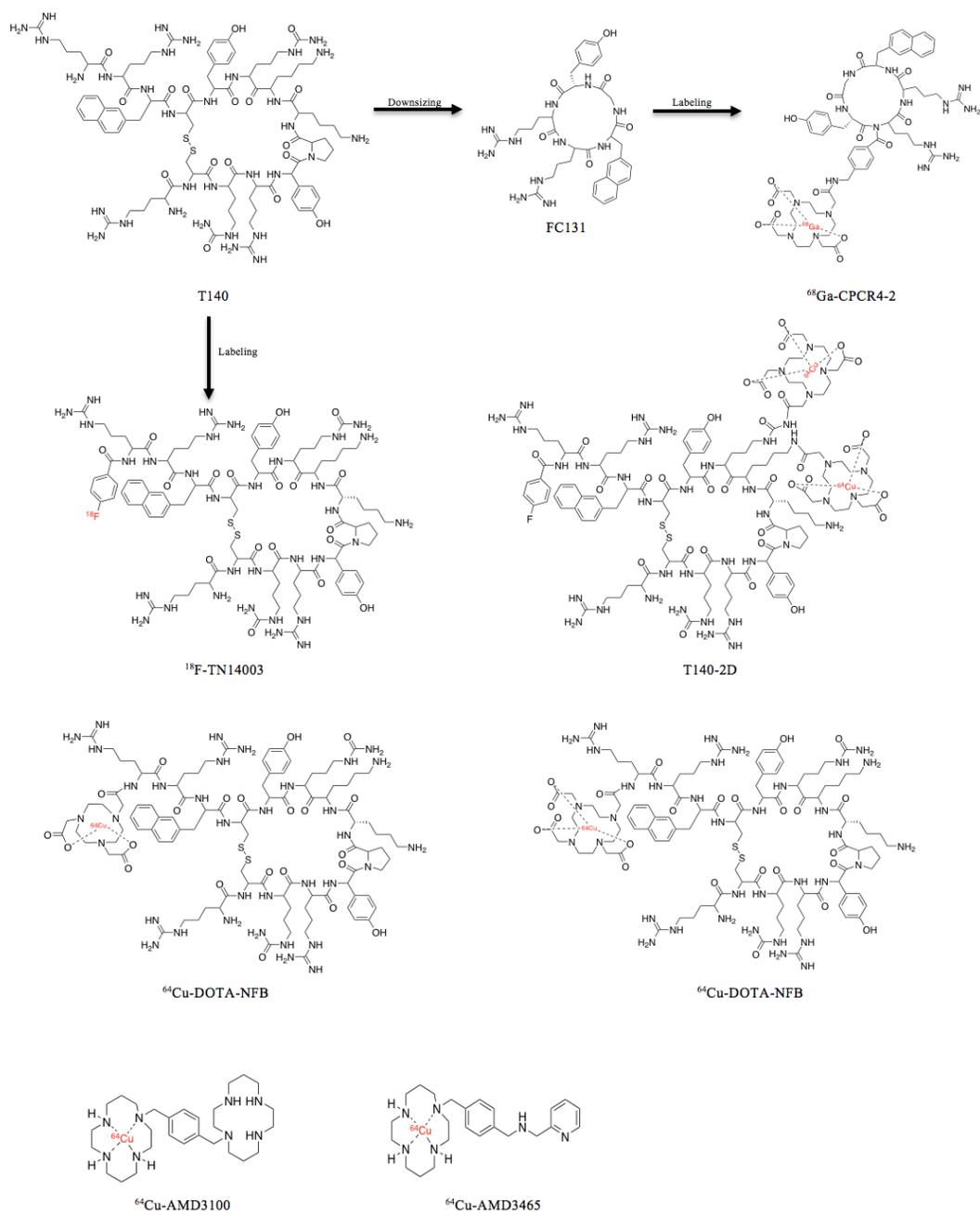


Chart 4.2. The chemical structure of CXCR4 ligands

4.2 Results and discussion

4.2.1 Design

Inspired by the promising results found for peptide **30**, its potential use as radiolabeled probe to targeting vectors was investigated. However, due to the expensive chemical synthesis of this peptide, we decide to work on its parent peptide **10**, which has been already reported as a remarkable CXCR4 antagonist. In fact, **30** contains, in fact, multiple unnatural amino acids and a low first coupling yield owing to the *L*-Pen⁷. On the other hand, peptide **10**, preserving the successful modification of D-Cys at position 3, appeared to be a robust model for the subsequent molecular imaging-oriented SAR study. Compared to the previously synthesized compounds, **10** combines remarkable CXCR4 affinity with a high the stability against enzymatic degradation, thanks to the N-terminal acetyl unit, and the commonly low toxicity of peptides.

With the goal of identifying a valid anchor point within the peptide sequence suitable for a chelating agent, we valued the introduction in different positions of multiple residues bearing the amino moiety. The resulting peptide bond was chosen over the others because of its versatile and advantageous conjugation strategy compatible with SPPS, avoiding in this way additional critical steps which may reduce the final synthesis efficiency. These efforts resulted in a library of cyclopeptides- and cyclopeptidemimics- based CXCR4 specific probe molecules. The Lys ϵ -amino group and the N-terminal of the aminohexanoic acid (6-Aha) and 4-(aminomethyl)benzoic acid (AMBS) were selected as moorings, as it is often reported in the literature. A Lys residue has been included in the C-terminal region of the peptide template, with or without the spacer aminohexanoic acid. The chirality of the side-chain carrying amino acid has been also investigated in the same position. Moreover, the Ala² has been replaced by a Lys with the aim of explore putative non sterically hindered rearrangements within the receptor. We tested then the suitability of a heteroalkyl spacer at the amino terminus of the sequence introducing a AMBS group. Consequently, the labelling would be appended to the Lys side-chains using the orthogonal group Alloc or linked to the NH₂ termini by Fmoc, depending on the activities of the tested compounds.

4.2.2 Chemistry

Chain assembly of all the linear peptides was performed manually employing standard Fmoc-based SPPS. The loading of the first amino acid building block onto 2-Cl chlorotrityl resin was

achieved using an excess of resin (2 equivalent) with respect to the amino acid, in order to increase the yield of the desired product. To obtain the radiolabeled molecules, the selective peptides were conjugate to the appropriate chelator for the trivalent metal Ga.

Table 4.1. Half-Maximal Inhibitory Concentration (IC₅₀, μM, Mean ± SD) of CXCR4 Antagonist Peptides Necessary to Reduce by 50% The Binding to CXCR4 of a) CXCR4-Specific mAb 12G5 in CCRF-CEM Cells or b) [¹²⁵I]FC-131 in Jurkat T-cell. Each experiment was performed in triplicate.

comp	sequence	IC ₅₀ (μM)
30	Ac-Arg-Ala-[DCys-Arg-2Nal-His-Pen]-COOH ^a	0.0015 ± 0.0005
10	Ac-Arg-Ala-[DCys-Arg-Phe-Phe-Cys]-COOH ^a	0.053±0.004
31	Ac-Arg-Ala-[DCys-Arg-Phe-Phe-Cys]-Lys(Alloc)-COOH ^a	1.43±0.02175
32	Ac-Arg-Ala-[DCys-Arg-Phe-Phe-Cys]-DLys(Alloc)-COOH ^a	1.27±0.0244
33	Ac-Arg-Ala-[DCys-Arg-Phe-Phe-Cys]-Aha-Lys(Alloc)-COOH ^a	0.92±0.0134
34	Ac-Arg-Lys(Alloc)-[DCys-Arg-Phe-Phe-Cys]-COOH ^a	1.430±0.01864
35	Fmoc-Aha-AMBS -Arg-Ala-[DCys-Arg-Phe-Phe-Cys]-COOH ^a	0.118±0.00289
36	Fmoc- AMBS-Aha-Arg-Ala-[DCys-Arg-Phe-Phe-Cys]-COOH ^a	0.013±0.00028
37	DOTA- AMBS-Aha-Arg-Ala-[DCys- Arg-2Nal-His-Pen]-COOH ^b	0.203±0.00406
38	NOTA- AMBS-Aha-Arg-Ala-[DCys-Arg-2Nal-His-Pen]-COOH ^b	0.042±0.00098
39	^{nat} Ga- DOTA-AMBS- Aha-Arg-Ala-[DCys-Arg-2Nal-His-Pen]-COOH ^b	0.049±0.00084
40	^{nat} Ga- NOTA-AMBS- Aha-Arg-Ala-[DCys-Arg-2Nal-His-Pen]-COOH ^b	0.015±0.00032

We used the active-ester tri-*t*-butyl-protected DOTA or the active-ester di-*t*-butyl-protected NOTA, because their multiple potential sites for conjugation could lead to a mixture of cross-linked conjugates. The low water solubility of those compounds is ideal for reacting peptides in non-aqueous solvents. Thus, the conjugation reaction was achieved preincubating the DOTA-tri-*t*-butyl ester/ NOTA-di-*t*-butyl ester in DMF with equimolar amounts of the coupling agents, PyAOP, HOAt and DIPEA. The peptide was then added to this mixture in a peptide:chelator ratio of 1:3. Once assembled on the solid support peptides **31-36** were cleaved off the resin keeping the acid stable orthogonal group Alloc appended to the Lys side-chains (**31-34**) or Fmoc moiety linked to the NH₂ termini (**35-36**). Those elements were employed to mimic the steric hindrance of the bifunctional chelators in the CXCR4 binding assay. Alternatively, the linear peptides **37-40** were removed from the resin as fully deprotected crude peptide. Cyclization via side chain sulfur atoms was achieved in solution using N-Chloro succinimide to form the desired disulphide bridge. Additionally, the conjugated DOTA/NOTA peptide were labeled with the radionuclide ^{nat}Ga via the metal chloride salt.

4.2.3 CXCR4 Binding Assay

Initially, the receptor binding inhibition of PE-conjugated-12G5 anti-CXCR4 antibodies was assayed in CEM-CCRF human T leukemia CXCR4 expressing cells for the **31-36** (Table 4.1) synthesized peptides. As shown in Table 1, analysis of the binding data revealed that peptide **36** possesses an IC₅₀ value of around 0.013 μM, which is significantly lower than that of the other analogues and of the lead compound **10**. Thus, we hypothesize that the presence of the aminohexanoic acid maximizes the distance between the label and the peptide core part, avoiding negative effects of the former on the binding affinity. Besides, the spacer flexible six carbons chain is efficiently stabilized by the aromatic ring of the AMBS that permits the right accommodation of the peptide at the receptor binding cavity. Thus, we translocated our promising adjustments to the our most promising candidate **30**, introducing at the C-terminal of the sequence the spacer linkers. Carboxymethyl pendant-armed derivates of tetraza- (DOTA) or triaza- (NOTA) macrocyclic rings have been adopted as stable chelating agents to complex ^{nat}Ga. Currently, DOTA remains the most frequently used chelator because of its availability and well-recognized coordination chemistry.⁶² However, it has been noticed that its six coordinate analog NOTA forms slightly deformed octahedral complexes with Gallium-68 which display higher thermodynamically stability (log K_{NOTA-Ga(III)} = 31.0 vs. log K_{DOTA-Ga(III)} = 21.3).^{63,64} We decided to assay both DOTA and NOTA-derived

bifunctional chelators and the resulting compounds have been used for the competition binding experiments of ^{125}I -FC131 to CXCR4-expressing Jurkat cells. Notably, NOTA labeled **40** (Table 4.1), compared to its DOTA counterparts, showed high affinity to the chemokine receptor with a half maximal inhibitory concentration (IC_{50}) of $0.015 \pm 0.32 \mu\text{M}$.

4.3 Conclusion

Due to its overexpression in a variety of human cancer type and its frequently associated tumor aggressiveness and poor prognosis, CXCR4 represents a highly relevant diagnostic and therapeutic target. These observations underline and support the non-invasive imaging of CXCR4 as a highly interesting new diagnostic or prognostic biomarker through the *in vivo* quantification of CXCR4 expression levels in tumors.

To meet the clinical need for a highly specific and sensitive tool for the chemokine receptor 4 assessment and quantification *in vivo*, a CXCR4 targeted nuclear probe has been developed. Our previous reported potent, selective, functional antagonist, peptide **30**, has been tailored as a target-specific CXCR4 imaging probes with NOTA. The direct conjunction of the relative bulky radiometal chelate to the small cyclic peptide substantially affects the binding affinity. Thus, the optimal site of attachment and optimization of suitable linker units were carried on. Exploring the flexibility of the CXCR4 binding pocket, we had the best results employing 6-Aha acid and AMBS as moieties, that enabled us to preserve the micro-well bound of the peptide. Despite the additional moieties, the detectable compound **38** showed high affinity to the CXCR4 receptor and the collected results suggest that the ^{nat}Ga is a promising agent for preclinical non-invasive PET imaging.

4.4 References

1. (a) Mahajan, A.; Desai, S.; Kawthalkar, A. S.; Thakur, M. H. Molecular functional imaging in personalized clinical oncology: The road less traveled. *Indian Journal of Medical and Paediatric Oncology. Official Journal of Indian Society of Medical & Paediatric Oncology* **2016**, 37(1), 1–3. (b) 1. O'Farrell AC, Shnyder SD, Marston G, Coletta PL, Gill JH. Non-invasive molecular imaging for preclinical cancer therapeutic development. *Br J Pharmacol.* **2013**, 169(4), 719-735.
2. Van Heertum, R.; Scarimbolo, R.; Ford, R.; Berdougou, E.; O'Neal, M. Companion diagnostics and molecular imaging-enhanced approaches for oncology clinical trials. *Drug Des Devel Ther.* **2015**, 9, 5215-5223.
3. James, M.L.; Gambhir, S.S. A Molecular Imaging Primer: Modalities, Imaging Agents, and Applications. *Physiol Rev.* **2012**, 92(2), 897-965.
4. DiMasi, J.A.; Hansen, R.W.; Grabowski, H.G. The price of innovation: new estimates of drug development costs. *J Health Econ.* **2003**, 22, 151–185.
5. Slomka, P.J.; Pan, T.; Berman, D.S.; Germano, G. Advances in SPECT and PET hardware. *Prog Cardiovasc Dis.* **2015**, 57, 566–578.
6. Shukla, A. K.; Kumar, U. Positron emission tomography: An overview. *Journal of Medical Physics / Association of Medical Physicists of India* **2006**, 31(1), 13–21.
7. Turkington, T.G. Introduction to PET Instrumentation. *J Nucl Med Technol* 2001, 29(1), 4-11.
8. Fahey, F.H. Data Acquisition in PET Imaging. *J Nucl Med Technol* **2002**, 30(2), 39-49.
9. Almuhaideb, A.; Papatnasiou, N.; Bomanji, J. ¹⁸F-FDG PET/CT Imaging In Oncology. *Annals of Saudi Medicine* **2011**, 31(1), 3–13.
10. Kelloff, G.J.; Hoffman, J.M.; Johnson, B. Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. *Clin Cancer Res.* **2005**, 11(8), 2785-2808.
11. Kayani, A.M.; Groves, ¹⁸F-fluorodeoxyglucose PET/CT in cancer imaging. *Clin.Med.* **2006**, 6, 240–244.
12. Kelloff, G.J.; Hoffman, J.M.; Johnson, B.; Scher, H.I.; Siegel, B.A.; Cheng, E.Y.; Cheson, B.D.; O'Shaughnessy, J.; Guyton, K.Z.; Mankoff, D.A.; Shankar, L.; Larson, S.M.; Sigman, C.C.; Schilsky, R.L.; Sullivan, D.C. Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. *Clin. Cancer Res.* **2005**, 11, 2785–2808.
13. Rohren, E.M.; Turkington, T.G.; Coleman, R.E. Clinical applications of PET in oncology. *Radiology* **2004**, 231, 305–332.
14. Ben-Haim S, Ell P. ¹⁸F-FDG PET and PET/CT in the Evaluation of Cancer Treatment Response. *J Nucl Med* . 2009;50(1):88-99.
15. Wáng, Y.-X. J.; Choi, Y.; Chen, Z.; Laurent, S.; Gibbs, S. L. Molecular Imaging: From Bench to Clinic. *BioMed Research International* **2014**, 357258.
16. Weissleder, R.; Mahmood, U. Molecular imaging. *Radiology* 2001; 219(2), 316–333.
17. Thakur, M.; Lentle, B.C. Report of a summit on molecular imaging. *Radiology* **2005**, 236(3), 753–755.
18. Cai, W.; Rao, J.; Gambhir, S.S.; Chen, X. How molecular imaging is speeding up antiangiogenic drug development. *Mol. Cancer Ther.* **2006**, 5(11), 2624–2633.
19. Massoud, T.F.; Gambhir, S.S. Molecular imaging in living subjects: seeing fundamental biological processes in a new light. *Genes Dev.* **2003**, 17(5), 545–580.
20. Chen, K.; Chen, X. Design and Development of Molecular Imaging Probes. *Current Topics in Medicinal Chemistry* **2010**, 10(12), 1227–1236.
21. Meng, Q.; Li, Z. Molecular Imaging Probes for Diagnosis and Therapy Evaluation of

- Breast Cancer. *Int J Biomed Imaging* **2013**, 2013, 1-14.
22. Phelps, M.E. Positron emission tomography provides molecular imaging of biological processes. *Proc Natl Acad Sci USA* **2000**, 97(16), 9226-9233.
 23. Acton, Q.A. *Advances in Molecular Nanotechnology Research and Application: 2011 Edition*. ScholarlyEditions; 2012.
 24. Weissleder, R. *Molecular Imaging: Principles and Practice*; PMPH-USA, 2010.
 25. Chen X. Multimodality imaging of tumor integrin alphavbeta3 expression. *Mini Rev. Med. Chem.* 2006; 6(2):227-234.
 26. Liu, S.; Edwards, D.S. Bifunctional chelators for therapeutic lanthanide radiopharmaceuticals. *Bioconjug. Chem.* **2001**, 12(1), 7-34.
 27. Hargreaves, R.J. The role of molecular imaging in drug discovery and development. *Clin. Pharmacol. Ther.* **2008**, 83, 349-353.
 28. Gross, S.; Piwnica-Worms, D. Molecular imaging strategies for drug discovery and development. *Curr. Opin. Chem. Biol.* **2006**, 10, 334-342.
 29. Ogawa, M.; Regino, C.A.; Seidel, J.; Green, M.V.; Xi, W.; Williams, M.; Kosaka, N.; Choyke, P.L.; Kobayashi, H. Dual-modality molecular imaging using antibodies labeled with activatable fluorescence and a radionuclide for specific and quantitative targeted cancer detection. *Bioconjugate Chem.* **2009**, 20, 2177-2184.
 30. Velikyan, I. Prospective of ⁶⁸Ga-Radiopharmaceutical Development. *Theranostics* **2014**, 4(1), 47-80.
 31. Velikyan, I. Positron emitting [⁶⁸Ga] Ga-based imaging agents: chemistry and diversity. *Med Chem.* **2011**; 7, 338-72.
 32. Baum, R.P.; Kulkarni, H.R. Theranostics: from molecular imaging using Ga-68 labeled tracers and PET/CT to personalized radionuclide therapy - the Bad Berka experience. *Theranostics.* **2012**; 2, 437-47.
 33. Larson, S.M.; Pentlow, K.S.; Volkow, N.D.; Wolf, A.P.; Finn, R.D.; Lambrecht, R.M.; Graham, M.C.; Di Resta, G.; Bendriem, B.; Daghighian, F.; Yeh, S.D.J.; Wang, G.J.; Cheung, N.V. PET scanning of iodine-124-3F8 as an approach to tumor dosimetry during treatment planning for radioimmunotherapy in a child with neuroblastoma. *J. Nucl. Med.* **1992**, 33, 2020-2023.
 34. Verel, I.; Visser, G.W.; Boerman, O.C.; van Eerd, J.E.; Finn, R.; Boellaard, R.; Vosjan, M.J.; Stigter-van Walsum, M.; Snow, G.B.; van Dongen, G.A. Long-lived positron emitters zirconium-89 and iodine-124 for scouting of therapeutic radioimmunoconjugates with PET. *Cancer Biother. Radiopharm.* **2003**; 18, 655-661.
 35. Roesch, F.; Filosofov, D.V. Production, radiochemical processing and quality evaluation of Ge-68 suitable for production of a ⁶⁸Ge/⁶⁸Ga generator. Int. Atomic Energy Agency 2010.
 36. Nanda, P. K.; Lane, S. R.; Retzliff, L. B.; Pandey, U. S.; Smith, C. J. Radiolabeled regulatory peptides for imaging and therapy. *Current Opinion in Endocrinology, Diabetes, and Obesity* **2010**, 17(1), 69-76.
 37. Anderson, C. J.; Wadas, T. J.; Wong, E. H.; Weisman, G. R. Cross-bridged Macrocyclic Chelators for Stable Complexation of Copper Radionuclides for PET Imaging. *The Quarterly Journal of Nuclear Medicine and Molecular Imaging: Official Publication of the Italian Association of Nuclear Medicine (AIMN) [and] the International Association of Radiopharmacology (IAR), [and] Section of the Society Of...* **2008**, 52(2), 185-192
 38. Wu, Y.; Zhang, X.; Xiong, Z.; Chen, Z.; Fisher, D.R.; Liu, S. MicroPET imaging of glioma integrin αvβ3 expression using ⁶⁴Cu-labeled tetrameric RGD peptide. *J Nucl Med.* **2005**,

- 46, 1707–18.
39. McQuade, P.; Miao, Y.; Yoo, J.; Quinn, T.P.; Welch, M.J.; Lewis, J.S. Imaging of Melanoma Using ^{64}Cu -and ^{86}Y -DOTA-ReCCMSH(Arg11), a Cyclized Peptide Analogue of a-MSH. *J Med Chem.* **2005**, *48*, 2985–92.
 40. Chen, X.; Hou, Y.; Tohme, M.; Park, R.; Khankaldyyan, V.; Gonzales-Gomez, I.; Pegylated Arg-Gly-Asp peptide: ^{64}Cu labeling and PET imaging of brain tumor $\alpha\text{v}\beta 3$ -integrin expression. *J Nucl Med.* **2004**; *45*, 1776–1783.
 41. Chen, X.; Liu, S.; Hou, Y.; Tohme, M.; Park, R.; Bading, J.R. MicroPET imaging of breast cancer αv -integrin expression with ^{64}Cu -labeled dimeric RGD peptides. *Mol Imaging Biol.* **2004**, *6*, 350–9.
 42. Chen, X.; Sievers, E.; Hou, Y.; Park, R.; Tohme, M.; Bart, R. Integrin $\alpha\text{v}\beta 3$ -targeted imaging of lung cancer. *Neoplasia* **2005**, *7*, 271–9.
 43. Li, W.P.; Lewis, J.S.; Kim, J.; Bugaj, J.E.; Johnson, M.A.; Erion, J.L. DOTA-D-Tyr1-Octreotate: a somatostatin analog for labeling with halogen and metal radionuclides for cancer imaging and therapy. *Bioconjug Chem.* **2002**, *13*, 721–28.
 44. Chen, H.; Niu, G.; Wu, H.; Chen, X. Clinical Application of Radiolabeled RGD Peptides for PET Imaging of Integrin $\alpha\text{v}\beta 3$. *Theranostics* **2016**, *6*(1), 78-92.
 45. Schiller, E.; Bergmann, R.; Wunderlich, G.; Andreeff, M.; Jacob, Pietzsch H-J. Ga-68- and Cu-64-Labeled NOTA-Albumin Conjugates for PET Sentinel Lymph Node Imaging. *ISRN Molecular Imaging* **2013**, 358.
 46. Chakravarty, R.; Hong, H.; Cai, W. Positron Emission Tomography Image-Guided Drug Delivery: Current Status and Future Perspectives. *Molecular Pharmaceutics* **2014**, *11*(11), 3777–3797.
 47. Liu, Q.; Pan, D.; Cheng, C.; Zhang, D.; Zhang, A. Development of a Novel PET Tracer [^{18}F] AIF-NOTA-C6 Targeting MMP2 for Tumor Imaging. *PloS one* **2015**, *10*(11), p. e0141668
 48. Wadas, T. J.; Wong, E. H.; Weisman, G. R.; Anderson, C. J. Coordinating Radiometals of Copper, Gallium, Indium, Yttrium and Zirconium for PET and SPECT Imaging of Disease. *Chemical Reviews* **2010**, *110*(5), 2858–2902.
 49. Chem, M.; Knight, J.C.; Wuest, F.R. Nuclear (PET / SPECT) and optical imaging probes targeting the CXCR4 chemokine receptor. *MedChemComm* **2012**, 1039-1053.
 50. Woodard, L. E.; Nimmagadda, S. *J. Nucl. Med.* **2011**, *52*, 1665–1669.
 51. Tamamura, H.; Hiramatsu, K.; Mizumoto, M.; Ueda, S.; Kusano, S. Enhancement of the T140-based pharmacophores leads to the development of more potent and bio-stable CXCR4 antagonists. *Org Biomol Chem.* **2003**, *1*(21), 3663-9.
 52. Jacobson, O.; Weiss, I.D.; Kiesewetter, D.O.; Farber, J.M.; Chen, X. PET of tumor CXCR4 expression with 4-18F-T140. *J Nucl Med.* **2010**, *51*(11), 1796-804.
 53. Rot, A. Contribution of Duffy antigen to chemokine function. *Cytokine Growth Factor Rev.* **2005**, *16*(6), 687-694.
 54. Boswell, C.A.; Sun, X.; Niu, W.; Weisman, G.R.; Wong, E. H.; Rheingold, A.L., Anderson, C. J. Comparative in Vivo Stability of Copper-64- Labeled Cross-Bridged and Conventional Tetraazamacrocyclic Complexes. *J. Med. Chem.* **2004**, *47* (6), 1465-1474
 55. Jacobson, O.; Weiss, I.D.; Szajek, L. P.; Niu, G.; Ma, Y.; Kiesewetter, D. O.; Peled, A.; Eden, H. S.; Farber, J.M.; Chen. X. Improvement of CXCR4 tracer specificity for PET imaging. *J. Controlled Release* **2012**, *157*, 216–223.
 56. Demmer, O.; Gourni, E.; Schumacher, U.; Kessler, H.; Wester, H.J. PET imaging of CXCR4 receptors in cancer by a new optimized ligand. *ChemMedChem* **2011**, *6*, 1789-91.
 57. Gourni, E.; Demmer, O.; Schottelius, M.; D'Alessandria, C.; Schulz, S.; Dijkgraaf, I. PET of CXCR4 expression by a ^{68}Ga -labeled highly specific targeted contrast agent. *J Nucl Med.* **2011**, *52*, 1803-10.

58. Jacobson, O.; Weiss, I.D.; Szajek, L.; Farber, J.M.; Kiesewetter, D.O. ^{64}Cu -AMD3100--a novel imaging agent for targeting chemokine receptor CXCR4. *Bioorg Med Chem* **2009**, *17*, 1486-93.
59. Weiss, I.D.; Jacobson, O.; Kiesewetter, D.O.; Jacobus, J.P.; Szajek, L.P.; Chen, X. Positron emission tomography imaging of tumors expressing the human chemokine receptor CXCR4 in mice with the use of ^{64}Cu -AMD3100. *Mol Imaging Biol.* **2011**, *14*, 106-14.
60. Nimmagadda, S.; Pullambhatla, M.; Stone, K.; Green, G.; Bhujwalla, Z.M.; Pomper, M.G. Molecular imaging of CXCR4 receptor expression in human cancer xenografts with [^{64}Cu]AMD3100 positron emission tomography. *Cancer Res.* **2010**, *70*, 3935-44.
61. De Silva, R.A.; Peyre, K.; Pullambhatla, M.; Fox, J.J.; Pomper, M.G.; Nimmagadda, S. Imaging CXCR4 expression in human cancer xenografts: evaluation of monocyclam ^{64}Cu -AMD3465. *J Nucl Med.* **2011**, *52*, 986-93.
62. Breeman, W.A.P.; de Blois, E.; Chan, H.; Konijnenberg, M.; Kwekkeboom, D.J.; Krenning, E.P. ^{68}Ga -labeled DOTA-peptides and ^{68}Ga -labeled radiopharmaceuticals for positron emission tomography: current status of research, clinical applications, and future perspectives. *Semin Nucl Med* **2011**, *41*, 314-321.
63. Wadas, T.J. Coordinating radiometals of copper, gallium, indium, yttrium, and zirconium for PET and SPECT imaging of disease. *Chem Rev.* **2010**, *110*(5), 2858-902.
64. Banerjee, S.R.; Pomper, M.G. Clinical applications of gallium-68. *Appl Radiat Isot* **2013**, *76*, 2-13.

4.5 Experimental Section

4.5.1 Chemistry

Materials. N^α-Fmoc-protected amino acids, 2-Cl-trtCl resin, Fmoc-Rink Amide-Am resin, O-benzotriazole-N,N,N',N'-tetramethyl-uroniumhexafluorophosphate (HBTU), N,N-diisopropylethylamine (DIEA), triisopropylsilane (TIS), trifluoroacetic acid (TFA), and piperidine were purchased from IRIS Biotech, N-hydroxybenzotriazole (HOBt), N,N-dimethylformamide (DMF), dichloromethane (DCM), and from Iris-Biotech GmbH (Marktredwitz, Germany). Peptide synthesis solvents, reagents, H₂O and CH₃CN for HPLC were reagent grade and were acquired from commercial sources (Sigma-Aldrich, Milano, Italy) and used as received unless otherwise noted. Peptides were purified by preparative HPLC (Shimadzu HPLC system) equipped with a C18-bounded preparative RP-HPLC column (Phenomenex Kinetex 21.2 mm × 150 mm 5 μm). Peptides were analyzed by analytical HPLC (Shimadzu Prominence HPLC system) equipped with a C18-bounded analytical RP-HPLC column (Phenomenex Kinetex, 4.6 mm × 250 mm 5 μm) using a gradient elution (10–90% acetonitrile in water (0.1% TFA) over 15 min; flow rate = 1.0 mL/min; diode array UV detector). Molecular weights of compounds were confirmed by ESI-mass spectrometry using an Agilent 6110 quadrupole LC/MS system.

General Procedure for the Synthesis of Peptides. 2-Cl-trtCl resin (62.0 mg, 1.60 mmol/g) was swollen in DMF dry over 0.5 h, and a solution of the first amino acid (Fmoc-*L*-Pen(trt)-OH: 92.1 mg, 0.15 mmol, 0.5 equiv; or Fmoc-*L*-Cys(Trt)-OH: 87.9 mg, 0.15 mmol, 0.5 equiv; or Fmoc-*L*-Lys(Alloc)-OH (68 mg, 0.15 mmol, 0.5 equiv; or Fmoc-*D*-Lys(Alloc)-OH: 68 mg, 0.15 mmol, 0.5 equiv) and DIPEA (26 μl, 1.5 eq) in DMF (2 ml) was added. The mixture was stirred for 10 min. An additional amount of DIPEA (34.8 μl, 0.2 mmol, 2 equiv) was added and the mixture was then shaken for 1 h. The residual chloride groups contained in the resin were capped by adding MeOH (200 μl) in DCM (2 ml) and stirring for 1 h to avoid eventually parallel synthesis of side products. Fmoc group removal was performed using 20 % piperidine in DMF (1 × 5 min and 1 × 25 min). The peptide resin was then washed with DCM (3 × 0.5 min) and DMF (3 × 0.5 min) and positive Kaiser ninhydrine³⁵ and TNBS³⁶ tests were observed. Fmoc-*L*-Cys(Trt)-OH (234.3 mg, 0.4 mmol, 4 equiv) or Fmoc-*L*-Lys(Alloc)-OH (181.0 mg, 0.4 mmol, 4 equiv) or Fmoc-*D*-Lys(Alloc)-OH (181.0 mg, 0.4 mmol, 4 equiv), Fmoc-*L*-6-Ahx-OH (141.4 mg, 0.4 mmol, 4 equiv) or Fmoc-*L*-Phe-OH (155 mg, 0.4 mmol, 4 equiv) or Fmoc-*L*-4-Abz-OH (144 mg, 0.4 mmol, 4 equiv) Fmoc-*L*-1-Nal-OH (175 mg, 0.4 mmol, 4 equiv) or Fmoc-*L*-2-Nal-OH (175 mg, 0.4 mmol, 4 equiv) or Fmoc-*L*-Cys(Trt)-OH (234.3 mg, 0.4 mmol, 4 equiv) or Fmoc-*D*-Cys(Trt)-OH (234.3 mg, 0.4 mmol, 4 equiv), Fmoc-*L*-

Ala-OH (124.5 mg, 0.4 mmol, 4 equiv) or Fmoc-*L*-Arg-OH (260 mg, 0.4 mmol, 4 equiv) were sequentially added sequentially to the resin bound H-first amino acid. Each coupling reaction was achieved using a 4-fold excess of amino acid with HBTU (151.7 mg, 0.4 mmol, 4 equiv) and HOBT (61.2 mg, 0.4 mmol, 4 equiv) in the presence of DIPEA (140 μ l, 0.8 mmol, 8 equiv) in DMF. Coupling reaction with DOTA, 1,4,7,10-Tetraazacyclododecane-1,4,7,10-tetraacetic acid/NOTA, 1,4,7-triazacyclononane-1,4,7-trisacetic acid were performed using aa (DOTA: 121.3 mg, 0.3 mmol, 3 equiv; or NOTA: 91 mg, 0.3 mmol, 3 equiv), PyAOP (156 mg, 0.3 mmol, 3 equiv) and HOAt (41 mg, 0.3 mmol, 3 equiv) in the presence of DIPEA (140 μ l, 0.8 mmol, 8 equiv) in DMF. Fmoc deprotections were accomplished with 20% piperidine in DMF solution (1 x 5 min, 1 x 25 min). Washings with DMF (3 x 0.5 min) and DCM (3 x 0.5 min) were performed through every coupling/deprotection step. Kaiser ninhydrine and TNBS tests were employed for monitoring the progress of peptide synthesis. For the peptides bearing an acetyl group at the N-terminus, after removing the last Fmoc group, the resin bound peptide was treated with Ac₂O (19 μ l, 0.2 mmol, 2 equiv) and DIPEA (35 μ l, 0.2 mmol, 2 equiv) in DCM (2 ml) and the mixture was shaken for 1 h. A negative Kaiser ninhydrine and TNBS tests were observed.

General Procedure for Peptide Oxidation and Purification. The peptide was released from the solid support and all the protecting groups cleaved, treating the resin with TFA/DCM/TIS (80/15/5, v/v/v) (3 ml solvent/0.1 mmol) for 2 h. The resin was then filtered off and the crude linear peptide was recovered by precipitation with chilled ether to give a powder. The crude deprotected peptide (0.1 mmol) was dissolved in H₂O/acetonitrile (1:1) and a solution containing 1.5 equiv. of NCS was added. The mixture was then mechanically shaken for 15 min at room temperature, lyophilized and subsequently analysed. Oxidized cyclic eptapeptides were synthesized in high purity, and no linear precursors were present. Final peptide purification was achieved by preparative RP-HPLC in 0.1% TFA with an ACN gradient (10–90% ACN in H₂O over 15 min, flow rate of 15 mL/min) on a Phenomenex Kinetex C18 column (21.2 mm \times 150 mm 5 μ m). Analytical RP-HPLC were performed in 0.1% TFA with an ACN gradient (10–90% ACN in H₂O over 20 min, flow rate of 1.0 mL/min) on a Phenomenex Kinetex C18 column (0.46 mm \times 150 mm 5 μ m).

31: Purity > 95%, t_R 15.3 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 20 min, flow rate of 1.0 mL/min); molecular formula: C₅₁H₇₅N₁₅O₁₂S₂; calculated mass: 1153.5; found: 1154.4 (M+H⁺), 576.6 (M+2H⁺)/2, 1176.4 (M+Na⁺).

32: Purity > 95%, t_R 14.9 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 20 min, flow rate of 1.0 mL/min); molecular formula: C₅₁H₇₅N₁₅O₁₂S₂; calculated mass: 1153.5; found: 1154.5 (M+H⁺), 576.7 (M+2H⁺)/2, 1176.2 (M+Na⁺).

33: Purity > 95%, t_R 15.2 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 20 min, flow rate of 1.0 mL/min); molecular formula: C₅₆H₈₄N₁₆O₁₃S₂; calculated mass: 1252.6; found: 1253.6 (M+H⁺), 626.8 (M+2H⁺)/2.

34: Purity > 95%, t_R 14.7 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 20 min, flow rate of 1.0 mL/min); molecular formula: C₅₃H₇₉N₁₅O₁₂S₂; calculated mass: 1181.5; found: 1182.5 (M+H⁺), 590.8 (M+2H⁺)/2.

35: Purity > 95%, t_R 14.8 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 20 min, flow rate of 1.0 mL/min); molecular formula: C₆₈H₈₅N₁₅O₁₂S₂; calculated mass: 1367.6; found: 1368.5 (M+H⁺), 684.8 (M+2H⁺)/2.

36: Purity > 95%, t_R 15.2 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 20 min, flow rate of 1.0 mL/min); molecular formula: C₆₈H₈₅N₁₅O₁₂S₂; calculated mass: 1367.6; found: 1368.5 (M+H⁺), 683.8 (M+2H⁺)/2.

37: Purity > 95%, t_R 14.7 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula: C₇₂H₁₀₅N₂₁O₁₇S₂; calculated mass: 1599.7; found: 1600.4 (M+H⁺), 800.8 (M+2H⁺)/2.

38: Purity > 95%, t_R 15.4 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula: C₆₈H₉₈N₂₀O₁₅S₂; calculated mass: 1498.7; found: 1499.4 (M+H⁺), 749.8 (M+2H⁺)/2.

General Procedure for Peptide ^{nat}Ga labelling. DOTA-AMBS-Ahx-Arg-Ala-[DCys-Arg-2Nal-His-Pen]-COOH or NOTA-AMBS-Ahx-Arg-Ala-[DCys-Arg-2Nal-His-Pen]-COOH were dissolved in 0.1 N NaOAc (pH 4.5), and an equimolar amount of Ga(NO₃)₃ was added. After heating to 90°C for 30 min, the reaction mixture was cooled to RT, and ^{nat}Ga-DOTA-AMBS-Ahx-Arg-Ala-[DCys-Arg-2Nal-His-Pen]-COOH or ^{nat}Ga- NOTA-AMBS-Ahx-Arg-Ala-[DCys-Arg-2Nal-His-

Pen]-COOH were cooled and immediately purified using preparative HPLC (14-35% acetonitrile (0.1%TFA) in water (0.1%TFA) within 15 min).

39: Purity > 95%, t_R 12.6 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula: C₇₂H₁₀₅GaN₂₁O₁₇S₂; calculated mass: 1668.7; found: 1669.4 (M+H⁺), 834.8 (M+2H⁺)/2.

40: Purity > 95%, t_R 12.6 min (analytical HPLC, 10–90% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 1.0 mL/min); molecular formula: C₆₈H₉₈GaN₂₀O₁₅S₂; calculated mass: 1567.6; found: 1568.4 (M+H⁺), 784.2 (M+2H⁺)/2.

4.5.2 Biological studies

Binding Assay. CXCR4 binding of peptides **31-36** was evaluated as previously described. Briefly, 5×10^5 CCRF–CEM, HT29 cells were preincubated with increasing peptide concentrations (0.01 μ M, 0.1 μ M, 1 μ M, 10 μ M) in the binding buffer (PBS 1 \times plus 0.2% BSA and 0.1% NaN₃) for 30 min at 37 °C, 5% CO₂ and then labeled for 30 min using anti-CXCR4 PE-antibody (FAB170P, clone 12G5, R&D Systems, Minneapolis, MN, USA). To evaluate the specific peptide binding to CXCR4, the experiments were also conducted in COLO205, human colon cancer cells, overexpressing CXCR3, using anti-CXCR3 FITC-antibody (R&D FAB160F clone 49801) and in MCF-7, human breast cancer cell line, overexpressing CXCR7, using anti-CXCR7 APC-antibody (R&D FAB4227A clone 11G8). The cells were counted through a FACS Canto II cytofluorometer (Becton Dickinson Immunocytometry Systems, Mountain View, CA, USA). Each data point is the average of three determinations.

Competition binding experiments for peptides **37-40** were performed using the Jurkat cell line. In brief, cells were resuspended in PBS/0.2 % BSA. A total of 200 μ L of the suspension containing 400,000 Jurkat cells were incubated with 25 μ L of the desired labeled peptide solution, (containing 3.1 kBq, approx. 0.1 nM) and 25 μ L of the tested peptides at concentrations of 10⁻¹¹ to 10⁻⁵ M. Nonspecific binding was determined in the presence of 1 μ M cyclo(-D-Tyr1-Arg2-Arg3-Nal4-Gly5). After shaking for 2 h at room temperature, the incubation was terminated by centrifugation at 1300 rpm for 5 min. Cell pellets were washed twice with cold PBS. Cell bound radioactivity was

determined by using a 1480 Wizard3 gamma- counter from Wallac (Turku, Finland)). Experiments were repeated 2-3 times in triplicates. IC₅₀ values of the compounds were calculated by nonlinear regression using GraphPad Prism (GraphPad Prism 4.0 Software, Inc., San Diego, CA, USA). Each data point is the average of three determinations.

PART II

--



Rescue of C-terminal Destabilized Proteins by Use of D-Amino Acids

ABSTRACT

OBJECTIVES

In eukaryotic cells, the ubiquitin–proteasome pathway acutely regulates the protein turnover. The covalent attachment of ubiquitin to lysine residues of targeted proteins is a signal for the recognition by the proteasome. However, it's been recently discovered that efficient proteasomal proteolysis requires an additional intrinsically disordered region at which the enzyme complex engages the substrate and initiates degradation. On the other hand, although biomolecules composed of mirror image amino acids are of particular interest for their non immunogenic properties, the reason of their unique biological stability remains relatively unexplored.

METHODS AND RESULTS

The main challenge with probing the stability of proteins containing non-natural functionalities is the delivery of such proteins into the cytosol. Inspired by nature, researchers exploited the ability of toxins to spread through the body and started to investigate the sophisticated bacterial A-B toxin-mediated delivery nanomachines. Working as a macromolecular syringe, the pathogen injects his toxins into the cytoplasm of host cells in a directly and selectively way. Thus, I recruited the disarmed version of Bacillus anthracis toxin, the versatile LF_N/PA platform, for efficiently deliver chemical entities such as mirror image polypeptides across the bilayer phospholipid membrane.

I used the catalytic domain of sortase A from *Staphylococcus aureus* (SrtA) as a “molecular stapler”: this enzyme-mediated tailoring has allowed for the facile and specific fixing of protein-peptide fusions, safeguarding their functionality. Further, a standardized and rigorous western blot workflow has been drawn up to provide highly sensitive, consistent, reproducible and as quantitative as possible data of the cytosolic delivered fraction of chimera proteins. Thus, I related the rate of proteasomal degradation of proteins containing mixed chirality peptides. It has been shown that *L*-sequences, acting as an unstructured region, were able to achieve lower steady state concentrations when compared to *D*-cargos, which rescued the protein from the proteasomal-mediated degradation, suggesting that stereospecific interactions are required for the accessibility to the narrow catalytic pore. It was also found that the C-terminus group plays a role in the degradation pathway: peptides presenting a –CONH₂ as a C-termini are less stable compared to those with the – COOH. Protein quantification via Western blotting allowed us to detect even this feature, while using an SRM-based quantification we wouldn't be able to ascertain such a small difference in the molecular weight of the targeted proteins.

CONCLUSIONS

With the goal of understanding complex cell biological processes in the detailed quantitative manner, Anthrax toxin system has been successfully leveraged to deliver mixed chirality peptides into the cytosol. The resulted information could be used to discover novel regulatory principles and to develop strategies to interfere with cellular processes therapeutically or to engineer cellular metabolism

Chapter 1

Introduction

1.1 Survival through destruction

Protein degradation is as essential to the cell as protein synthesis. Non-performing wrongly folded and damaged proteins, prone to aggregation, are removed from the cellular milieu in the context of stress response and to avoid toxicity.^{1,2} The concentrations of regulatory proteins are adjusted by degradation at the appropriate time switching on/off many cellular pathways. Both unwanted foreign and native proteins are simply scavenged and digested into bits that the cell can re-use to make new proteins. In this way, the proteasome is maintained healthy and the interior of the cell is kept neat and tidy. Besides, this function is particularly important under starvation conditions.

According to one line of current thinking, there are two major, fundamentally different, intracellular devices for the breakdown of proteins within the environment of eukaryotic cells: the lysosome and the proteasome.

Cells rely on lysosomes to recycle old and worn out proteins enclosed in or part of intracellular structures. The lysosome working like a cellular stomach, contains acid and several types of relatively non-selective digestive enzymes through which disassembles and removes the entire anatomical structure, such various membrane-enclosed vesicles and organelles, and its contents, including proteins.

The proteasome is the central catalytic unit of the ubiquitin proteasome system (UPS) (Figure 1.1):³ it is the other cell's trash processor that in a discriminatory way degrades cytosolic and nuclear protein junk allowing a complete shutdown of function of a selected protein molecule due to its irreversible proteolysis.⁴ The abundance, widespread distribution, evolutionary conservation, long half-life and broad proteolytic specificity of the proteasome underline its key regulatory role in protein degradation. As Rock and coworkers demonstrated, the block of its function produces the almost complete inhibition of protein degradation.⁵ Cell cycle control and apoptosis, signal

transduction and DNA repair, inflammatory and immunity responses regulation, organismal development, cell metabolism and protein quality control and garbage disposal are biological processes, which are crucially dependent on the precise function of this highly-efficient checking station.⁶

These pathways are highly interconnected, making failures in proteasomal degradation a source of severe disturbances of the cellular function network. Defects of components within the system leads to severe human diseases, among them neurological disorders, inflammatory processes and cancer.⁷⁻

13

The UPS pathway works via a multistep mechanism, within which peptide bond cleavage being only the last step in a complex program of substrate manipulation.¹⁴ First, most proteins are targeted to the proteasome by the covalent attachment of ubiquitin molecules. Then the proteasome particle recognizes the ubiquitin signal and controlling access to them, encapsulates the proteins inside its structure. Subsequently, the substrate is then unfolded and translocated to the proteolytic sites in an ATP-dependent reaction.

1.2 Ubiquitination machinery

The discovery of the ubiquitin proteasome pathway emerged from efforts to understand why intracellular proteolysis requires metabolic energy. Although the hydrolysis of the peptide bond is exergonic, and there is no thermodynamic reason to use energy, Simpson first detected that protein degradation is an energy-dependent process measuring the release of amino acids from intact cells.¹⁵ Key elements of the answer became clear in the early 1980's as a result of the pioneering biochemical studies of Aaron Ciechanover, Avram Hershko and Irwin Rose, who were awarded jointly the Nobel Prize in Chemistry 2004.¹⁶⁻¹⁸ Using the cell-free proteolytic system from reticulocyte these investigators found that an energy source, in the form of ATP, participates in an early phase of degradation and is needed to modify proteolytic substrates with the attachment of a chain consisting of several copies of ubiquitin molecules to the protein substrate, which leads the recognition by the 26S proteasome.¹⁹⁻²⁰

Ubiquitin is a 76-amino-acid polypeptide that is highly conserved in all eukaryotes (for instance, yeast and human ubiquitin differ at only 3 residues), and it was originally assumed to participate in the differentiation of lymphocytes.

Conjugation of ubiquitin to the protein substrate proceeds via several step cascade mechanism involving three distinct classes of enzymes:²¹ ubiquitin-activating enzymes (E1), ubiquitin-conjugating enzymes or ubiquitin-carrier proteins (E2) and ubiquitin ligases (E3) (Figure 1.2).^{22,23} Those enzymes are organized in an hierarchial order. In humans, two E1 enzymes specifically recognize and activate ubiquitin and interact with a small number (~30) E2 enzymes which then face towards a huge and as yet unknown number of E3 enzymes.²⁴

Initially, ubiquitin is activated in its C-terminal Gly76 in a reaction catalyzed by the enzyme E1. In detail, the initial Ubl-adenylate intermediate reacts with the E1 active site cysteine to form an E1~Ubl thiol ester in an ATP-dependently manner.

Following activation, ubiquitin is transferred from the E1 cysteine to the E2 cysteine enzymes in a trans thioesterification reaction and then to a member of the E3 enzyme family, to which the substrate protein is specifically bound. This enzyme catalyzes the last step in the conjugation process, the covalent attachment of the COOH C-terminal Gly residue of ubiquitin to a ϵ -NH₂ group of a Lys residue (or N-terminus) of the target protein to generate an isopeptide bond.

There are three classes of E3 ligases. RING E3s serve more as a bridge binding to both E2~Ub thioester and substrate and catalyzing attack of the substrate lysine on the thioester without the formation of an intermediate

The RING domain is similar to the zinc finger in that Cys and/or His residues coordinate two zinc ions and appears to function as a protein-protein interaction domain.²⁵⁻²⁸ On the other hand, HECT and RBR E3s both have active site cysteines and catalyze substrate ubiquitination in a two-step reaction involving formation of an another high-energy thioester bond between the HECT or RBR E3 followed by attack of the substrate lysine (or N-terminus) on the E3~Ub thioester to form the isopeptide linkage.^{29,33}

In successive reactions, a polyubiquitin chain is synthesized by processive transfer of additional activated mono-ubiquitin moieties through their Gly to Lys of the previously conjugated ubiquitin molecule. Ub has seven lysines, making several 'mixed' topologies of a poly-Ub chain possible.

The polymerized ubiquitin chain acts as a signal that shuttles the target proteins to the proteasome, where the substrate is proteolytically broken down.

The first poly-Ub chain that was discovered had its Ub moieties conjugated through Lys48,³⁴ which turned out to be the most common linkage for targeting substrates for degradation by the 26S proteasome. More recently, it has become evident that protein modification by ubiquitin also has unconventional (non-degradative) functions such as the regulation of DNA repair and endocytosis.

These non-traditional functions are dictated by the number of ubiquitin units attached to proteins (mono versus poly-ubiquitination) and also by the type of ubiquitin chain linkage that is present, involving Lys63 and Lys69.^{22,35} The key to specific recognition of Lys48 linked Ub chains probably lays in the three-dimensional conformation of the polyubiquitin structure. This is supported by the finding that repeated patches in the chain formed by three hydrophobic residues on the surface of Ub, Leu8, Ile44 and Val70, are essential for the recognition of Lys48 linked chains by the proteasome.³⁶ The conformation of Lys48 linked (Ub)₄, which is the minimal signal for efficient targeting to the proteasome,³⁷ gives better access to the hydrophobic patches of the Ub units than the (Ub)₂ conformation.³⁸

1.3 The proteasome: a complex of two halves

In eukaryotes, the proteasome particle constitutes *ca.* 1% of the cellular protein pool³⁹ and is equally distributed all over the cyto- and nucleoplasm as demonstrated by *in vivo* localization studies on tissue culture cells using GFP-labeling techniques.⁴⁰ Also, in mammalian organisms proteasomal degradation of regulatory proteins such as p27, p53 and I κ B was found to depend on nuclear export suggesting a general principle for proteasomal turnover of short-lived nuclear proteins.⁴¹ However, its concentration varies considerably among cell types and is greater in organs, including liver, in which average rates of protein breakdown are higher than in other tissues, such muscle.

The 26S protease complex is functionally and structurally divided into two parts. The proteolytic active sites of the proteasome are housed in a hollow cylindrical chamber, the 20S core (core particle, CP),⁴²⁻⁵¹ which has a 19S regulator complex at either end (regulatory particle, RP or PA700).⁵²⁻⁶⁰

1.3.1 The proteolytic core particle

Early electron microscopic (EM) studies of the 730-kDa catalytic core of this enzymatic assembly, purified from different tissues and species, revealed its barrel-shape built from four axially stacked heteroheptameric rings, that collectively form a cylinder 14.8 nm in length and 11.3 nm in diameter.⁶¹ Such “chambered” protease complexes of different composition have been found in many organisms including bacteria.⁶²⁻⁶⁵ The ancestor of the eukaryotic 20S proteasome is the eubacterial and archaeobacterial proteasome.

Atomic-resolution crystal structures of the 20S proteasome showed that the subunits of the core particle are arranged into four seven-subunit rings: two rings consisting of related α -subunits and two consisting of related β -subunits (Figure 1.3).⁶⁶ The rings are stacked on top of each other to form a compact cylinder with two β -subunit rings in the center that are topped at each end by a ring of α -subunits.⁶⁷⁻⁷⁰

The two outer sections contain evolutionarily related, but non identical α -subunits ($\alpha 1$ - $\alpha 7$), that with their N-terminal tails form the ‘gates’, the α -annulus) through which substrates enter and products are released.⁷¹ It projects across and orthogonal to the pseudo seven-fold axis and forms contacts with every other subunit. The deletion of the nine residue tail from the N-terminus of $\alpha 3$ of the yeast proteasome results in the $\alpha 3\Delta N$ mutant (Figure 1.4)⁷² that has an axial channel through the molecule whose dimensions are comparable to those of the *Thermoplasma* CP channels.⁷³ Crystal structure analysis of this mutant highlights the loss of defined electron density arising from both the elimination of the N-terminal residues from $\alpha 3$ and the disorder in the tails of the other subunits. In particular, $\alpha 3$ is unperturbed within the mutant particle, indicating that the deletion does not cause an assembly defect.⁷⁴ Moreover, the CP of $\alpha 3\Delta N$ shows strongly enhanced peptidase activity using three different fluorogenic peptide substrates, each specific to a different active site, probably because the access is facilitated (Figure 1.4 e-f).

Antechambers are developed jointly by one α and one β subunit, connecting the two outer cavities with the central nucleus, which can only be reached by two axial pores of about 2 nm.^{75,71} The inner section ($\beta 1$ – $\beta 7$) harbor three distinct catalytically active subunits — $\beta 1$, $\beta 2$ and $\beta 5$ — in duplicate, giving a total of six sites that are responsible for the proteolytic action.

Therefore, the proteasome architecture occludes its catalytic site within the lumen of the cylinder and this arrangement prevents well-folded proteins from entering the constricted annulus, protecting them from degradation.

The formation of the active site undergoes a two-step mechanism (Figure 1.5).⁷⁶ First, β s initially possess an N-terminal pro-sequence which is cleaved off by neighboring sites during 20S particle formation. The autocatalytic second step leads to exposing a terminal Threonine residue that is necessary for activity, positioning it just at the open cleft between the two layers of β -sheets. As predicted from sequence comparisons,⁷⁷ site-directed mutagenesis and the crystal structure analysis of a proteasome in complex with the small peptide aldehyde inhibitor Ac-LLnL-al (commercially available as calpain inhibitor I) identified the amino-terminal threonine (Thr1) as both the catalytic nucleophile and the primary proton acceptor,^{73,77} excluding the role of serine or cysteine.

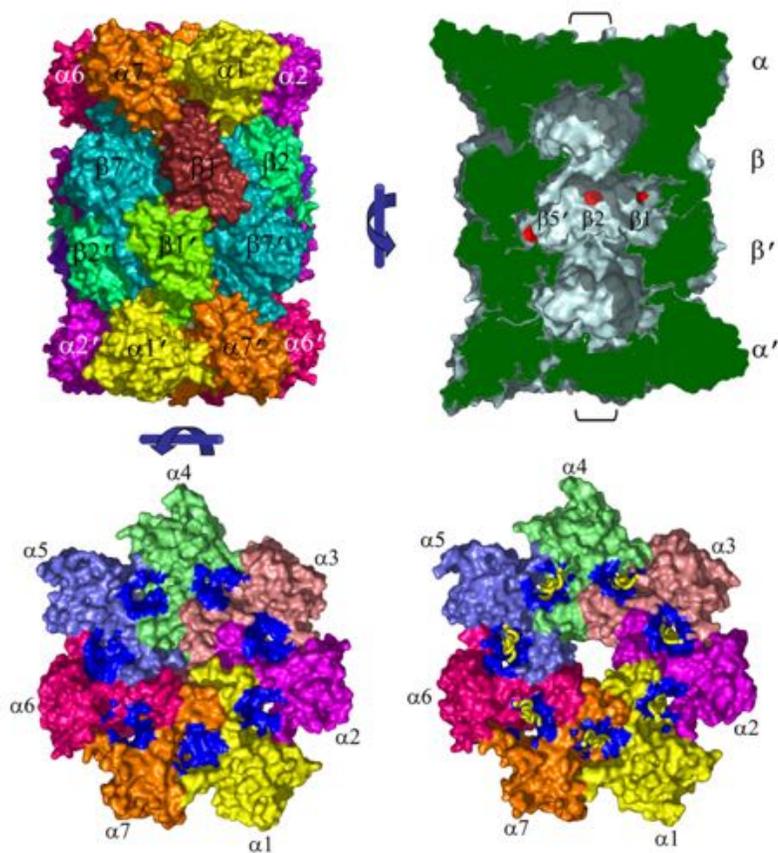


Figure 1.3. Side and top views of a yeast CP provide a perspective on the basic CP structure.

Moreover, Thr1 replacement by a residue of Ser in proteasomes from the archaeon *Thermoplasma acidophilum* (T1S mutant) does not alter the hydrolysis rate of small fluorogenic peptide substrates, but the mutant is significantly slower in degrading larger peptides and proteins. In addition, it was shown that deletion of the N-terminal Thr1 or mutation to Ala, used to study substrate specificity and the hierarchy of the proteasome active sites, leads to inactivation of the enzyme and the mutants became deficient in autolysis. Thus, N-terminal threonine functions as a single-residue active site characterizing the proteasome as a member of the family of amino-terminal nucleophile (NTN) hydrolases.⁷⁸

The hydroxyl group O_γ of Thr1 as nucleophile attacks the carbonyl carbon of the scissile peptide bond, while its amino-terminal group serves as the primary proton acceptor enhancing the

nucleophilicity by stripping the proton from the side-chain hydroxyl; for steric reasons, a water molecule is likely to mediate the proton transfer.

Aside from the N-terminal Thr1, structural and mutational studies on proteasomes from various species identified the highly conserved residues Asp17 (Glu17 in the *Thermoplasma* subunit) and Lys33. In the catalytic center, Lys33 is most probably charged at neutral pH and therefore unsuitable to function as the proton acceptor. Hence, it is expected that the role of Lys33-Nε is to lower the pKa of the N-terminal amino group of Thr1 by its electrostatic potential, so that this group can act as the proton acceptor in proteolysis. Thr'Oy-H ... N may form a five-membered ring structure in favorable hydrogen bonding geometry.^{47,79} Surprisingly, the conservative exchange of Lys33 to arginine in subunit 5 of the yeast CP abolishes proteolysis against chromogenic substrates. Other conserved residues in the vicinity of the active site Thr1 also appear to be required for structural integrity of the proteolytic site. The atoms Ser129-Oγ, Asp166-O, and Ser169-Oγ stabilize the conformation of Thr1 via hydrogen bonds.

Although N-terminal residue of the macromolecular complex is always a threonine for the nucleophilic attack on peptide bonds, its substrate specificity is determined by broad and complementary substrate-binding channels allowing to degrade many different sequences. Each of the three mature β-subunits has indeed a different preference for cleaving after hydrophobic, basic or acidic residues. Accordingly, their cleavage-site specificity can be roughly characterized as chymotrypsin-like (CL, catalyzed by subunit 5), trypsin-like (TL, by 2), and post-glutamylpeptide hydrolysing or caspase-like (PG, by 1).⁸⁰

The three catalytic sites interact allosterically. For example, inhibitors of the caspase-like activity stimulate the trypsin-like activity but do not affect the chymotrypsin-like activity. In addition, substrates of the caspase-like activity allosterically inhibit the chymotrypsin-like. Moreover, the eucarya enzyme also catalyzes cleavage after branched-chain amino acids (BrAAP activity) and small neutral amino acids (SNAAP activity).⁸¹

1.3.2 The 19S Regulatory Particle

The 19 S regulatory particle is a ~930 kDa complex composed of about 17 subunits and two catalytic activities: ATP hydrolysis and a specialized proteolytic cleavage that removes ubiquitin chains from substrate proteins.⁸²

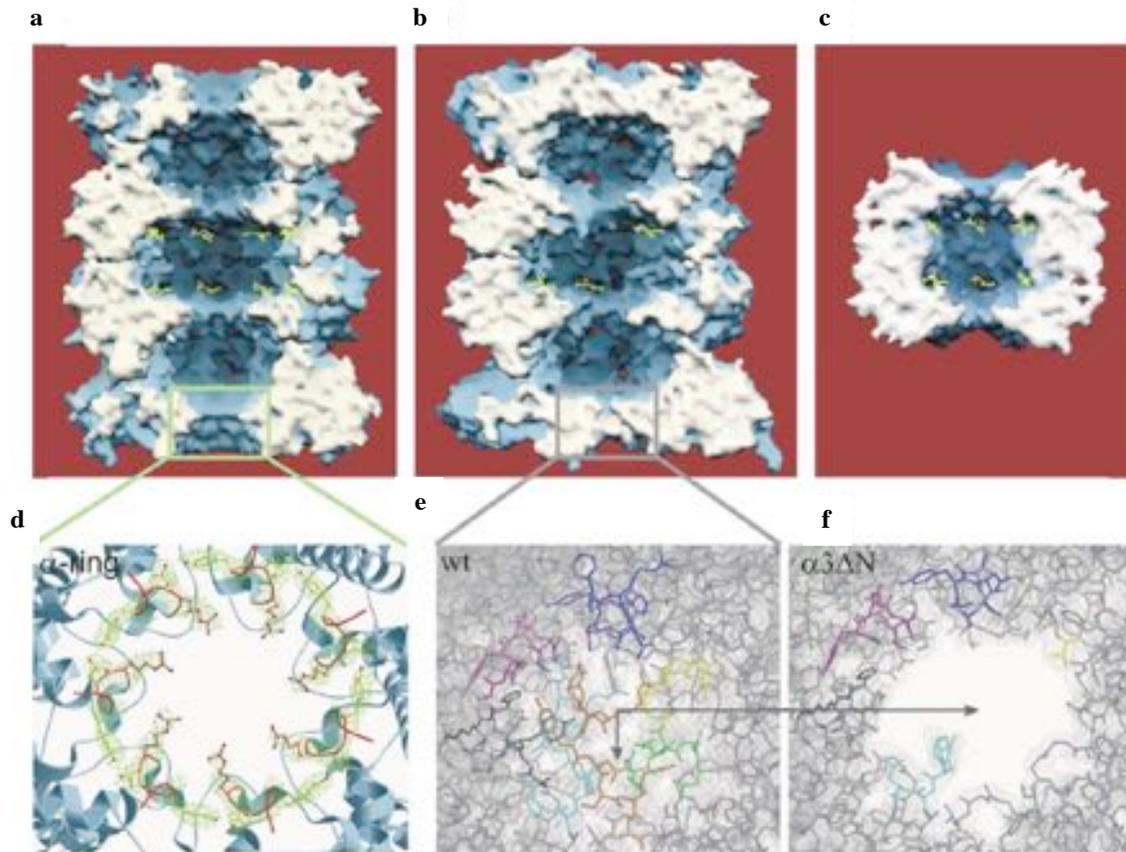


Figure 1.4. Surface view of the a) archaeal; b) eukaryotic 20S proteasome; and c) hslV molecule in complex with calpain-inhibitor I. The inhibitor molecules are shown as yellow balls-and-sticks and mark the proteolytic active sites. The disorder of the first N-terminal residues in the archaeal-subunits creates a channel in the crystal structures of these CPs (a), whereas the irregular but well defined arrangement of the N-terminal α -subunits seals the chamber in eukaryotic CPs (b). d) Ribbon diagram of the crystal structure of the free α -ring from *A. fulgidus* with its defined N-terminus in focus (colored in red). Tyr8 of each N-terminal part makes hydrogen bonds to Asp9 of the adjacent α -subunit (yellow balls-and-sticks), Arg10 (red balls-and-sticks) points toward the channel, generating a 13Å entrance. (e, f) Electron density maps (colored in gray) of the yeast core particle (gray sticks) from (e) wild type and (f) $\alpha 3\Delta N$. (Adapted from figure in Groll M, *Nature* 1997)

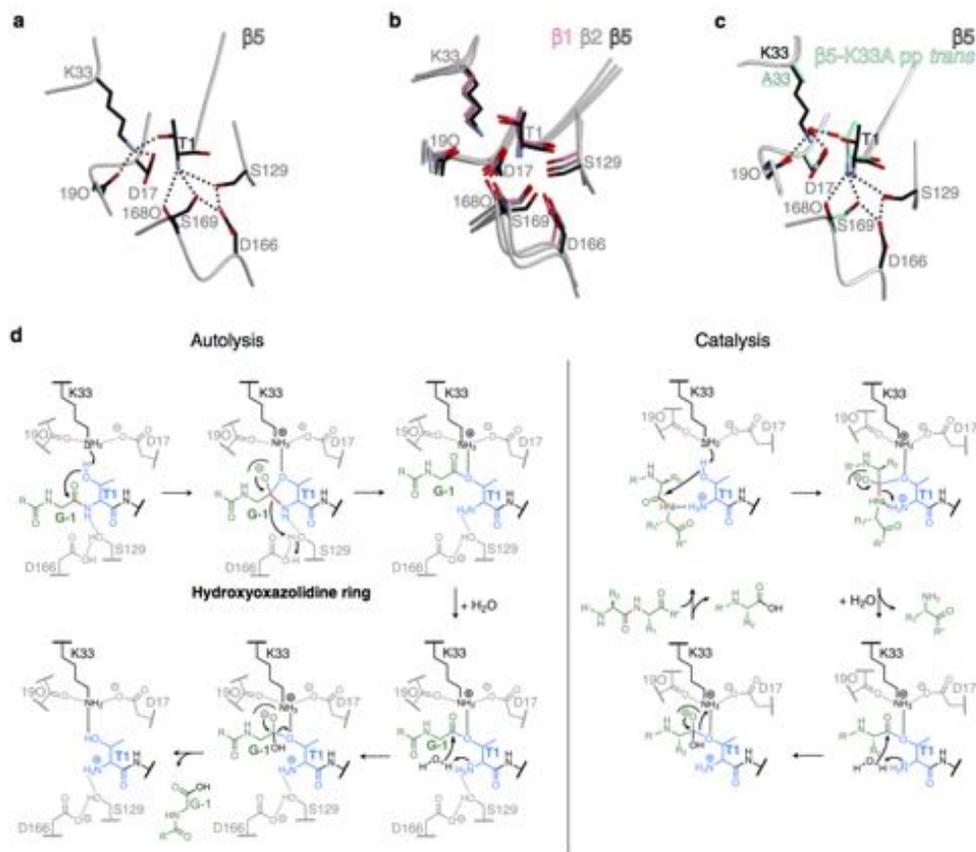


Figure 1.5. Architecture and proposed reaction mechanism of the proteasomal active site. a) Hydrogen-bonding network at the mature WT b5 proteasomal active site (dotted lines). Thr1OH is hydrogen-bonded to Lys33NH₂ (2.7 Å), which in turn interacts with Asp17O. The Thr1 N terminus is engaged in hydrogen bonds with Ser129O_γ, the carbonyl oxygen of residue 168, Ser169O_γ and Asp166O_δ. b) The orientations of the active-site residues involved in hydrogen bonding are strictly conserved in each proteolytic centre, as shown by superposition of the b subunits. c) Structural superposition of the WT b5 and the b5-K33A pp trans mutant active site. In the latter, a water molecule (red sphere) is found at the position where in the WT structure the side chain amine group of Lys33 is located. Similarly to Lys33, the water molecule hydrogen bonds to Arg190, Asp17O_δ and Thr1OH. Note, the strong interaction with the water molecule causes a minor shift of Thr1, while all other active-site residues remain in place. d) Proposed chemical reaction mechanism for autocatalytic precursor processing and proteolysis in the proteasome. (Adapted from figure in Lfwe J, *Science* 1995)

The 19S cap can be split into two different subcomplexes, base and lid (Figure 1.6).⁸³ The lid is composed of nine different non-ATPase subunits (Rpn: Regulatory Particle Non-ATPase, Rpn3, Rpn5–Rpn9, Rpn11, Rpn12 and Rpn15) (Figure 1.7a).⁸⁴ The subunits Rpn3, Rpn5, Rpn6, Rpn7, Rpn9 and Rpn12 contain a C-terminal PCI (Proteasome–CSN–eIF3) domain that interacts with each other to form a horseshoe-shaped anchor from which the aminoterminal domains extend radially.^{83,85,86} This arrangement is assumed to have scaffolding functions and allow inter-subunit contacts. Rpn11 and its inactive counterpart Rpn8 contain a Jab1/Mpn domain that allows them to dimerize. Rpn11 metalloproteinase is located at the entrance of the translocation channel and works as a deubiquitylating enzyme. This DUB has been shown to remove entire ubiquitin chains from the substrate by cleaving the isopeptide bond of the proximal ubiquitin moiety and thus recycle the ubiquitins; this chain is further cleaved into monomeric ubiquitins by other DUBs. In addition, in mammalian cells, two other DUBs that are physically associated with the base complex cleave the ubiquitin moiety at a distal site. Usp14 (yeast UBP6) is associated with Rpn1 and Uch37 binds to the C-terminal domain of Rpn2-bound Rpn13. Notably, Rpn11's deubiquitination activity was found to depend on the action of ATPases, suggesting a potential coupling with substrate translocation. Thus, these properties prevent premature deubiquitination of the substrate before engagement by the ATPase ring.^{87,88} However, the mechanism for this coupling remains unknown. The lid might be partly anchored to the base by Rpn10, because deletion of this subunit in yeast cells results in lability of the 19S RP.

Moreover, Rpn6 seems to function as a molecular clamp holding the core and regulatory subcomplexes together.⁸⁹

The second subcomplex of the proteasome, called the base, has three functional roles: capturing client proteins via ubiquitin recognition, promoting substrate unfolding and opening the channel in the α -ring.⁹⁰

The base consists of six proteins with ATP-hydrolysing activity, (Figure 7b)⁸³ of the AAA-type designated Rpt1–Rpt6 (ATPase associated with various cellular activities; Rpt, Regulatory Particle Triple A protein). Those subunits are essential for the survival of yeast cells and their C-termini form a hexameric ring-shaped motor that dock onto the outer α -rings on the both ends of the digestive cylinder. The ring is not planar but has a spiral staircase-like topology, with Rpt3 at the highest and Rpt2 at the lowest position. The heterohexamer forms a trimer of dimers with the pairs Rpt1/Rpt2, Rpt6/Rpt3, and Rpt4/Rpt5 each held together by coiled coils formed by near N-terminal regions of these subunits. The coiled coils protrude

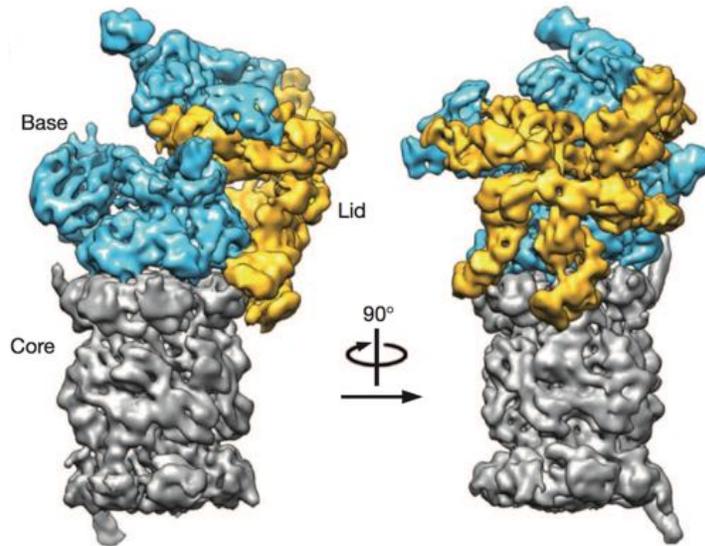


Figure 1.6. Architecture of the proteasomal 19S.
(Adapted from figure in Lander G, *Nature* 2012)

from a ring formed by oligosaccharide binding (OB) domains of the six subunits (N-ring). The latter is positioned atop the hexameric ring formed by the AAA-ATPase domains (AAA- ring), but the two rings are not in register.⁹¹

Four additional components of the base are non-ATPase subunits, Rpn1 and Rpn2, the largest proteins of the 19S RP, and two established ubiquitin receptors (Rpn10 and Rpn13). Rpn1 and Rpn2 act as scaffolds that assist in assembly of the base subcomplex and crucially contact other proteasome subunits. Rpn1 docks to the ATPase subunits Rpts 1 and 2, and is known to interact with a number of proteasome factors including ubiquitin shuttle receptors and the deubiquitinase Ubp6, which use their ubiquitin-like (Ubl) domain to dock to Rpn1.^{92,93} Rpn2 binds to the N-terminal coiled-coil of Rpts 3 and 6 and is positioned high above the unfoldase pore. Rpn2 binds to ubiquitin receptor Rpn13, and also appears to contact the critical deubiquitinase Rpn11.

Rpn10 and Rpn13 function as integral ubiquitin receptors and efficiently trap polyubiquitylated substrates. Rpn10 achieves this function via a C- terminal ubiquitin-interacting motif (UIM). More recently, Rpn13 was identified as a second ubiquitin receptor. The N-terminal domain of Rpn13 shows no similarity to known ubiquitin-binding motifs, but instead, contains the novel ‘pleckstrin-like receptor for ubiquitin’ (Pru) domain. The Pru domain in human Rpn13 shows a high affinity for diubiquitin.

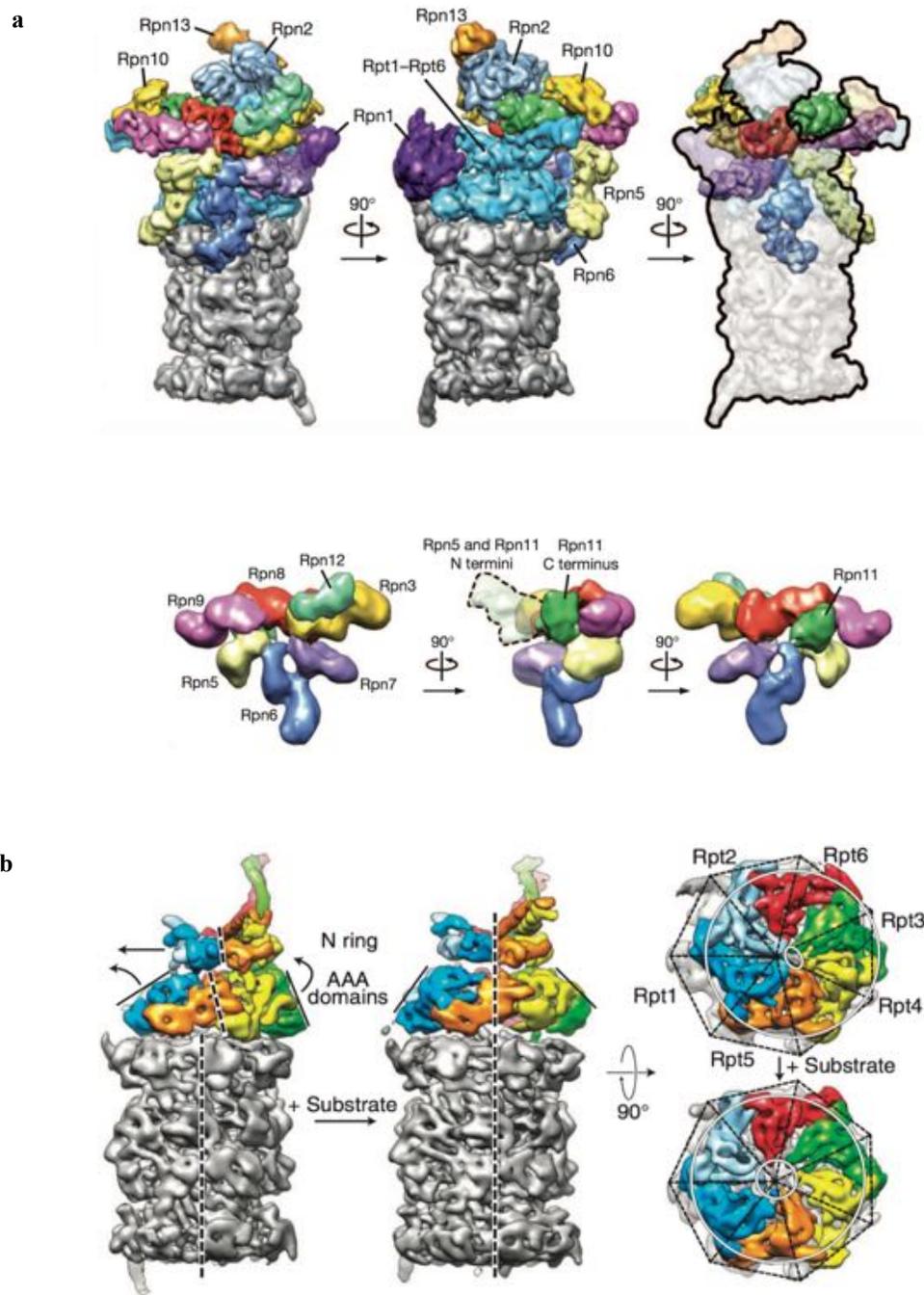


Figure 1.7. The subnanometer-resolution structure of the 26S proteasome. Three dimensional reconstruction of the base and the lid subcomplexes. (Adapted from figure in Lander G, *Nature* 2012)

Matyskiela and coworkers provide the first insights into the structure of the actively translocating 26S proteasome and outline the transitions that accompany substrate engagement. In this preengaged state, ubiquitin chain of an incoming substrate is tethered to a proteasomal receptor and the entrance to the N ring is accessible to its unstructured initiation region.

However, the central pore of the ATPase ring is constricted and not coaxially aligned with the subjacent peptidase. Furthermore, the DUB Rpn11 active site is occluded, and this prevents premature deubiquitination of the substrate before engagement by the ATPase ring. In this state, the AAA+ domains are arranged in a pronounced spiral staircase. Substrate interactions with Rpt subunits at the top of this spiral trigger the switching of the regulatory particle into a translocation-competent conformation that is characterized by a reorganized AAA+ ring with an alternative spiral arrangement, more uniform AAA-domain interfaces and a continuous central channel to the peptidase (Figure 1.8).⁸⁴ Rpn11 shifts to a central location directly above the N-ring pore, where its active site is accessible and ideally positioned to scan translocating poly-peptides for ubiquitin chains and ensure complete deubiquitination. This substrate-engaged conformation of the regulatory particle is stabilized by an alternative set of lid-base interactions.

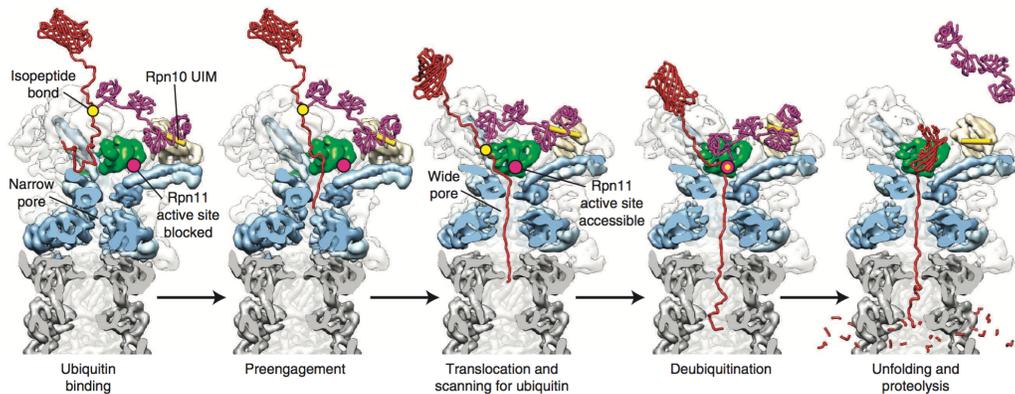


Figure 1.8. Structure-based model for substrate engagement and degradation by the 26S proteasome.
(Adapted from figure in Matyskiela ME, *Nat Struct Mol Biol* 2012)

1.4 Defining the geometry of the two-component proteasome degron

A fundamental question about intracellular proteolysis is how specific proteins are recognized by the proteolytic machinery, resulting in proteins being degraded only under specific conditions with highly characteristic degradation rates. Native substrates carry recognition tags that, when present, deliver proteins for degradation. The tags confer specificity, distinguishing substrates from other proteins. Early work had suggested that global structural features determine the metabolic stability of individual proteins. For instance, mutant proteins or proteins that had incorporated amino acid analogues during their synthesis were found to have shorter half-lives *in vivo* than their wild-type counterparts.^{94,95}

Indeed, most short-lived proteins are distinguished by localized structure determinants ('signals') that target them to the ubiquitin ligase machinery and then to the proteasome. A degradation signal or 'degron',⁹⁶ is usually defined as a minimal element within a protein that is sufficient for recognition and degradation by a proteolytic apparatus.

Different physiological requirements for the degradation of specific proteins dictate the design of different degrons. Distinct determinants comprise a degron, and they have different roles in the degradation pathway. Specifically, primary recognition determinants are those sequences or structures within the degron that bind directly to the E3-E2 ubiquitin-ligase complex. Another determinant in ubiquitin-dependent degrons is the presence of an appropriate acceptor site(s) for attachment of the polyubiquitin chain, such as a lysine residue. Finally, the degron must be situated within the substrate such that the protein is then threaded into the proteolytic chamber where it is hydrolyzed into short peptides.^{97,98}

1.5 N-degrons and the N-end rule pathway

The notion that covalent ubiquitin conjugation commits proteins for degradation led the discoverers of ubiquitin-mediated proteolysis to propose that substrate selection takes place mainly at the stage of ubiquitin ligation.^{99,100} By adding a variety of proteins to a rabbit reticulocyte lysate, Hershko and colleagues noted an apparent correlation between the presence of a free α -amino group in the proteins and their ubiquitin-dependent degradation.¹⁰¹ They subsequently isolated a 180 kD protein with E3 ubiquitin ligase activity that appeared to have a higher affinity for proteins with a free α -amino group than those with a blocked N-terminus.¹⁰⁰

This particular E3 distinguishes proteins not only by a free N-terminal α -amino group but also the side chain of the N-terminal residue. Varshavsky and co-workers systematically changed the N-terminal residue in an otherwise identical series of *Escherichia coli* β -galactosidase test substrates and expressed them in yeast, where they displayed a remarkable range of degradation rates.¹⁰² Half-lives ranged from a few minutes to greater than 20 hours. Thus, an E3 is able to bind protein substrates with very high selectivity, in this case being able to distinguish substrates by recognizing a specific residue at the N-terminus of a protein. This degradation pathway, termed the ‘N-end rule pathway,’ states that the half-life of a protein is determined by the nature of its N-terminal residue. Peptide sequences within the N-terminal region of the substrate that are sufficient for ubiquitin-dependent turnover constitute the ‘N-degron’.

1.6 A second component of the proteasome targeting code

The polyubiquitin chain mediates the productive binding of the substrate to the proteasome.^{103,104} During degradation, the ubiquitin moieties are removed by deubiquitinating activities while the substrate remains bound to the proteasome.^{104,105} Although ubiquitination of the substrate leads to recognition by the proteasome, it does not ensure rapid degradation of all proteins.^{104,106}

Fishbain et al. showed that Rad23 binds the proteasome and ubiquitinated proteins, which are subsequently degraded, but Rad23 itself escapes to shuttle the next round of substrates for degradation. The introduction of unstructured linkers into various regions of Rad23 serves as artificial initiation regions led to Rad23’s rapid proteolysis *in vitro* (Figure 1.9)¹⁰⁷ Thus, he demonstrated that Rad23 lacks a critical part of the degron.

Matouschek and co-workers reached a conclusion: an effective substrate requires not only a means of docking to the proteasome, but that proteins must also an unstructured region that acts as a site for initiation of unfolding.¹⁰⁸

The unstructured regions work as a degradation initiation site and proteolysis continues from there along the polypeptide chain. Thus, ubiquitin tagging allows the protease to recognize its substrate proteins, and degradation then begins with proteolysis of initiation site. The unstructured region functions to engage the unfolding machinery of the proteasome and it is indispensable for the degradation of folded proteins. Thus, protein targeting to the proteasome appears to have two components: an ubiquitination signal and an initiation site. By assaying degradation of a set of ubiquitin-modified DHFR variants they investigate how the initiation site contributes to proteasome

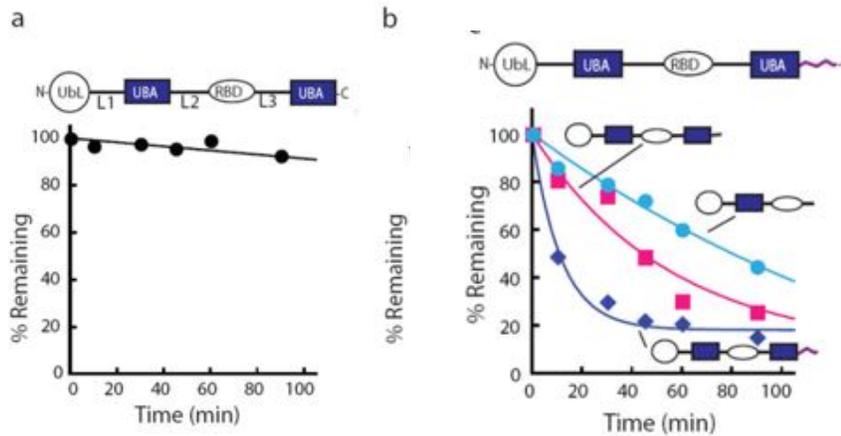


Figure 1.9. Rad23 escapes proteasome degradation *in vitro* because it lacks an initiation region. a) Degradation kinetics of Rad23. Incubation of Rad23 with purified yeast proteasome did not lead to its degradation at 30 °C. b) Adding an initiation region to Rad23 led to its rapid degradation. (Adapted from figure in Fishbain S, *Nat Commun* 2011)

targeting. They find that proteolysis of tightly folded proteins is accelerated greatly when an unstructured region is attached to the substrate. A chain of four ubiquitin moieties linked through Lys48 is a sufficient targeting signal for degradation by the proteasome.¹⁰⁴ However, attachment of tetraubiquitin to an ubiquitin-DHFR fusion protein led to notably slow degradation of this protein in an *in vitro* assay. Indeed, N-Ub4-DHFR-C lacks any unstructured region and was degraded very inefficiently. Instead, degradation was accelerated several-fold when an unstructured region was added to the C terminus of the protein, N-Ub4-DHFR~~~~-C and N-Ub4-DHFR^^-C. (Figure 1.10a).¹⁰⁸

Prakash evaluated also the eventually variation of the position of the ubiquitination site mediated by the addition of the unstructured region.¹⁰⁸ It has been shown that ubiquitination of the substrate with the C-terminal unstructured region still occurred at the N-terminal ubiquitination signal. Instead, the protein was stabilized when the N-end rule tag was inactivated by the mutation of the N-end residue from arginine to methionine27, by mutating the ubiquitin acceptor lysines, Lys15 and Lys17, in the linker region to arginine27, or by deleting the N-terminal ubiquitination site entirely (Figure 1.10b).¹⁰⁸ Taken together, these results underline the fact that rapidly degraded proteasome substrates must contain a degradation signal that has two parts, a proteasome-binding tag and a proteasome initiation region.

1.7 Defining the geometry of the two component proteasome degron

The initiation region is reminiscent of the linear targeting signals found in substrates of the archaeal and bacterial analogues of the proteasome.¹⁰⁹ Bacterial AAA+ proteases recognize their linear degrons through loops that line the pore at the center of the ring of ATPase subunits and it seems likely that the proteasome recognizes its initiation sites similarly.¹¹⁰ In the proteasome, the diameter of the pore is too narrow to allow folded proteins to pass through it so that a disordered polypeptide tail would have to be a certain length *g* to be able to reach the ATPase loops. This length requirement agrees roughly with the results of *in vitro* degradation experiments with model proteasome substrates, where proteins become degraded rapidly by purified yeast proteasome once they contain an unstructured tail of approximately 30 amino acids in length.¹¹¹

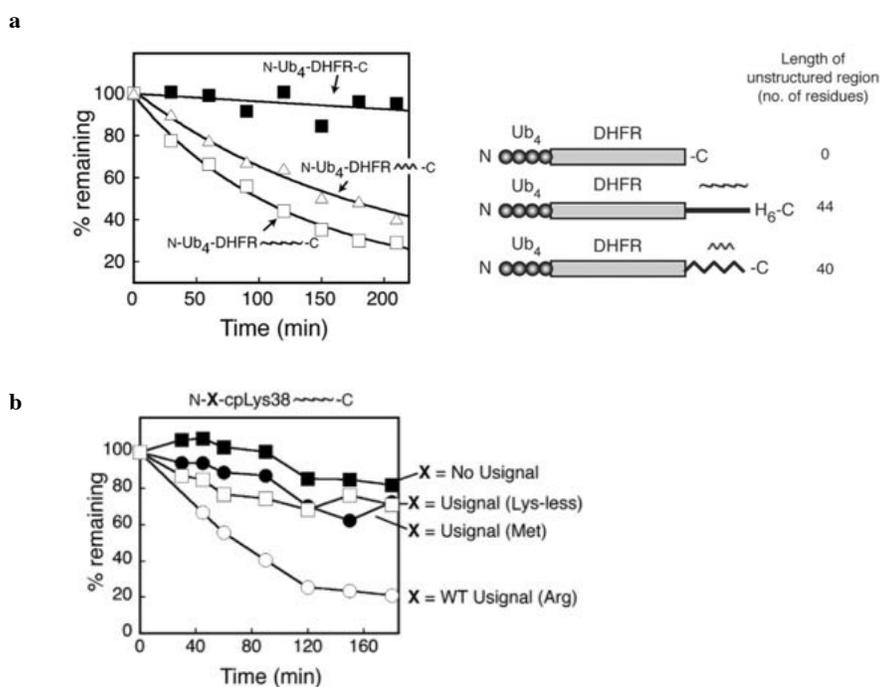


Figure 1.10. a) Degradation of N-Ub₄-tagged DHFR-C by proteasome at 25 °C; b) Degradation of cpLys38 containing a C-terminal unstructured region depends on ubiquitination at Lys15 and Lys17. N-Usignal-cpLys38~~~~-C (○), N-cpLys38~~~~-C (■), N-Usignal(Met)-cpLys38~~~~-C (●) in which the N-end residue has been mutated from arginine to methionine, and N-Usignal(Lys-less)-cpLys38~~~~-C (□) in which Lys15 and Lys17 in the N-end rule tag were mutated to arginine, were degraded by the proteasome at 25 °C in the presence of 320 μM methotrexate. (Adapted from figure in Prakash S, *Nat Struct Mol Biol* 2004)

Further, Inobe *et al.* drew the same conclusions. With the goal of defining the rules that determine how well unstructured regions in proteins can serve as proteasome initiation sites, they measured the minimum length required to support rapid degradation using again series of proteasome substrates containing a central dihydrofolate reductase (DHFR) domain.¹¹²

The proteasome-targeting part of the degron was located at the N-terminus of DHFR, and the initiation region was located at the C-terminus (Figure 1.11).

In one set of constructs, the proteasome-binding tag consisted of four ubiquitin moieties fused in frame to the N-terminus of the substrate (Ub4 tag).¹¹³ This tag mimics a polyubiquitin chain and probably most closely resembles ubiquitin moieties linked through Lys63 of ubiquitin^{114,115} In the other set of constructs, the targeting tag was a single UbL domain derived from yeast Rad23 (UbL tag).^{116,117}

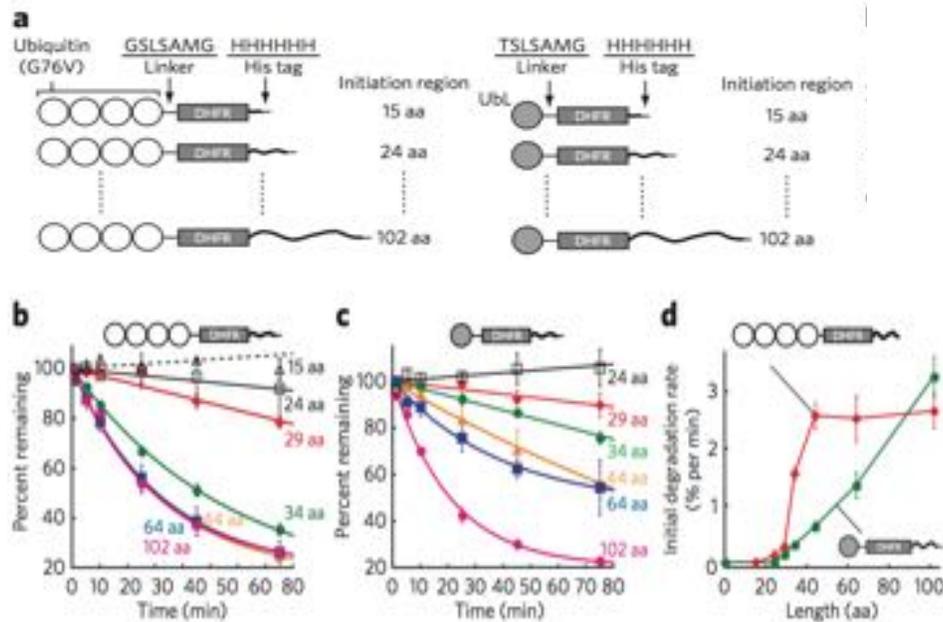


Figure 1.11. Linear representation of substrate proteins with initiation regions of different lengths. The substrates contained an *E. coli* DHFR domain and were targeted to the proteasome by an N-terminal Ub4 tag or a UbL tag. (Adapted from figure in Inobe T, *Nat Chem Biol* 2011)

The unstructured tails were derived from either yeast cytochrome b2 or subunit 9 of mold (*Neurospora crassa*) Fo ATPase because these peptides are known to be unstructured and soluble under standard solution conditions. The tails consisted of 15 to 102 amino acids derived from either protein, and the cytochrome b2 sequences ended in a hexahistidine tag (Figure 1.11a). Shortening the unstructured region in both Ub4- and UbL-tagged substrates decreased degradation rates and finally led to the stabilization of the proteins, once a certain critical length of the initiation region was reached (Figure 1.11b–d). For Ub4-tagged DHFR substrates, the transition occurred between 29- and 34-residue-long initiation regions: proteins with initiation regions of 34 amino acids or longer were degraded, whereas substrates with initiation regions of 29 amino acids or less escaped degradation (Figure 11b,d), in agreement with earlier studies. For UbL-tagged substrates, the transition occurred between 34 and 44 amino acids: constructs with initiation regions of 44 amino acids or more were degraded rapidly, whereas constructs with initiation regions of 34 amino acids or less were not (Figure 1.11c,d).

1.8 References

1. Goldberg, A. L.; Dice, J. F. *Annu. Rev. Biochem.* **1974**, *43*, 835–869.
2. Rabinovitz, M. Translational repression in the control of globin chain initiation by hemin. *Ann N Y Acad Sci* **1974**, *241*, 322–333.
3. Glickman, M. H. A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9–signalosome and eIF3. *Cell* **1998**, *94*, 615–623.
4. Wolf, D.H.; Hilt, W. The proteasome: A proteolytic nanomachine of cell regulation and waste disposal. *Biochim Biophys Acta - Mol Cell Res.* **2004**, *1695*(1-3), 19-31.
5. Rock, K.L.; Gramm, C.; Rothstein, L.; Clark, K.; Stein, R.; Dick, L.; Hwang, D.; Goldberg, A.L. Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell* **1994**, *78*, 761 – 771.
6. Wolf, D.H.; Sommer, T.; Hilt, W. Death gives birth to life: The essential role of the ubiquitin-proteasome system in biology. *Biochim Biophys Acta - Mol Cell Res.* **2004**, *1695*(1-3), 1-2.
7. Hershko, A.; Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **1998**, *67*, 425–479.
8. Ravid, T.; Hochstrasser, M. Diversity of degradation signals in the ubiquitin-proteasome system. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 679– 690.
9. Ventii, K. H.; Wilkinson, K. D. Protein partners of deubiquitinating enzymes. *Biochem. J.* **2008**, *414*, 161–175.
10. Varshavsky, A. Regulated protein degradation. *Trends Biochem. Sci.* **2005**, *30*, 283–286.
11. Ciechanover, A. The ubiquitin proteolytic system: from a vague idea, through basic mechanisms, and onto human diseases and drug targeting. *Neurology* **2006**, *66*, S7-19.
12. Tai, H. C.; Schuman, E. M. Ubiquitin, the proteasome and protein degradation in neuronal function and dysfunction. *Nat. Rev. Neurosci.* **2008**, *9*, 826–838.
13. Hershko, A.; Ciechanover, A. *Annu. Rev. Biochem.* **1998**, *67*, 425–479.
14. Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu Rev Biochem.* **2009**; *78*, 477-513.
15. Simpson, M. V. *J. Biol. Chem.* **1953**, *201*, 143–154
16. Hershko, A.; Tomkins, G. M. *J. Biol. Chem.* **1971**, *246*, 710-714.
17. Ciechanover, A.; Heller, H.; Elias, S.; Haas, A. L.; Hershko, A. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 1365–1368.
18. Hershko, A.; Ciechanover, A.; Heller, H.; Haas, A. L.; Rose, I. A. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 1783–1786.
19. Etlinger, J. D.; Goldberg, A. L. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 54 – 58.
20. Hershko, A.; Heller, H.; Ganoth, D.; Ciechanover, A. Protein Turnover and Lysosome Function (Eds.: H. L. Segal, D. J. Doyle), Academic Press, New York, 1978, pp. 149 – 169.
21. Finley, D. Recognition and processing of ubiquitin–protein conjugates by the proteasome. *Annu. Rev. Biochem.* **2009**, *78*, 477–513.
22. Kravtsova-Ivantsiv, Y.; Sommer, T.; Ciechanover, A. The Lysine48-Based polyubiquitin chain proteasomal signal: Not a single child anymore. *Angew Chemie - Int Ed.* **2013**, *52*(1), 192-198.
23. Ciechanover, A. Proteolysis: From the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol.* **2005**, *6*, 79-86.
24. Hänzelmann, P.; Schäfer, A.; Völler, D.; Schindelin, H. Structural Insights into Functional Modes of Proteins Involved in Ubiquitin Family Pathways. In: Dohmen RJ, Scheffner M, eds. Ubiquitin Family Modifiers and the Proteasome: Reviews and Protocols. Totowa, NJ: Humana Press; 2012:547-576.

25. Lovering, R.; Hanson, I.M.; Borden, K.L.; Martin, S.; O'Reilly, N.J.; Evan, G.I.; Rahman, D.; Pappin, D.J.; Trowsdale, J.; Freemont, P.S. Identification and preliminary characterization of a protein motif related to the zinc finger. *Proc Natl Acad Sci USA* **1993**, *90*, 2112–2116.
26. Borden, K.L. RING domains: master builders of molecular scaffolds? *J Mol Biol.* **2000**, *295*(5), 1103-1112.
27. Borden, K.L.; Boddy, M.N.; Lally, J.; O'Reilly, N.J.; Martin, S.; Howe, K.; Solomon, E.; Freemont, P.S. The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncogene protein PML. *EMBO J* **1995**, *14*: 1532–1541.
28. Barlow, P.N.; Luisi, B.; Milner, A.; Elliott, M.; Everett, R. Structure of the C3HC4 domain by 1H-nuclear magnetic resonance spectroscopy. A new structural class of zinc-finger. *J Mol Biol* **1994**, *237*, 201–211.
29. Albert, T.K.; Hanzawa, H.; Legtenberg, Y.I.A.; van den Heuvel, F.A.J.; Collart, M.A.; Boelens, R.; Timmers, H.T.M. Identification of a ubiquitin-protein ligase subunit within the CCR4–NOT transcription repressor complex. *EMBO J.* **2002**, *21*, 355 – 364.
30. Lorick, K.L.; Jensen, J.P.; Fang, S.; Ong, A.M.; Hatakeyama, S.; Weissman, A.M. RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11364 – 11369.
31. Seol, J.H.; Feldman, R.M.; Zachariae, W.; Shevchenko, A.; Correll, C.C.; Lyapina, J.S.; Chi, Y.; Galova, M.; Claypool, J.; Sandmeyer, S.; Nasmyth, K.; Shevchenko, A.; Deshaies, R.J. Cdc53/cullin and the essential Hrt1 RING-H2 subunit of SCF define a ubiquitin ligase module that activates the E2 enzyme Cdc34. *Genes Dev.* **1999**, *13*, 1614 – 1626.
32. Zheng, N.; Schulman, B.A.; Song, L.; Miller, J.J.; Jeffrey, P.D.; Wang, P.; Chu, C.; Koeppe, D.M.; Elledge, S.J.; Pagano, M.; Conaway, R.C.; Conaway, J.W.; Harper, J.W.; Pavletich, N.P. Structure of the Cull1–Rbx1– Skp1–F boxSkp2 SCF ubiquitin ligase complex. *Nature* **2002**, *416*, 703 – 709.
33. Brzovic, P.S.; Keefe, J.R.; Nishikawa, H.; Mayamoto, K.; Fox, D.; Fukuda, M.; Ohta, T.; Klevit, R. Binding and recognition in the assembly of an active BRCA1–BARD1 ubiquitin ligase complex. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 5646–5651.
34. Chau, V.; Tobias, J.W.; Bachmair, A.; Marriott, D.; Ecker, D.J.; Gonda, D.K.; Varshavsky, A. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* **1989**, *243*, 1576–1583.
35. Spence, J.; Sadis, S.; Haas, A.L.; Finley, D. A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol Cell Biol* **1995**, *15*, 1265-1273.
36. Beal, R.; Deveraux, Q.; Xia, G.; Rechsteiner, M.; Pickart, C. Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. *Proc Natl Acad Sci U S A* **1996**, *93*, 861-6.
37. Thrower, J. S.; Hoffman, L.; Rechsteiner, M.; Pickart, C. M. Recognition of the polyubiquitin proteolytic signal. *Embo J* **2000**, *19*, 94-102.
38. Varadan, R.; Walker, O.; Pickart, C.; Fushman, D. Structural properties of polyubiquitin chains in solution. *J Mol Biol* **2002**, *324*, 637-47.
39. Coux, O.; Tanaka, K.; Goldberg, A.L. Structure and function of 20S and 26S proteasomes. *Annu Rev Biochem.* **1996**, *65*, 801–847.
40. Reits, E.A.J.; Benham, A.M.; Plougastel, B.; Neefjes J.; Trowsdale, J. Dynamics of proteasome distribution in living cells. *EMBO J.* **1997**, *16*, 6087–6094.
41. Scheffner, M. Moving protein heads for breakdown. *Nature* **1999**, *398*, 103–104.
42. Zwickl, P.; Grziwa, A.; Puhler, G.; Dahlmann, B.; Lottspeich, F.; Baumeister, W. Primary structure of the Thermoplasma proteasome and its implications for the structure, function, and evolution of the multicatalytic proteinase. *Biochemistry* **1992**, *31*, 964 – 972.

43. Pqhler, G.; Weinkauff, S.; Bachmann, L.; Mqller, S.; Engel, A.; Hegerl, R.; Baumeister, W. Subunit stoichiometry and three-dimensional arrangement in proteasomes from *Thermoplasma acidophilum*. *EMBO J.* **1992**, 11, 1607 – 1616.
44. Grziwa, A.; Baumeister, W.; Dahlmann, B.; Kopp, F. Localization of subunits in proteasomes from *Thermoplasma acidophilum* by immunoelectron microscopy. *FEBS Lett.* **1991**, 290, 186 – 190.
45. Heinemeyer, W.; Trfndle, N.; Albrecht, G.; Wolf, D.H. PRE5 and PRE6, the last missing genes encoding 20S proteasome subunits from yeast? Indication for a set of 14 different subunits in the eukaryotic proteasome core, *Biochemistry* **1994**, 33, 12229 – 12237.
46. Lfwe, J.; Stock, D.; Jap, B.; Zwickl, P.; Baumeister, W.; Huber, R. Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **1995**, 268, 533 – 539.
47. Groll, M.; Ditzel, L.; Lowe, J.; Stock, D.; Bochtler, M.; Bartunik, H.D.; Huber, R. Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **1997**, 386, 463 – 471.
48. Seemqller, E.; Lupas, A.; Stock, D.; Lowe, J.; Huber, R.; Baumeister, W. Proteasome from *Thermoplasma acidophilum*: a threonine protease. *Science* **1995**, 268, 579 – 582.
49. Heinemeyer, W.; Fischer, M.; Krimmer, T.; Stachon, U.; Wolf, D.H. The active sites of the eukaryotic 20 S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.* **1997**, 272, 25200 – 25209.
50. Chen, P.; Hochstrasser, M. Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell* **1996**, 86, 961 – 972.
51. Hilt, W.; Wolf, D.H. Proteasomes: destruction as a programme. *Trends Biochem. Sci.* **1996**, 21, 96 – 102.
52. Ciechanover, A. The ubiquitin-proteasome proteolytic pathway. *Cell* **1994**, 79(1), 13-21.
53. Glickman, M.H.; Rubin, D.M.; Larsen, C.N.; Schmidt, M.; Finley, D. The regulatory particle of the yeast proteasome, in: W. Hilt, D.H. Wolf (Eds.), *Proteasomes: The World of Regulatory Proteolysis*, Landes Bioscience/Eurekah.com, Georgetown/Austin, USA, 2000, pp. 71 – 90.
54. Gorbea, C.; Rechsteiner, M. The mammalian regulatory complex of the 26S proteasome, in: W. Hilt, D.H. Wolf (Eds.), *Proteasomes: The World of Regulatory Proteolysis*, Landes Bioscience, Georgetown/ Eurekah.com, Austin, USA, Georgetown/Austin, USA, 2000, pp. 91 – 128.
55. Baumeister, W.; Walz, J.; Zuhl, F.; Seemqller, E. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **1998**, 92, 367 – 380.
56. Peters, J.M. Proteasomes: protein degradation machines of the cell. *Trends Biochem. Sci.* **1994**, 19, 377 – 382.
57. Chu-Ping, M.; Vu, J.H.; Proske, R.J.; Slaughter, C.A.; DeMartino, G.N. Identification, purification, and characterization of a high molecular weight, ATP-dependent activator (PA700) of the 20 S proteasome. *J. Biol. Chem.* **1994**, 269, 3539 – 3547.
58. Udvardy, A. Purification and characterization of a multiprotein component of the *Drosophila* 26 S (1500 kDa) proteolytic complex. *J. Biol. Chem.* **1993**, 268, 9055 – 9062.
59. Peters, J.M.; Franke, W.W.; Kleinschmidt, J.A. Distinct 19 S and 20 S subcomplexes of the 26 S proteasome and their distribution in the nucleus and the cytoplasm. *J. Biol. Chem.* **1994**, 269, 7709 – 7718.
60. Deveraux, Q.; Ustrell, V.; Pickart, C.; Rechsteiner, M. A 26 S protease subunit that binds ubiquitin conjugates. *J. Biol. Chem.* **1994**, 269 7059 – 7061.
61. Kleinschmidt, J.A.; Hugle, B.; Grund, C.; Franke, W.W. *Eur. J. Cell Biol.* **1983**, 32, 143-156.

62. Rohrwild, M.; Coux, O.; Huang, H.C.; Moerschell, R.P.; Yoo, S.J.; Seol, J.H.; Chung, C.H.; Goldberg, A.L. HslV–HslU: a novel ATP- dependent protease complex in Escherichia coli related to the eukaryotic proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 5808 – 5813.
63. Yoo, S.J.; Seol, J.H.; Shin, D.H.; Rohrwild, M.; Kang, M.S.; Tanaka, K.; Goldberg, A.L.; Chung, C.H. Purification and characterization of the heat shock proteins HslV and HslU that form a new ATP-dependent protease in Escherichia coli. *J. Biol. Chem.* **1996**, *271*, 14035 – 14040.
64. Kessel, M.; Wu, W.; Gottesman, S.; Kocsis, E.; Steven, A.C.; Maurizi, M.R. Sixfold rotational symmetry of ClpQ, the E. coli homolog of the 20S proteasome, and its ATP-dependent activator, ClpY. *FEBS Lett.* **1996**, *398*, 274 – 278.
65. Missiakas, D.; Schwager, F.; Betton, J.M.; Georgopoulos, C.; Raina, S. Identification and characterization of HslV HslU (ClpQ ClpY) proteins involved in overall proteolysis of misfolded proteins in Escherichia coli. *EMBO J.* **1996**, *15*, 6899 – 6909.
66. Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **2009**, *78*, 477-513.
67. Lowe, J. Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 Å resolution. *Science* **1995**, *268*, 533–539.
68. Groll, M. Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **1997**, *386*, 463–471
69. Unno, M. The structure of the mammalian 20S proteasome at 2.75 Å resolution. *Structure* **2002**, *10*, 609–618
70. Bhattacharyya, S.; Yu, H.; Mim, C.; Matouschek, A. Regulated protein turnover: snapshots of the proteasome in action. *Nat Rev Mol Cell Biol.* **2014**, *15*(2), 122-133.
71. Groll, M.; Ditzel, L.; Lowe, J.; Stock, D.; Bochtler, M.; Bartunik, H.D.; Huber, R. Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature.* **1997**; *386*, 463–471.
72. Groll, M.; Huber, R. Substrate access and processing by the 20S proteasome core particle. *Int J Biochem Cell Biol.* **2003**, *35*(5), 606-616.
73. Lowe, J.; Stock, D.; Jap, B.; Zwickl, P.; Baumeister, W.; Huber, R. Crystal structure of the 20S proteasome from the archeon T. acidophilum at 3.4 Å resolution. *Science* **1995**, *268*, 533–539.
74. Groll, M.; Kim, K.B.; Kairies, N.; Huber, R.; Crews, C.M. Crystal structure of epoxomicin : 20S proteasome reveals a molecular basis for selectivity of alpha 'beta '-epoxyketone proteasome inhibitors. *J. Am. Chem. Soc.* **2000**; *122*, 1237–1238.
75. Lfwe, J.; Stock, D.; Jap, B.; Zwickl, P.; Baumeister, W.; Huber, R. Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 Å resolution. *Science* **1995**, *268*, 533 – 539.
76. Huber, E.M.; Heinemeyer, W. Li. X.; Arendt, C.S.; Hochstrasser, M.; Groll, M. A unified mechanism for proteolysis and autocatalytic activation in the 20S proteasome. *Nat Commun.* **2016**, *7*, 10900.
77. Seemüller, E.; Lupas, A.; Stock, D.; Lfwe, J.; Huber, R.; Baumeister, W. Proteasome from Thermoplasma acidophilum: a threonine protease. *Science* **1995**, *268*, 579-582.
78. Stock, D.; Ditzel, L.; Baumeister, W.; Huber, R.; Lfwe, J. The catalytic mechanism of the 20S proteasome of Thermoplasma acidophilum revealed by X-ray crystallography. *Co/d Spring Harb Symp Quant Bio* **1996**, *60*, 525-532.
79. Arendt, C.S.; Hochstrasser, M. Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active-site formation. *Proc. Natl Acad. Sci. USA.* **1997**, *94*, 7156–7161.
80. Groll, M.; Bochtler, M.; Brandstetter, H.; Clausen, T.; Huber, R. Molecular machines for protein degradation. *ChemBioChem.* **2005**, *6*, 222–256.

81. Farout, L.; Lamare, M.C.; Cardozo, C.; Harrisson, M.; Briand, Y.; Briand, M. Distribution of proteasomes and of the five proteolytic activities in rat tissues. *Arch Biochem Biophys.* **2000**, 374(2), 207-212.
82. Glickman, M.H.; Rubin, D.M.; Coux, O.; Wefes, I.; Pfeifer, G.; Cjeka, Z.; Baumeister, W.; Fried, V.A.; Finley, D. A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* **1998**, 94, 615 – 623.
83. Lander, G.C.; Estrin, E.; Matyskiela, M.E.; Bashore, C.; Nogales, E.; Martin, A. Complete subunit architecture of the proteasome regulatory particle. *Nature* **2012**, 482(7384), 186-191.
84. Matyskiela, M.E.; Lander, G.C.; Martin, A. Conformational switching of the 26S proteasome enables substrate degradation. *Nat Struct Mol Biol.* **2013**, 20(7), 781-788.
85. Kim, T.; Hofmann, K.; Arnim, von, A. G.; Chamovitz, D. A. PCI complexes: pretty complex interactions in diverse signaling pathways. *Trends in plant science* **2001**, 6, 379–386.
86. Lingaraju, G. M.; Bunker, R.D.; Cavadini, S.; Hess, D.; Hassiepen U.; Renatus, M.; Fischer, E.S.; Thoma, N.H. Crystal structure of the human COP9 signalosome. *Nature* **2014**, 512, 161-165.
87. Yao, T.; Cohen, R.E. A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* **2002**, 419, 403–407.
88. Verma, R.; Aravind, L.; Oania, R.; McDonald, W.H.; Yates, J.R. 3rd; Koonin, E.V.; Deshaines, R.J. Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **2002**, 298, 611–615.
89. Pathare, G.R.; Nagy, I.; Bohn, S. The proteasomal subunit Rpn6 is a molecular clamp holding the core and regulatory subcomplexes together. *Proc Natl Acad Sci.* **2012**, 109(1), 149-154.
90. Tanaka, K. The proteasome: overview of structure and functions. *Proc Jpn Acad Ser B Phys Biol Sci.* **2009**, 85(1), 12-36.
91. Beck, F.; Unverdorben, P.; Bohn, S. Near-atomic resolution structural model of the yeast 26S proteasome. *Proc Natl Acad Sci U S A.* **2012**, 109(37), 14870-14875.
92. Rosenzweig, R.; Bronner, V.; Zhang, D.; Fushman, D.; Glickman, M. H. Rpn1 and Rpn2 coordinate ubiquitin processing factors at proteasome. *J. Biol. Chem.* **2012**, 287, 14659–14671.
93. Elsasser, S. Proteasome subunit Rpn1 binds ubiquitin-like protein domains. *Nat. Cell Biol.* **2002**, 4, 725–730
94. Platt, T.; Miller, J.H.; Weber, K. In vivo degradation of mutant lac repressor. *Nature* **1970**, 228, 1154–1156.
95. Rabinovitz, M. Translational repression in the control of globin chain initiation by hemin. *Ann NY Acad Sci* **1974**, 241, 322–333.
96. Varshavsky, A. Naming a targeting signal. *Cell* **1991**, 64, 13–15.
97. Finley, D. Recognition and Processing of Ubiquitin-Protein Conjugates by the Proteasome. *Annu Rev Biochem.* **2009**, 78, 477–513.
98. Tomko, R.J.; Hochstrasser, M. Molecular architecture and assembly of the eukaryotic proteasome. *Annu Rev Biochem.* **2013**, 82(1), 415–45.
99. Hershko, A.; Ciechanover, A.; Heller, H.; Haas, A.L.; Rose, I.A. Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis. *Proc Natl Acad Sci U S A* **1980**, 77, 1783–1786.
100. Hershko, A.; Heller, H.; Eytan, E.; Reiss, Y. The protein substrate binding site of the ubiquitin-protein ligase system. *J Biol Chem* **1986**, 261, 11992–11999.

101. Hershko, A.; Leshinsky, E.; Ganoth, D.; Heller, H. ATP-dependent degradation of ubiquitin-protein conjugates. *Proc. Natl. Acad. Sci. U S A* **1984**, 81, 1619–1623.
102. Bachmair, A.; Finley, D.; Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **1986**, 234, 179–186.
103. Deveraux, Q.; Ustrell, V.; Pickart, C.; Rechsteiner, M. A 26S protease subunit that binds ubiquitin conjugates. *J. Biol. Chem.* **1994**, 269, 7059–7061.
104. Thrower, J.S.; Hoffman, L.; Rechsteiner, M.; Pickart, C.M. Recognition of the polyubiquitin proteolytic signal. *EMBO J.* **2000**, 19, 94–102.
105. Verma, R. Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **2002**, 298, 611–615.
106. Petroski, M.D.; Deshaies, R.J. Context of multiubiquitin chain attachment influences the rate of Sic1 degradation. *Mol. Cell* **2003**, 11, 1435–1444.
107. Fishbain, S.; Prakash, S.; Herrig, A.; Elsasser, S.; Matouschek, A. Rad23 escapes degradation because it lacks a proteasome initiation region. *Nat Commun.* **2011**, 2, 192.
108. Prakash, S.; Tian, L.; Ratliff, K.S.; Lehotzky, R.E.; Matouschek, A. An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat Struct Mol Biol* **2004**, 11, 830–837.
109. Sauer, R.T.; Baker, T.A. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem.* **2011**, 80, 587–612.
110. Beckwith, R.; Estrin, E.; Worden, E.J.; Martin, A. Reconstitution of the 26S proteasome reveals functional asymmetries in its AAA+ unfoldase. *Nat Struct Mol Biol.* **2013**, 20, 1164–1172.
111. Takeuchi, J.; Chen, H.; Coffino, P. Proteasome substrate degradation requires association plus extended peptide. *EMBO J.* **2007**, 26, 123–131.
112. Inobe, T.; Fishbain, S.; Prakash, S.; Matouschek, A. Defining the geometry of the two-component proteasome degron. *Nat Chem Biol* **2011**, 7, 161–167.
113. Stack, J.H.; Whitney, M.; Rodems, S.M.; Pollok, B.A. A ubiquitin-based tagging system for controlled modulation of protein stability. *Nat. Biotechnol.* **2000**, 18, 1298–1302.
114. Komander, D.; Reyes-Turcu, F.; Licchesi, J.D.F.; Odenwaelder, P.; Wilkinson, K.D.; Barford, D. Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. *EMBO Rep.* **2009**, 10, 466–473.
115. Saeki, Y.; Kudo, t.; Sone, T.; Kikuchi, Y.; Yokosawa, H.; Toh-e, A.; Tanaka, K. Lysine 63-linked polyubiquitin chain may serve as a targeting signal for the 26S proteasome. *EMBO J.* **2009**, 28, 359–371.
116. Watkins, J.F.; Sung, P.; Prakash, L.; Prakash, S. *Saccharomyces cerevisiae* DNA repair gene RAD23 encodes a nuclear protein containing a ubiquitin-like domain required for biological function. *Mol. Cell. Biol.* **1993**, 13, 7757–7765.
117. Schaubert, C.; Chen, L.; Tongaonkar, P.; Vega, I.; Lambertson, D.; Potts, W.; Madura, K. Rad23 links DNA repair to the ubiquitin/proteasome pathway. *Nature* **1998**, 391, 715–718.

Chapter 2

Rescue of C-terminal destabilized proteins by use of D-amino acids

2.1 Introduction

2.1.1 Delivery of therapeutic proteins: challenges and strategies

The roll call of the bio-therapeutics comprises some of the most innovative and highly successful drugs in treating a wide range of serious disease.

First of all, insulin therapy revolutionized the treatment of diabetes. Since then a research-intensive enterprise has allowed scientist to develop modern insulins mimicking the physiological insulin secretion closely.¹ The clinical use of enzyme has also a laundry list of widespread applications. Exogenous enzymes, for example, are the only cure for a large group of disorders (lysosomal storage diseases, LSD_s) caused by a lack of degradation of substances within the lysosomes.² More broadly, currently biotech products approved for clinical use are more than 100 including anticoagulants, growth factors, interferons, antibody-based drugs, and their efficacy, specificity and selectivity are constantly improved trying to satisfy key parameters for targeting, uptake, translocation, and activity of the payload.³ However, we are far from taking full advantage of these potent bio-machines.

Molecular methodologies are still required to address and truly perturb intracellular trafficking environment, minimizing side effects from off-target interactions, in a range of research and human therapeutic purpose.

The cellular membrane, which restricts, as a rule, the spontaneous crossing of macromolecules, is a real roadblock. High molecular weight and surface properties of biologically active molecules are the key issues and hamper these therapeutic candidates from finding their way to the cytoplasm and to desired subcellular locations of interest.

Bioavailability barrier of plasma membrane has two major hurdles: delivery and endosomal escape. First, the protein, modified or conjugated, must be carried into the cell exploiting, the cell internalization machine via endocytosis and phagocytosis.

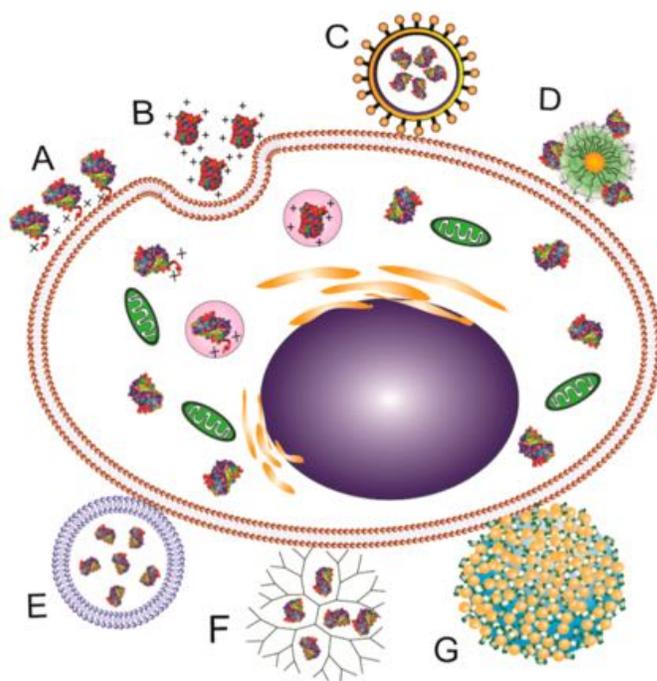


Figure 2.1. Schematic illustration of protein delivery systems. (A) cell- penetrating peptides, (B) supercharged proteins, (C) virus-like particle, (D) nanocarrier, (E) liposomes, (F) polymer, and (G) nanoparticle-stabilized nanocapsule. (Adapted from figure in Fu A., *Bioconjug Chem* 2014)

Once inside, the vesicular entrapment must be decomposed and release the encapsulated protein to reach the desired specific targeted cytosolic compartments. Several methods to achieve an effective and general internalization of functional protein into cells have been explored and developed over the past decade, focusing on two different approaches: mechanical/physical methods and carried-based delivery system.⁴⁻¹² The former group includes invasive techniques such as microinjection and electroporation, which have, respectively, downsides of needing high skills to operate the equipment and significant rate of physical damage caused by rapid and high- voltage electric pulse.^{13,14} Also, the cell-to-cell variability and the influx of further proteins are important limitations.¹⁵ To gain the access to the interior of the cell of foreign biotherapeutics a carried-based delivery approach is an attractive alternative. There are two typologies: covalent protein modification and supramolecular delivery system, each of which consists of several subcategories including cell-penetrating peptides (CPP_s), virus-like particles, supercharged proteins, nanocarriers,

supramolecular carrier-based delivery system and nanoparticle-stabilized nanocapsules (Figure 2.1).¹²

CPPs, or Protein Transduction Domain (PTD) are short, cationic and/or amphipathic peptides of less than 30 residues in length which fully penetrate through the cell membrane.¹⁶ The molecular mechanism of the transduction is not entirely clarified and is still under debate.¹⁷⁻¹⁹ Caveolar, clathrin-mediated and macropinocytosis are some of the postulated pathways.¹⁸⁻²³ Moreover, several game plans for raising endosomal lysis have been adopted (pH- and temperature- elements, synthetic vesicular lysis agents) with the aim of preventing the CPP-tagged vehicle's disassembly within the lysosome before reaching the wanted target. Therefore, the efficiency of cytosolic access is hard to quantify, as in the case of Virus-like particles.²⁴⁻²⁶ For the latter, these vectors are also affected by immunogenicity and safety problems. A class of engineered or naturally occurring human proteins, with unusually high net negative or positive charge, able to escort other macromolecules into the cell have been identified.²⁷ However, a limitation of this delivery platform should not be underestimated: modifications proteins in terms of charge might alter the properties and the activity of the protein.

Physical aggregates of amphiphilic molecules, such as liposome, lipo- and nano- plexes, nanoparticle-stabilized nanocapsule have been also considered as powerful carriers, allowing the uptake of unmodified proteins into mammalian cells.²⁸⁻³³ Whilst a lower dose of drug together with a reduction of side-effects are encouraged features, insufficient specific cell localization of targeted systems, diffusion and redistribution of released drugs, higher cost are influential disadvantages.³⁴ Thus, each of the above-mentioned intracellular delivery tools has limitations and challenges that still need to be overcome for optimal application.

2.1.2 Break on through to the other side: the bacterial toxin mediated delivery

Reviewing the many approaches taken to date, a promising route to achieve the release of macromolecules in its active form into the cytosol is believed to be the synergistic coengineering of proteins and carriers as integrated vectors.

Recently, with the goal of developing a highly effective and receptor-specific intracellular carrier, many efforts have been made in the engineering of protein/peptide “pro-form” and “chimeras” linked with a peptide/proteinaceous domain.

Inspired by nature, researchers exploited the ability of toxins to spread through the body and started to investigate the sophisticated bacterial toxin-mediated delivery nanomachines.^{35,36} Some pathogens

have already evolutionary solved the trafficking barrier problem: AB-type toxins, in fact, are able to interfere with specific intracellular substrate gained because of it this increasing interest.

AB toxins consist of at least two functional polypeptides, the subunit A, responsible for the enzymatic activity of the toxin, and the receptor binding subunit B. Once the B domain binds to a specific cell surface receptor, triggering endocytosis and internalization, the A moiety is translocated across the cell membrane into the cytosol.

Working as a macromolecular syringe, the pathogen injects his toxins into the cytoplasm of host cells in a directly and selectively way. Thus, the disarmed version of the toxin has been recruited for this highly efficient and flexible targeting system to introduce non-naturally occurring agents and chemical entities such as D-peptides/proteins, cyclic peptides/proteins.

In this way, the translocation domain *Pseudomonas aeruginosa* exotoxin A (PE) has been employed with CPP for the delivery of a protein cargo in order to facilitate the endosomal escape's payload;³⁷ heat-labile Enterotoxin IIa was engineered to deliver a heterologous protein into neurons.³⁸

In particular, the protective antigen (PA) of Anthrax Toxin coadministered with the N-terminal domain of anthrax toxin lethal factor (LF_N) has been fused to several payloads of interest to facilitate their delivery.³⁹

2.1.3 AT- a paradigm for toxin-mediated protein translocation

Bacillus anthracis is a well-adapted and very efficient pathogen. The bacillus secretes three non-toxic monomeric proteins self assemble into a toxic non covalent complex that is crucial in causing anthrax disease symptoms. Anthrax toxic can be classified as a binary A₂B toxin, in which the Protective Antigen (PA) [PA83; 83 kDa] functions as the receptor binding moiety, allowing entry into the cytosol of mammalian cells for either of the two acting enzymes Lethal Factor and Edema Factor.⁴⁰ Lethal Factor [LF; 90 kDa] is a Zinc metalloproteinase that cleaves MAPK_s family members in their amino termini interfering with one or more signaling pathways leading to apoptosis. Edema Factor [EF; 89kDa] is a Ca²⁺ and calmodulin- dependent adenylate cyclase which increase the cAMP concentration in cells protecting bacteria from phagocytic destruction and helping spreading *B. anthracis* in the host by affecting signaling pathways and modulating immunologic responses.

At the molecular level, PA is recognized by and binds to one of its receptors on the host cells surface: Tumor Endotelial Marker 8 (TEM8, or ANTXR1) and Capillary Morphogenesis Gene 2 (CMG2, or

ANTXR2) with nanomolar or picomolar affinity, respectively.⁴¹⁻⁴³ Furthermore, these receptors are expressed on almost every type of human cell line at approximately 2000-50000 receptors per cell.⁴⁴ Once bound, PA₈₃ undergoes cleavage of the amino terminal 20 kDa fragment by a furin-family protease.⁴⁵⁻⁴⁸ The larger PA₆₃ subunit remains attached to the receptor and readily oligomerizes into a ring-shaped heptamer or octamer (Figure 2.2).⁴⁹ The PA pre-pore is capable of binding in a competitive fashion and with nanomolar affinity multiple copies of LF and/or EF. EF and LF share significant sequence similarity with each other over the N-terminal ~250 residues. Given their different catalytic activities, this led to the hypothesis that this region contains the determinants required for binding to, and translocation by, PA. Arora et al. supported this proposal showing that the protein resulting from the fusion of the amino terminal of LF and the catalytic domain of Pseudomonas exotoxin A was efficiently internalized to the cytosol in a PA-dependent manner.⁵⁰ The resulting assembled complex is internalized and delivered to the endocytic vesicle. Within this acidic environment, the PA pre-channel is converted into membrane-spanning channel and the destabilized LF and EF native structures are dissociated and translocated.

The “flower-on-a-stem” shaped pore serves as an active transporter: ATP-driven cellular factors are not required. The pathway could be described in three successive steps: docking, protein unfolding and translocation of the unstructured chain (Figure 2.3).⁵¹ Upon channel conversion and subsequent binding of the substrate in the narrow ring ϕ clamp, pore chaperones the advance of the protein to the cytoplasm. Due to the high number of glutamic and aspartic acid residues, protonated in that condition, the canal is strongly cation-selective and accommodates the anion unfolded side-chains.^{52,53} Also, biochemical and electrophysiological results indicated that the proton gradient across the membrane (ΔpH), cooperatively with a charge state-dependent Brownian ratchet, supports the free translocation of LF and EF in an N- to C- terminal direction. Furthermore, mutagenesis analyses pointed out that substitution of aromatic residues at position 427 with smaller aliphatic or hydrophilic substitutions (Ala, Ile, Ser, Asp) decreases the efficiency of the translocation mechanism, suggesting that the solvent-exposed, lumen-facing ϕ clamp effectively provides to preserve the positive gradient by acting as a hydrophobic sigil.⁴⁹

Over the past two decades, because it has proven to be one of the most tractable toxins, the nontoxic PA/LF_N nanomachine has been widely developed. The toxin from Bacillus anthracis is, in fact, an ideal model system for the study of key structural requirements of functional protein/peptide's translocation across membranes.

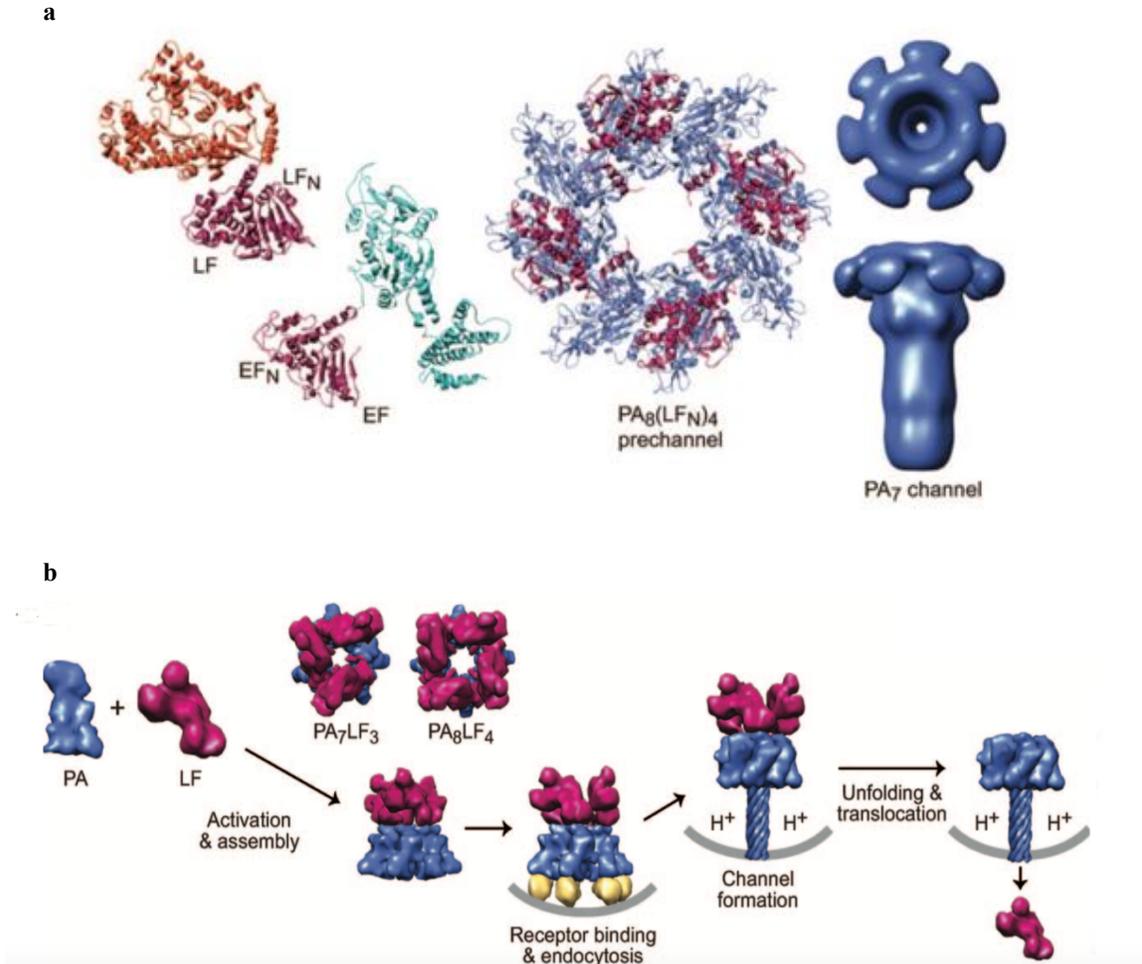


Figure 2.2. a) amino-terminal PA binding domains (LF_N and EF_N , respectively) are colored red-violet and their catalytic domains colored orange and cyan, respectively. b) Anthrax toxin assembly and transport. PA (denim) is proteolytically nicked and assembles with LF (red-violet) and forms prechannel complexes, which bind cellular receptors (gold) triggering endocytosis; acidic pH conditions in the endosome induce PA to form a transmembrane channel; the pH gradient that develops across the endosomal membrane destabilizes LF, drives LF unfolding and translocation through the PA channel. (Adapted from figure in Feld GK et al., *Protein Sci.* 2012)

LF_N has been demonstrated to be important for initiating translocation into living cells and several proteins have been combined with Lethal Factor N-terminal domain: the A chain of diphtheria toxin (DTA),⁵⁴⁻⁵⁶ Pseudomonas exotoxin A (PE),⁵⁷ A chain of Shiga toxin (STA),⁵⁴ dihydrofolate reductase (DHFR),⁵⁶ and β lactamase,⁵⁸ proving the promiscuity of the PA pore.

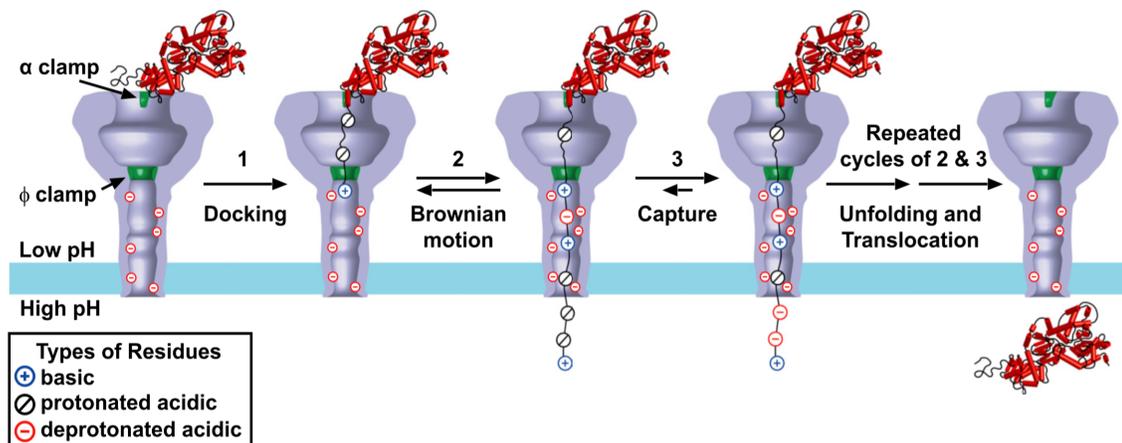


Figure 2.3. The charge state Brownian ratchet model for pH-driven translocation of anthrax toxin. At the time of endosome acidification, the substrate is bound to the top of the channel, with its 1 helix in the clamp. Upon channel conversion, the unstructured amino terminus docks in the clamp. The low pH of the endosomal compartment will protonate most of the acidic residues, whereas positive charges chaperone any remaining deprotonated aspartic or glutamic acids. This ensures that the translocating polypeptide will have a net positive charge, allowing it to move freely through the cation-selective channel. Because of Brownian motion, a portion of the substrate will eventually emerge in the cytosol. There, the higher pH of this region will result in frequently deprotonated acidic residues, thereby giving the emerged portion of the polypeptide a net negative charge and capturing it on the cytosolic side of the membrane. Repeated cycles of emergence from the channel through Brownian motion and capture via deprotonation allow the remaining portion of the substrate to translocate across the membrane. (Adapted from a figure in Brown MJ, *J Biol Chem* 2011)

These hybrid proteins must be able to adopt unfolded or extended conformation. However, the measured enzymatic activity of each reported payload confirms that, once in the cytosol, they perform their duties against the desired target, attesting their refolded conformation. On the other hand, Wesche et al. have demonstrated that LF_N-DHFR bound to methotrexate, as well as LF_N-DTA containing an artificial disulfide, are not able to penetrate a cytosol-contiguous membrane.⁵⁹ Non-natural/cyclic peptides and small drugs have been also employed. In general, cargos do not disrupt the translocation machinery, with the exception of moieties with low pK_a values that cannot be protonated in the endosome.⁶⁰

2.1.4 Fusion and conjugation approaches

Recombinant expression has enabled the processing of chimeric proteins with the aim of explore the delivery pathway and so investigate the biological function of key payloads inside the cytosol of the cells. However, this manipulating technique is not suitable for site specific post-translational modification and incorporation of non-canonical features, such as modified backbone moieties or amino acids with inverted stereochemistry, into proteins.

To overcome this issue and meticulously probe their bioactivity, stability, or affinity properties, powerful semisynthetic (such as native chemical ligation, NCL) and enzymatic (like enzyme-mediated ligation using sortase A, SrtA) strategies have been drawn up. Pentelute et al. used NCL to ligate non-natural peptide sequences onto the N-terminus of LF_N in order to deepen the translocation initiation PA-mediated.^{60,61} This chemoselective approach can be usefully applied to incorporate a wide range of modular chains within multi-functional proteins, but oftentimes it is performed under denaturing conditions with the aim of obtain optimal substrate concentrations. Consequently, ligated products may not refold properly and preserve their structure, resulting in lack of activity.⁶²

The catalytic domain of sortase A from *Staphylococcus aureus* (SrtA) has been found to be useful as a “molecular stapler”: this enzyme-mediated tailoring has allowed for the facile and specific fixing of protein-protein fusions, safeguarding their functionality.⁶³

2.1.5 SrtA-catalyzed ligation

SrtA is widely present in bacteria and catalyzes the surface protein covalent anchoring to the peptidoglycan of the cell wall envelope, a process closely involved in the control of the *S. Aureus* virulence into the host.^{64,65}

A full-length precursor protein equipped with an N-terminal signal peptide is exported from the cytoplasm (Figure 2.4).⁷¹ The C-terminal consists of an LPxTG motif followed by a hydrophobic region and a cluster of cationic amino acid residues, which retain the protein anchored to the membrane.⁶⁶ Sortase A acts as a cysteine protease and transpeptidase. The enzyme is able to scan secreted polypeptides for its recognition and cleaves the scissile peptide bond between the Threonyl and the Glycyl of LPxTG motif sorting signal which, protruding from the outside of the bacterium, is hosted in direct proximity of the active-site sulphhydryl Cys¹⁸⁴ generating a sortase-protein complex via thioester-linked intermediate.⁶⁷ The sortase acyl intermediate is resolved by the nucleophilic attack of the free G₅ residues amino group within the membrane anchored cell-wall

precursor (Lipid II) bounding the protein to the cross-bridge peptide. Transglycosilation and Transpeptidations reactions then incorporate this product into the peptidoglycan and the enzyme active site is regenerated.

SrtA-mediated ligation is an attractive and resourceful engineering tool for specific biologics modification and incorporation of several probes into protein fusions, with application ranging from protein conjugation and surface labelling.

While the introduction of the LPxTG moiety do not represent a problem with LF_N, it might alter the properties of other protein and could be the only restriction to the application of the sortagging. However, Piotuck et al. recently envolved SrtA able to identify FPxTG or APxTG motif, prompting that even this limitation may be overcome with different and possibly shorter sequences.⁶⁸

Furthermore, Chen et al. developed an improved SrtA with increased efficiency for the conjugation reaction overcoming wild-type enzyme's poor kinetics. SrtA* variant achieves comparable conjugation yields while requires significantly lower amount of the enzyme and abbreviated times carrying out the reaction under non denaturing conditions.^{69,70}

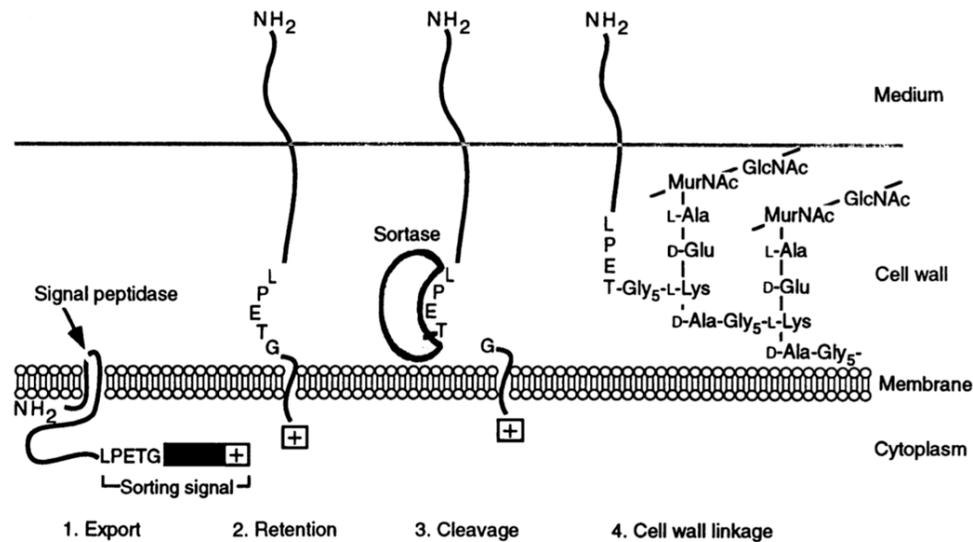


Figure 2.4. Diagram showing the structure of the cell wall anchor of surface proteins in *S. Aureus*.
(Adapted from *Science* 1995)

2.2 Results and discussion

2.2.1 Design

Biomolecules composed of mirror image amino acids are of particular interest for their unique biological stability and non immunogenic properties. However, the biological impact of polypeptides containing amino acids with non-natural side chains, backbone composition, mirror image chirality, and many others remains relatively unexplored.⁷²

The main challenge with probing the stability of proteins containing non-natural functionalities is the delivery of such proteins into the cytosol.

Moreover, the investigation of the mechanism of intracellular ATP-dependent, ubiquitin-dependent proteasomal degradation has been so far carried on using only rabbit reticulocyte lysate (RRL) for *in vitro* degradation assay. Further, the physical size of initiation regions of some 30 amino acids has been established to be roughly on the appropriate length scale to fit the proteasome structure.

Facing this demanding situation, we asked ourselves how can we study in cells the degradation proteasome-mediated and if we would obtain the same results of in previous reported *in vitro* assays, and also what minimum functional and structural properties make a protein the target for degradation by the proteasome.

Performing in cell experiment could provide new insights into the molecular mechanism regarding the selection of substrates and so their stability toward the proteasome.

We thought to utilize the LF_N/PA platform derived from nature for studying the intracellular stability of proteins containing D-amino acids. We used sortase A (SrtA) from *Staphylococcus aureus* to ligate mixed chirality peptide sequences onto the C-terminus of L-proteins that can then be delivered in a PA dependent manner.⁷⁰ Furthermore, we incorporated a cleavable linker that releases the cargo protein from LF_N after translocation into the cytosol further allowing us to characterize C-terminal D-amino acids on proteins other than LF_N, including A-chain of diphtheria toxin (DTA).⁷³

2.2.2 Chemistry

2.2.2.1 Peptide synthesis

All the mixed chirality peptides were synthesized adopting a classical Fmoc/Boc solid phase peptide strategy (Table 2.1). Each sequence contains a penta-glycine bridge at the N-terminus for the following SrtA*-mediated reaction.

2.2.2.2 Sortase A attaches peptide sequences onto the C-terminus of LF_N and LF_N-DTA

Here we used SrtA*- mediated ligation to attach the N-terminal of mixed chirality peptides containing an oligoglycine moiety on the LPSTGG C-terminus of LF_N (Figure 2.5, Table 2.2). Our one-pot chemoenzymatic bioconjugation protocol was carried out over only 30 min at room temperature without any further purification. In addition, in order to explore the behavior of cargo on protein different from LF_N but still able to translocate through the PA channel, we incorporated a cleavable linker, the A-chain of diphtheria toxin (DTA) (Figure 2.5, Table 2.3), that releases the payloads from LF_N after translocation into the cytosol.

2.2.3 Biological Studies

2.2.3.1 Translocation of LF_N-DTA-X and LF_N-X Constructs

The LF_N and LF_N-DTA constructs were delivered into Chinese hamster ovary cells (CHO-K1) in presence of PA for 6 h (Figure 2.6) Digitonin lysis buffer, supplemented with protease inhibitor cocktail, was used for the aptitude of the steroid glycoside for permeabilizing the plasma membrane via the bind with the cholesterol and other β -hydroxy steroids, leading to the formation of pores within the phospholipid bilayer and its subsequent disruption.⁷⁴ Digitonin thereby releases cytoplasmatic proteins leaving the other intracellular membrane organelles structurally intact, as the cholesterol composition of these membranes is lower. Lysates were analyzed by western blot. The membranes were immunoblotted with the targeted antibodies for the stability analysis, and then labeled with Erk1/2 for the cytosolic fraction and Rab5, an early endosome marker. That allowed the evaluation of the extraction efficiency which depends on the amount of contamination of early endosome and high level of Erk1/2. Moreover, Lactacystin, a proteasome inhibitor, was added to the cells treated with designate conjugates as a control. Regarding the mechanism, Lactacystin itself does not react with the proteasome;⁷⁵ the organic compound, isolated for the first time as a microbial metabolite from Streptomyces, spontaneously undergoes an intramolecular reaction forming the cell-permeable clastro-lactacystin β -lactone, which binds the proteasome active site Threonines and irreversibly inhibits via acylation its activity.

Table 2.1.

comp	sequence
1	H - Gly ₅ -Ala-Lys-Phe-Arg-Pro-Asp-Ser-Asn-Val-Arg-Gly-CONH ₂
2	H - Gly ₅ -Ala-Lys-Phe-Arg-Pro-Asp-Ser-Asn-Val-Arg-Gly-COOH
3	H - Gly ₅ -Ser-Arg-Ala-Phe-Asn-Gly-Val-Arg-Asp-Pro-Lys- CONH ₂
4	H - Gly ₅ -Ser-Arg-Ala-Phe-Asn-Gly-Val-Arg-Asp-Pro-Lys- COOH
5	H - Gly ₅ -DAla-DLys-DPhe-DArg-DPro-DAsp-DSer-DAsn-DVal-DArg-Gly-CONH ₂
6	H - Gly ₅ -DAla-DLys-DPhe-DArg-DPro-DAsp-DSer-DAsn-DVal-DArg-Gly-COOH
7	H - Gly ₅ -DSer-DArg-DAla-DPhe-DAsn-Gly-DVal-DArg-DAsp-DPro-DLys- CONH ₂
8	H - Gly ₅ -DSer-DArg-DAla-DPhe-DAsn-Gly-DVal-DArg-DAsp-DPro-DLys- COOH
9	H - Gly ₅ -Ser-Arg-Ala-Phe-Asn-Gly-Val-Arg-Asp-DPro-DLys- CONH ₂
10	H - Gly ₅ -Ser-Arg-Ala-Phe-Asn-Gly-Val-Arg-Asp-Pro-DLys- CONH ₂
11	H - Gly ₅ -Ala-Lys-Phe-Arg-Pro-Asp-Ser-Asn-Val-DArg-Gly-CONH ₂
12	H - Gly ₅ -DAla-Lys-Phe-Arg-Pro-Asp-Ser-Asn-Val-Arg-Gly-CONH ₂
13	H - Gly ₅ -Ala-Lys-Phe-Arg-DPro-Asp-Ser-Asn-Val-Arg-Gly-CONH ₂

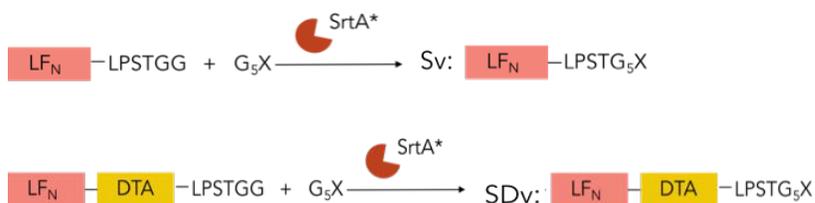


Figure 2.5. G_5X peptides, where X represents mixed chirality sequences, are ligated onto LF_N -LPSTGG or LF_N -DTA-LPSTGG using sortase A* ($SrtA^*$) to yield sortagged variants (SV's, LF_N -X constructs) or sortagged DTA variants (SDV's, LF_N -DTA-X constructs).

Table 2.2

comp	sequence
Sv1	LF _N - Gly ₅ -Ala-Lys-Phe-Arg-Pro-Asp-Ser-Asn-Val-Arg-Gly-CONH ₂
Sv2	LF _N - Gly ₅ -Ala-Lys-Phe-Arg-Pro-Asp-Ser-Asn-Val-Arg-Gly-COOH
Sv3	LF _N - Gly ₅ -Ser-Arg-Ala-Phe-Asn-Gly-Val-Arg-Asp-Pro-Lys- CONH ₂
Sv4	LF _N - Gly ₅ -Ser-Arg-Ala-Phe-Asn-Gly-Val-Arg-Asp-Pro-Lys- COOH
Sv5	LF _N - Gly ₅ -DAla-DLys-DPhe-DArg-DPro-DAsp-DSer-DAsn-DVal-DArg-Gly-CONH ₂
Sv6	LF _N - Gly ₅ -DAla-DLys-DPhe-DArg-DPro-DAsp-DSer-DAsn-DVal-DArg-Gly-COOH
Sv7	LF _N - Gly ₅ -DSer-DArg-DAla-DPhe-DAsn-Gly-DVal-DArg-DAsp-DPro-DLys- CONH ₂
Sv8	LF _N - Gly ₅ -DSer-DArg-DAla-DPhe-DAsn-Gly-DVal-DArg-DAsp-DPro-DLys- COOH
Sv9	LF _N - Gly ₅ -Ser-Arg-Ala-Phe-Asn-Gly-Val-Arg-Asp-DPro-DLys- CONH ₂
Sv10	LF _N - Gly ₅ -Ser-Arg-Ala-Phe-Asn-Gly-Val-Arg-Asp-Pro-DLys- CONH ₂
Sv11	LF _N - Gly ₅ -Ala-Lys-Phe-Arg-Pro-Asp-Ser-Asn-Val-DArg-Gly-CONH ₂
Sv12	LF _N - Gly ₅ -DAla-Lys-Phe-Arg-Pro-Asp-Ser-Asn-Val-Arg-Gly-CONH ₂
Sv13	LF _N - Gly ₅ -Ala-Lys-Phe-Arg-DPro-Asp-Ser-Asn-Val-Arg-Gly-CONH ₂

Table 2.3

comp	sequence
SDv1	LF _N -DTA-Gly ₅ -Ala-Lys-Phe-Arg-Pro-Asp-Ser-Asn-Val-Arg-Gly-CONH ₂
SDv2	LF _N - DTA-Gly ₅ -Ala-Lys-Phe-Arg-Pro-Asp-Ser-Asn-Val-Arg-Gly-COOH
SDv3	LF _N - DTA-Gly ₅ -DAla-DLys-DPhe-DArg-DPro-DAsp-DSer-DAsn-DVal-DArg-Gly-CONH ₂
SDv4	LF _N - DTA-Gly ₅ -DAla-DLys-DPhe-DArg-DPro-DAsp-DSer-DAsn-DVal-DArg-Gly-COOH

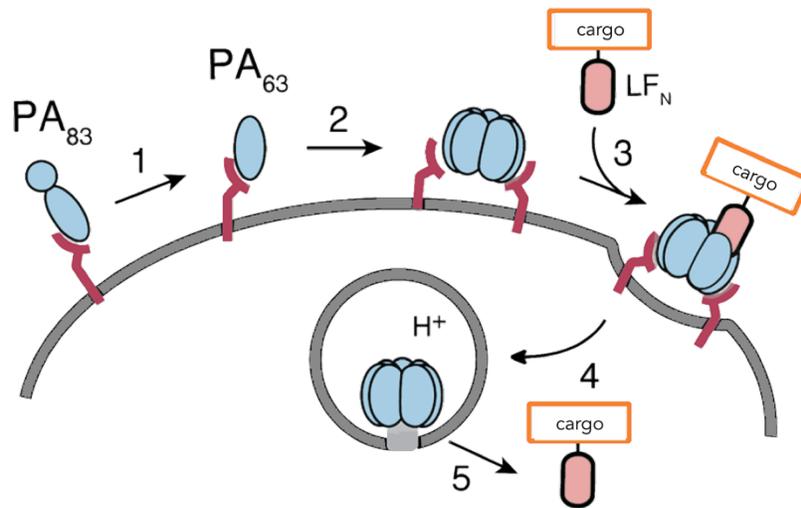


Figure 2.6. Delivery of cargo sortagged onto LF_N (or LF_N-DTA). Translocation of cargo conjugated to LF_N is mediated by protective antigen (PA) from anthrax toxin.

2.2.3.2 Optimization of western blot workflow

To support our immunoblot analysis, we needed highly sensitive, consistent, reproducible and as quantitative as possible data. We draw up a standardized and rigorous western blot protocol, allowing us to correlate and compare our outcomes across different experiment.⁷⁶ Therefore, we determined the span of signal intensities of our proteins, the linear dynamic range (Figure 2.6). A curve of several proteins extract loads versus band intensity has been produced, establishing the midpoint in order to generate bands linearly related to their protein abundance, without overestimating or underestimating real differences. We treated all antibodies in order to set the proper dilution factor required for each one and to verify the specific and selective interaction with the targeted antigen.

Also, we introduced Biotin on a lysine residue of LF_N. Each of the four subunit of streptavidin binds to one molecule of biotin with remarkably high affinity and the resulting complex is the strongest know non-covalent interaction between a protein and a ligand. The bond is unaffected by temperature, organic solvents and other denaturing agents. These features, combined with coupling high sensitivity and short reaction times, are particularly useful for detecting the protein.

We compared then multiple exposures in an effort to determine the optimal acquisition condition using a fluorescent-based-detection. In enzyme-substrate system (chemiluminescence) light is a result of a chemical reaction, depending on the oxidation catalyzed by horseradish peroxidase (HRP) of luminol in the presence of peroxide. By contrast, during fluorescent blotting reaction, light is emitted transiently by a fluorophore after excitation and then releases protons. This difference in signal optimization guaranties more accurate and quantitative results. Additionally, using fluorophore-conjugated antibodies multiple targets can be detected simultaneously on the same blot. This advantage has been used for data normalization. Ideally, each lane should contain the same amount of protein, but immunoblotting involves several numbers of handy steps. Normalization is then indispensable to identify and rectify any possible errors. We used Erk 1/2 to reduce the sample-to-sample source of variability and improve the accuracy of the results.

Blocking	Primary antibody incubation: Streptavidin	Primary antibody incubation: Anti-Erk 1/2	Secondary antibody incubation
2 h	o.n., 4°C ; 0.05% TBST	o.n., 4°C; 0.05% TBST	1h, RT; 0.05% TBST

*TBST: 150 mM NaCl, 20 mM Tris amino; pH 7.5; 0.05% Tween 20

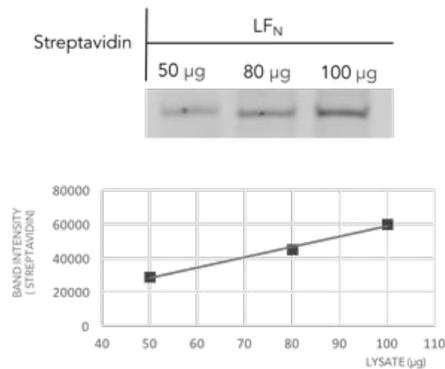


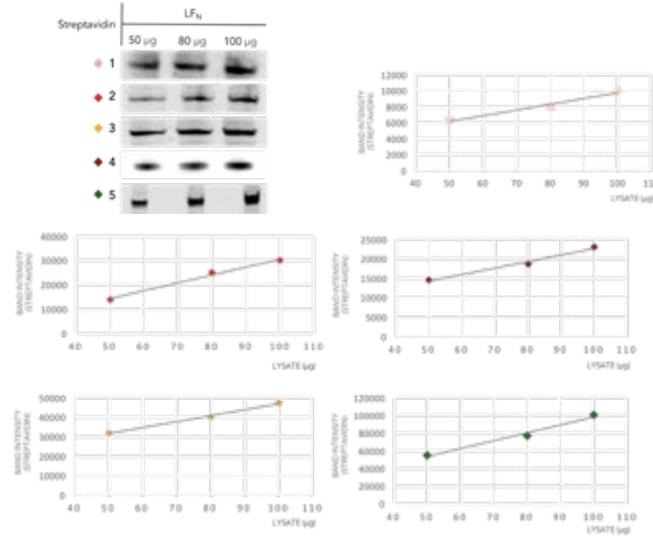
Figure 2.6. Determination of the linear dynamic range.

a

	Blocking	Primary antibody incubation: Streptavidin	Primary antibody incubation: Anti-Erk 1/2	Secondary antibody incubation
1	2 h	1h, RT in 0.05% TBST	1h, RT in 0.05% TBST	1h, RT in 0.05% TBST
2	2 h	2 h, RT in 0.05% TBST	2 h, RT in 0.05% TBST	1h, RT in 0.05% TBST
3	2 h	4h, RT in 0.05% TBST	4h, RT in 0.05% TBST	1h, RT in 0.05% TBST
4	2 h	o.n., 4°C; 0.05% TBST	o.n., 4°C; 0.05% TBST	1h, RT, 0.05% TBST
5	1 h	2 h, RT in TBST 0.05% + 5% Li-COR blocking buffer	2 h, RT in TBST 0.05% + 5% Li-COR blocking buffer	1h, RT in TBST 0.05% + 5% Li-COR blocking buffer

*TBST: 150 mM NaCl, 20mM Tris amino, pH 7.5, 0.05% Tween 20; Primary antibody concentration: 1/2000; Secondary antibody concentration: 1/5000

b



c

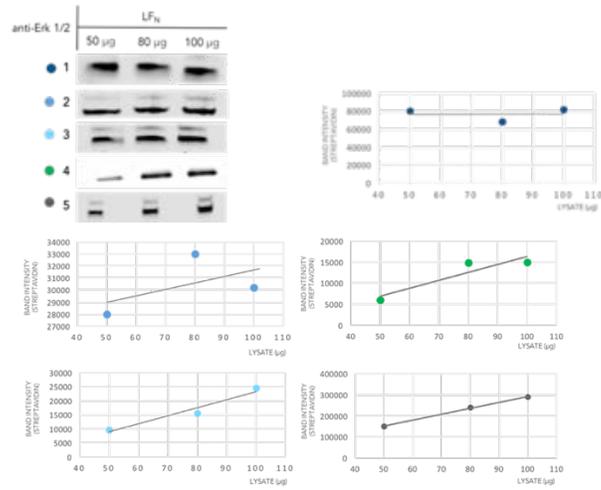


Figure 2.7. Optimization of Streptavidin and Erk1/2 incubation.

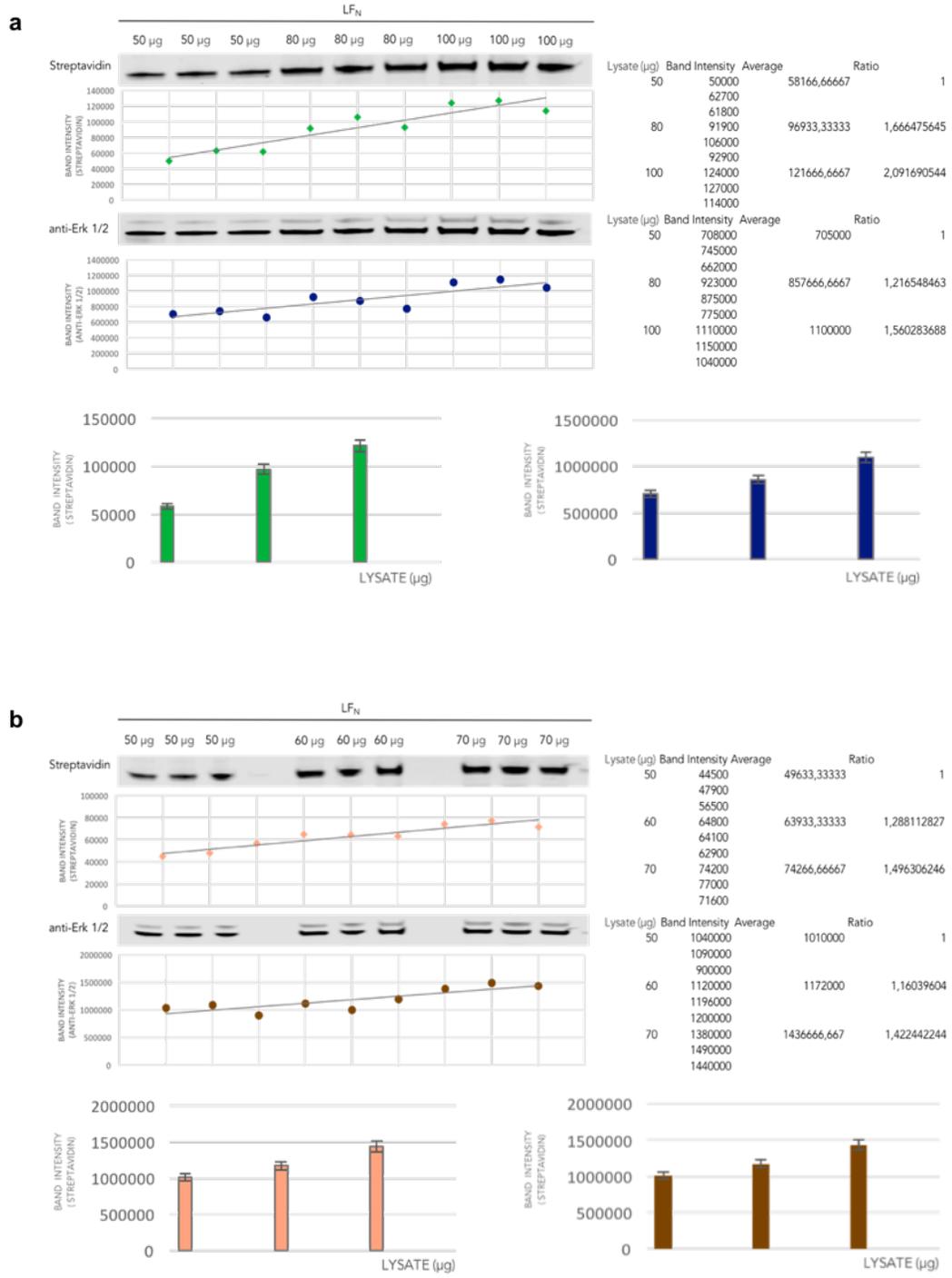


Figure 2.8. Compare signals.

We had the best results blocking the membrane for 1 h and incubating it with primary antibodies at RT for 2 h in 5% TBST + 5% Licor-buffer. Then we performed the incubation with the appropriate secondary antibody over 1 h at RT using 5% TBST + 5% Licor-buffer (Figure 2.7). Further, to examine whether densitometric ratios accurately reflect ratios between protein amounts, gels were loaded with incremental amounts (50, 80 and 100 μg ; 50, 60 and 70 μg) of total cellular proteins from control samples. Each loading was performed in triplicate, and all the loadings were applied to the same gel to insure identical treatment during the western blotting procedure. In addition, to avoid band distortion, sample loadings were brought to the same final volume with sample buffer. The densitometric values of LF_N bands in the incremental loadings of control samples were plotted against the corresponding total lysate protein amounts. We found that the average of the band intensity resulting from each triplicate samples was mathematically proportional to the loaded amount (Figure 2.8).

2.2.3.3 Western Blot Analysis with LF_N-DTA-X and LF_N-X Constructs

Strengthening our western blot workflow, we could convert each band intensity into a reliable and meaningful assessment of abundance. Our goal was to relate the rate of proteasomal degradation of proteins containing mixed chirality peptides.

Immunoblot analyses showed significant differences in cytosolic proteins levels highlighting the substantial impact of the presence of D-amino acid on the polypeptide stretch stability. The results underscored that LF_N -^LX constructs were rapidly degraded in the cell cytosol (Figure 2.9a); nevertheless, LF_N -^DX conjugates reported to be stabilized to degradation (Figure 2.9b). We demonstrated that the stabilization is not protein specific: our data suggest that the presence of the ^LX and ^DX stretches attached to LF_N-DTA affects, in the same way, the protein half-life. The samples treated with unstructured D-peptide payloads on the C-terminus of LF_N (Figure 2.9) or LF_N-DTA (Figure 2.10) all showed Streptavidin or anti-LF, respectively, strong bands and their signal intensity were equivalent to that resulting from the treatment with Lactacystin.

It was found that the C-terminus group plays a role in the degradation pathway: peptides presenting a -CONH₂ as a C-termini are less stable compared to those with the -COOH. Protein quantification via Western blotting allowed us to detect even this feature, while using an SRM-based quantification we wouldn't be able to ascertain such a small difference in the molecular weight of the targeted proteins.

Moreover, capping the C-terminus with one or two amino acids (Dcap) stabilize the L-peptide cargo (Figure 2.9c), while the presence of one random mirror image amino acid in the mid of the sequence doesn't rescue the conjugate from the degradation (Figure 2.9c).

2.2.3.4 Protein Synthesis Inhibition Assay with LF_N-DTA-X Constructs

Aimed at understanding the structural requirements of the translocation mechanism of mixed chirality peptide cargos and proving the efficiency and adaptability of PA/LF_N delivery platform, we generated a chimeric protein composed of the PA recognition domain of LF (LF_N) fused to the amino terminus of the catalytic A chain of diphtheria toxin (DTA).

The diphtheria toxin (DTX) (58 kDa), secreted by *Corynebacterium Diphtheria* as a single mature precursor (535 residues), is a three domain protein. The N-terminal protein represents the A fragment and is responsible for the catalytic properties; linked to this by a disulfide loop, the C-terminal domain, termed B fragment consist of two regions, receptor binding domain (R) and the translocation domain (T), which binding to its receptor on the eukaryotic surface cells and arranging the hydrophobic insertion into the endosome bilayer membrane ensure the release of A.

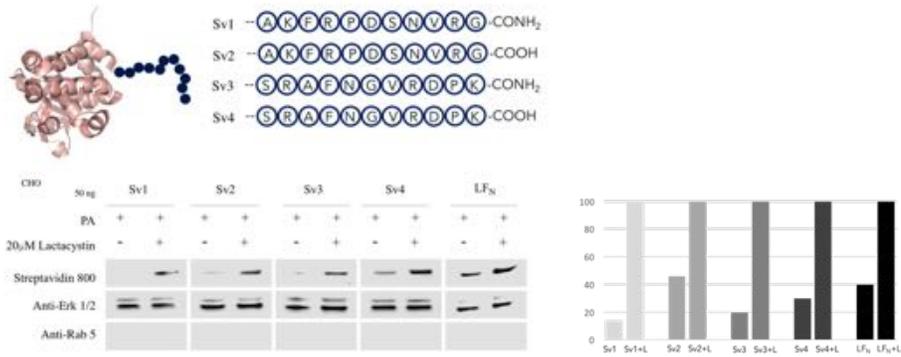
As a result of the proteolytic nicking, at the trypsin-sensitive site and the consequent reduction of the disulfide bridge, the free DTA catalyzes the transfer of ADP-ribose portion of NAD to a posttranslationally modified histidine, diphthamide, of the elongation factor 2 (EF2), which it thereby become no longer able of participating in protein synthesis.

Given that DTA or LF_N alone has no effect on cell, it's been taken advantage of the ability of the assembled protein to inhibit protein synthesis in the presence of PA with the purpose of test their translocation into the cytosol.

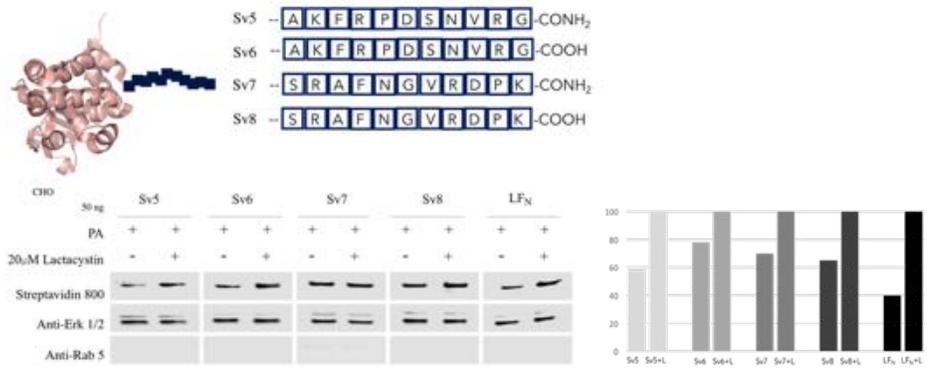
For the protein synthesis inhibition assay, Chinese hamster ovary (CHO-K1) cells were treated with 10-fold serial dilutions of each construct for 6 h to allow for sufficient buildup of the translocated material in the cytosol and to observe DTA activity. After translocation, the cells were washed and treated with ³H-Leu in leucine-free medium for 1 h to detect DTA activity. The fraction of protein synthesis with respect to DTA was measured with a scintillation counter.

Unfortunately, DTA wild-type is likely to dispatch its substrate in minutes and didn't allowed us to value any variability between the constructs translocation (Figure 2.10a). Indeed, a more suitable translocation profile is available incubating cells over 30 min (Figure 2.10b). For this reason, for

a



b



c

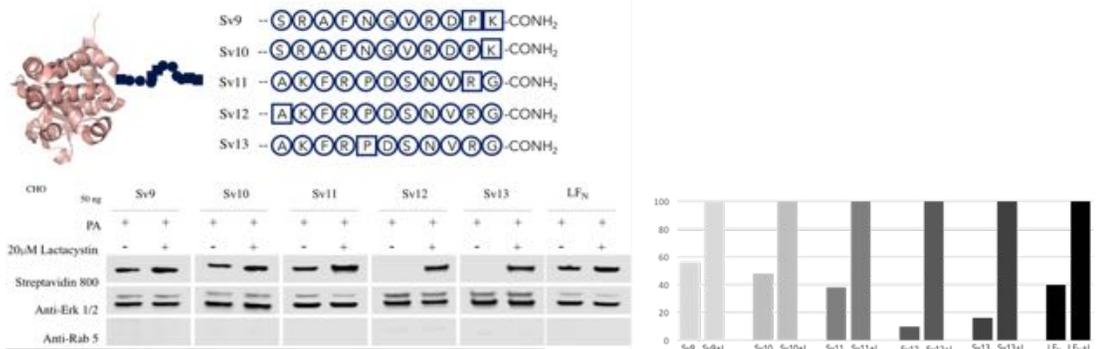


Figure 2.9. Degradation experiment for Sv's.
 L-amino acids are represented by circles; D-amino acids are represented by squares.

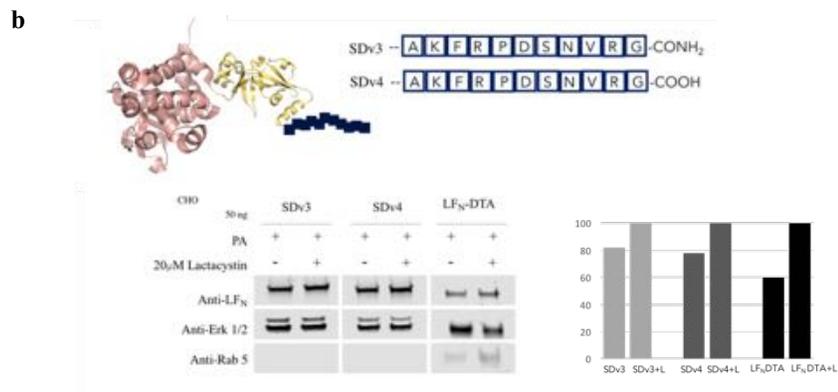
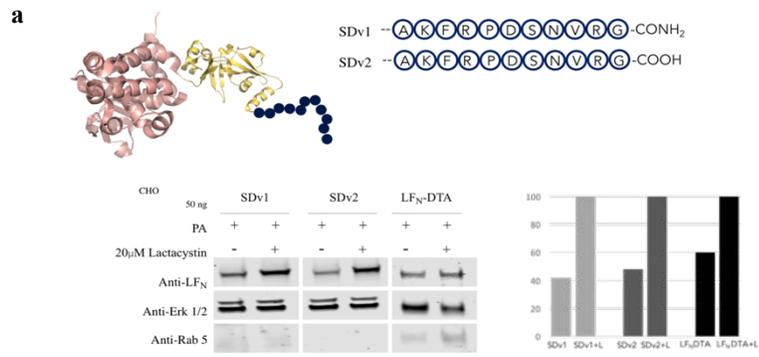


Figure 2.10. Degradation experiment for SDv's.
 L-amino acids are represented by circles; D-amino acids are represented by squares.

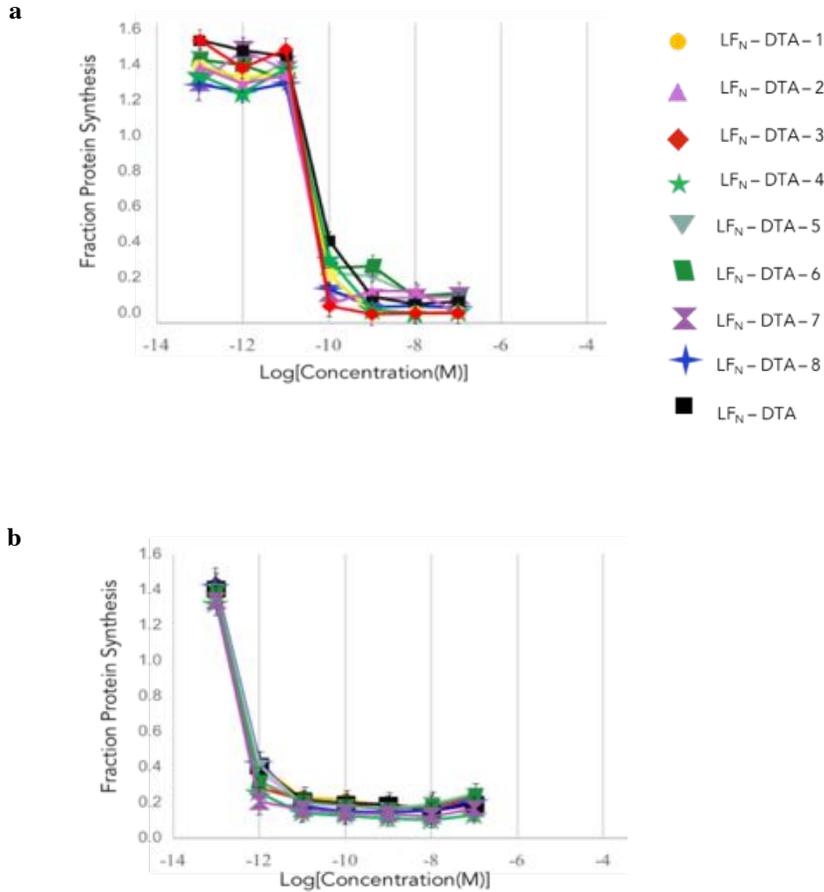


Figure 2.10. Control experiments validating the translocation mechanism of LF_N -DTA constructs (SDv's). Protein synthesis inhibition in CHO-K1 cells treated with variants in the presence of 20 nm PA for a) 6h; b) 30 min.; incubation at 37 °C.

our experiments we decide to use a mutant form of DTA (E148S, DTA_{mut}), by which it's possible determining differences in cytosolic lifetime of each LF_N - DTA_{mut} -X over a wider dynamic range. The point mutation involves, in fact, the active site glutamic acid residue drastically reducing (~ 300 fold) ADP-ribosyltransferase activity compared to the wild-type.

So far, the mutation has been performed on LF_N -DTA and the resulting protein LF_N - DTA_{mut} has been purified. Thus, we have to repeat the experiment using the mutant form of the protein over the

6 h of incubation. This experiment will remarkably provide that the degradation effect of mixed chirality peptides is not an artifact of the translocation.

2.3 Conclusion

The translocation of functional cell-impermeable substance into the cytosol of mammalian cells is a key issue and several methods have been developed to overcome this restriction for both basic research and therapeutic applications.

In this work, the versatile LF_N/PA platform has been employed for efficiently deliver mirror image polypeptides across the bilayer phospholipid membrane. Mixed chirality cargo featuring on N-terminal oligoglycine motif have been attached to the C-terminus of LF_N or LF_N-DTA containing the LPXTG recognition site via SrtA*-mediated ligation. Sortagged DTA variants confirmed the internalization of the chimera proteins blocking the protein synthesis once in the cytosolic compartment. Further, immunoblot analysis of the cytosolic delivered fractions underlined the high intracellular stability of variants featuring a mirror image sequence or a D-cap at the C-terminus. Indeed, acting as an unstructured region, the *L*-sequence promoted proteasomal-mediated degradation, while the *D*-cargo rescued the protein from the proteolysis suggesting that stereospecific interactions are required for the accessibility to the narrow catalytic pore.

Moreover, we demonstrated that a sequence of eleven amino acid in length is sufficient to work as an initiation site, disproving the results collected so far in cell free assays.

Several questions remain unanswered regarding the rate of protein degradation, including the impact of the amino acid composition of the tails.

Scouting the proteolytic machinery in cells will provide new insight into the potential therapeutic use of engineered proteins.

2.4 References

1. Tomko, R.J.; Hochstrasser, M. Molecular Hollon T. Delivery Technologies for Therapeutic Proteins: Assessment and Outlook. Pastores, G. M. Therapeutic approaches for lysosomal storage diseases. *Therapeutic Advances in Endocrinology and Metabolism* **2010**, 1 (4), 177–188.
2. Pastores, G. M. Therapeutic approaches for lysosomal storage diseases. *Therapeutic Advances in Endocrinology and Metabolism* **2010**, 1 (4), 177–188.
3. (a) Dimitroy, D.S. Therapeutic proteins. *Methods Mol Biol.* **2012**, 899, 1-26. (b) Campbell, P.L.; McCluskey, J.; Yeo, J.P.; Toh, B.H. Electro-poration of antibodies into mammalian cells. *Methods Mol Biol* **1995**, 48, 83–92.
4. Bar-Sagi, D.; Feramisco, J.R. Microinjection of the ras oncogene protein into PC12 cells induces morphological differentiation. *Cell* **1985**, 42, 841–848.
5. Morris, M.C.; Depollier, J.; Mery, J.; Heitz, F.; Divita, G. A peptide carrier for the delivery of biologically active proteins into mammalian cells. *Nat Biotechnol* **2001**, 19, 1173–1176.
6. Prochiantz, A. Messenger proteins: homeoproteins, TAT and others. *Curr Opin Cell Biol* **2000**, 12, 400–406
7. Raptis, L.H.; Liu, S.; Firth, K.L.; Stiles, C.D.; Alberta, J.A. Electroporation of peptides into adherent cells in situ. *Biotechniques* **1995**, 18, 104, 106, 108, 110 passim
8. Schwarze, S.R.; Dowdy, S.F. In vivo protein transduction: intracellular delivery of biologically active proteins, compounds and DNA. *Trends Pharmacol Sci* **2000**, 21, 45–48.
9. Tinsley, J.H.; Hawker, J.; Yuan, Y. Efficient protein transfection of cultured coronary venular endothelial cells. *Am J Physiol* **1998**, 275, H1873–H1878.
10. Zelphati, O.; Wang, Y.; Kitada, S.; Reed, J.C.; Felgner, P.L.; Corbeil, J. Intracellular delivery of proteins with a new lipid-mediated delivery system. *J Biol Chem* **2001**, 276, 35103–35110.
11. Fu, A.; Tang, R.; Hardie, J.; Farkas, M.E.; Rotello, V.M. Promises and Pitfalls of Intracellular Delivery of Proteins. *Bioconjug Chem.* **2014**, 25(9), 1602-1608.
12. Choi, S. O.; Kim, Y. C.; Lee, J. W.; Park, J. H.; Prausnitz, M. R.; Allen, M. G. Intracellular protein delivery and gene transfection by electroporation using a microneedle electrode array. *Small* **2012**, 8, 1081–1091.
13. Daugimont, L.; Baron, N.; Vandermeulen, G.; Pavselj, N.; Miklavcic, D.; Jullien, M.C.; Cabodevila, G.; Mir, L.M; Préât, V. Hollow microneedle arrays for intradermal drug delivery and DNA electroporation. *J. Membr. Biol.* **2010**, 236, 117–125.
14. Sharei, A.; Zoldan, J.; Adamo, A.; Sim, W. Y.; Cho, N.; Jackson, E.; Mao, S.; Schneider, S.; Han, M. J.; Lytton-Jean, A.; Basto, P. A.; Jhunjhunwala, S.; Lee, J.; Heller, D. A.; Kang, J. W.; Hartoularos, G. C.; Kim, K. S.; Anderson, D. G.; Langer, R.; Jensen, K. F. A vector-free microfluidic platform for intracellular delivery. *Proc. Natl. Acad. Sci.U.S.A.* **2013**, 110, 2082–2087.
15. Jo, J.; Hong, S.; Choi, W. Y.; and Lee, D. R. Cell-penetrating peptide (CPP)-conjugated proteins is an efficient tool for manipulation of human mesenchymal stromal cells. *Sci. Rep.* **2014**, 4, 4378.
16. Wadia, J. S.; Stan, R. V.; Dowdy, S. F. Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med.* **2004**, 10, 310–315.
17. Madani, F.; Lindberg, S.; Langel, U.; Futaki, S.; Grašlund, A. Mechanisms of cellular uptake of cell-penetrating peptides. *J. Biophys.* **2011**, 2011, 414729.
18. Duchardt, F.; Fotin-Mleczek, M.; Schwarz, H.; Fischer, R.; Brock, R. A comprehensive model for the cellular uptake of cationic cell-penetrating peptides. *Traffic* **2007**, 8, 848–866.
19. Brock, R. The uptake of arginine-rich cell-penetrating peptides: putting the puzzle together. *Bioconjugate Chem.* 2014, 25, 863–868.

20. Rodrigues, M.; de la Torre, B. G.; Radis-Baptista, G.; Santos, N. C.; Andreu, D. Efficient cellular delivery of β -galactosidase mediated by NrTPs, a new family of cell-penetrating peptides. *Bioconjugate Chem.* **2011**, *22*, 2339–2344.
21. Gaj, T.; Guo, J.; Kato, Y.; Sirk, S. J.; Barbas, C. F. 3rd. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nat. Methods* **2012**, *9*, 805–807.
22. Bersani, S.; Salmaso, S.; Mastrotto, F.; Ravazzolo, E.; Semenzato, A.; Caliceti, P. Starlikeoligo-arginyl-maltotriosyl derivatives as novel cell-penetrating enhancers for the intracellular delivery of colloidal therapeutic systems. *Bioconjugate Chem.* **2012**, *23*, 1415–1425.
23. Zhao, Q.; Chen, W.; Chen, Y.; Zhang, L.; Zhang, J.; Zhang, Z. Self-assembled virus-like particles from rotavirus structural protein VP6 for targeted drug delivery. *Bioconjugate Chem.* **2011**, *22*, 346–352.
24. Muratori, C.; Bona, R.; Federico, M. Lentivirus- based virus-like particles as a new protein delivery tool. *Methods Mol. Biol.* **2010**, *614*, 111–124.
25. Kaczmarczyk, S. J.; Sitaraman, K.; Young, H. A.; Hughes, S. H.; Chatterjee, D. K. Protein delivery using engineered virus- like particles. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16998–17003.
26. Cronican, J. J.; Beier, K. T.; Davis, T. N.; Tseng, J. C.; Li, W.; Thompson, D. B.; Shih, A. F.; May, E. M.; Cepko, C. L.; Kung, A. L.; Zhou, Q.; Liu, D. R. A class of human proteins that deliver functional proteins into mammalian cells in vitro and in vivo. *Chem. Biol.* **2011**, *18*, 833–838.
27. Algar, W. R.; Prasuhn, D. E.; Stewart, M. H.; Jennings, T. L.; Blanco-Canosa, J. B.; Dawson, P. E.; Medintz, I. L. The controlled display of biomolecules on nanoparticles: a challenge suited to bioorthogonal chemistry. *Bioconjugate Chem.* **2011**, *22*, 825–858.
28. Rao, N. V.; Ganivada, M. N.; Sarkar, S.; Dinda, H.; Chatterjee, K.; Dalui, T.; Das Sarma, J.; Shunmugam, R. Magnetic norbornene polymer as multiresponsive nanocarrier for site specific cancer therapy. *Bioconjugate Chem.* **2014**, *25*, 276–285.
29. (a) Doane, T.; Burda, C. Nanoparticle mediated non- covalent drug delivery. *Adv. Drug Delivery Rev.* **2013**, *65*, 607–621. (b) Kim, J. Y.; Choi, W. I.; Kim, Y. H.; Tae, G. Brain-targeted delivery of protein using chitosan- and RVG peptide- conjugated, pluronic-based nano-carrier. *Biomaterials* **2012**, *34*, 1170–1178.
30. Sarker, S. R.; Hokama, R.; Takeoka, S. Intracellular delivery of universal proteins using a lysine headgroup containing cationic liposomes: deciphering the uptake mechanism. *Mol. Pharmacol.* **2014**, *11*, 164–174.
31. Furuhashi, M.; Kawakami, H.; Toma, K.; Hattori, Y.; Maitani, Y. Intracellular delivery of proteins in complexes with oligoarginine-modified liposomes and the effect of oligoarginine length. *Bioconjugate Chem.* **2006**, *17*, 935–942.
32. Li, P.; Nielsen, H. M.; Müllertz, A. Oral delivery of peptides and proteins using lipid-based drug delivery systems. *Expert Opin. Drug Delivery* **2012**, *9*, 1289–1304.
33. Shaik, T.; Sundaram, M.A.; Umasankar, K. Review on Gastro-retentive Drug Delivery System. *Pharmaceutical and Nano Sciences.* **2014**, *3*(3), 177-185.
34. Ryou, J.H.; Sohn, Y.K.; Hwang, D.E. Engineering of bacterial exotoxins for highly efficient and receptor-specific intracellular delivery of diverse cargos. *Biotechnol Bioeng.* **2016**, *113*(8), 1639-1646.
35. Guillard, S.; Minter, R.R.; Jackson, R.H. Engineering therapeutic proteins for cell entry: The natural approach. *Trends Biotechnol.* **2015**, *33*(3), 163-171.
36. Mohammed, A.F.; Abdul-Wahid, A.; Huang, E.H.; Bolewska-Pedyczak, E.; Cydzik, M.; Broad, A.E.; Gariépy, J. The *Pseudomonas aeruginosa* exotoxin A translocation domain facilitates the routing of CPP-protein cargos to the cytosol of eukaryotic cells. *J Control Release.* **2012**, *164*(1), 58-64.

37. Chen, C.; Przedpelski, A.; Tepp, W. H.; Pellett, S.; Johnson, E. A.; Barbieri, J. T. Heat-Labile Enterotoxin IIa, a Platform To Deliver Heterologous Proteins into Neurons. *mBio* **2015**, 6(4), e00734–15.
38. Rabideau, A.E.; Pentelute, B.L. Delivery of Non-Native Cargo into Mammalian Cells Using Anthrax Lethal Toxin. *ACS Chem Biol.* **2016**, 11(6), 1490-1501.
39. Young, J. A. T.; Collier, R. J. Anthrax toxin: Receptor binding, internalization, pore formation, and translocation. *Annu. Rev. Biochem.* **2007**, 76, 243–265.
40. Scobie, H. M.; Rainey, G. J. A.; Bradley, K. A.; Young, J. A. T. Human capillary morphogenesis protein 2 functions as an anthrax toxin receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, 100, 5170–5174.
41. Bradley, K. A.; Mogridge, J.; Mourez, M.; Collier, R. J.; Young, J. A. T. Identification of the cellular receptor for anthrax toxin. *Nature* **2001**, 414, 225–229.
42. Lacy, D. B.; Wigelsworth, D. J.; Scobie, H. M.; Young, J. A. T.; Collier, R. J. Crystal structure of the von Willebrand factor A domain of human capillary morphogenesis protein 2: An anthrax toxin receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 6367–6372.
43. Abi-Habib, R. J.; Urieto, J. O.; Liu, S. H.; Leppla, S. H.; Duesbery, N. S.; Frankel, A. E. BRAF status and mitogen- activated protein/extracellular signal-regulated kinase kinase 1/2 activity indicate sensitivity of melanoma cells to anthrax lethal toxin. *Mol. Cancer Ther.* **2005**, 4, 1303–1310.
44. Kintzer, A. F.; Thoren, K. L.; Sterling, H. J.; Dong, K. C.; Feld, G. K.; Tang, I.I.; Zhang, T. T.; Williams, E. R.; Berger, J. M.; and Krantz, B. A. The protective antigen component of anthrax toxin forms functional octameric complexes. *J. Mol. Biol.* **2009**, 392, 614–629.
45. Klimpel, K. R.; Molloy, S. S.; Thomas, G.; Leppla, S. H. Anthrax toxin protective antigen is activated by a cell-surface protease with the sequence specificity and catalytic properties of furin. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, 89, 10277–10281.
46. Milne, J. C.; Furlong, D.; Hanna, P. C.; Wall, J. S.; Collier, R. J. Anthrax protective antigen forms oligomers during intoxication of mammalian-cells. *J. Biol. Chem.* **1994**, 269, 20607–20612.
47. Petosa, C.; Collier, R. J.; Klimpel, K. R.; Leppla, S. H.; Liddington, R. C. Crystal structure of the anthrax toxin protective antigen. *Nature* **1997**, 385, 833–838.
48. Feld, G.K.; Brown, M.J.; Krantz, B.A. Ratcheting up protein translocation with anthrax toxin. *Protein Sci.* **2012**, 21(5), 606-624.
49. Arora, N.; Leppla, S.H. Residues 1-254 of anthrax toxin lethal factor are sufficient to cause cellular uptake of fused polypeptides. *J Biol Chem.* **1993**, 268(5), 3334-3341.
50. Brown, M.J.; Thoren, K.L.; Krantz, B.A. Charge requirements for proton gradient-driven translocation of anthrax toxin. *J Biol Chem.* **2011**, 286(26), 23189-23199.
51. Thoren, K.L.; Krantz, B.A. The unfolding story of anthrax toxin translocation. *Mol Microbiol.* **2011**, 80(3), 588-595.
52. Jiang, J.; Pentelute, B.L.; Collier, R.J.; Zhou, Z.H. Atomic structure of anthrax protective antigen pore elucidates toxin translocation. *Nature.* **2015**, 521(7553), 545-549.
53. Arora, N.; Leppla, S. H. Fusions of anthrax toxin lethal factor with shiga toxin and diphtheria toxin enzymatic domains are toxic to mammalian cells. *Infect. Immun.* **1994**, 62, 4955–4961.
54. Milne, J. C.; Blanket, S. R.; Hanna, P. C.; Collier, R. J. Protective antigen-binding domain of anthrax lethal factor mediates translocation of a heterologous protein fused to its amino- or carboxy-terminus. *Mol. Microbiol.* **1995**, 15, 661–666.
55. Wesche, J.; Elliott, J. L.; Falnes, P. O.; Olsnes, S.; Collier, R. J. Characterization of membrane translocation by anthrax protective antigen. *Biochemistry* **1998**, 37, 15737–15746.
56. Arora, N.; Klimpel, K. R.; Singh, Y.; Leppla, S. H. Fusions of anthrax toxin lethal factor to the ADP-ribosylation domain of Pseudomonas exotoxin A are potent cytotoxins which are translocated to the cytosol of mammalian cells. *J. Biol. Chem.* **1992**, 267, 15542–15548.

57. Hu, H. J.; Leppla, S. H. Anthrax Toxin Uptake by Primary Immune Cells as Determined with a Lethal Factor-beta- Lactamase Fusion Protein. *PLoS One* **2009**, 4, e7946.
58. Wesche, J.; Elliott, J.L.; Falnes, P.O.; Olsnes, S.; Collier, R.J. Characterization of membrane translocation by anthrax protective antigen. *Biochemistry* **1998**, 37(45), 15737-46.
59. Liao, X.; Rabideau, A.E.; Pentelute, B.L. Delivery of antibody mimics into mammalian cells via anthrax toxin protective antigen. *ChemBioChem*. **2014**, 15(16), 2458-2466.
60. McMaster, M.L.; Kristinsson, S.Y.; Turesson, I.; Bjorkholm, M.; Landgren, O. Novel aspects pertaining to the relationship of Waldenström macroglobulinemia, IgM monoclonal gammopathy of undetermined significance (MGUS), polyclonal gammopathy and hypoglobulinemia. *Clin Lymphoma* **2010**, 9(1), 19-22.
61. Pentelute, B. L.; Sharma, O.; Collier, R. J. Chemical dissection of protein translocation through the anthrax toxin pore. *Angew. Chem., Int. Ed.* **2011**, 50, 2294–2296.
62. Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. Sortagging: a versatile method for protein labeling. *Nat. Chem. Biol.* **2007**, 3, 707–708.
63. Schneewind, O.; Missiakas, D.M. Protein secretion and surface display in Gram-positive bacteria. *Philos.trans.r.soc.l B BiolSci.* **2012**, 19;367(1592), 1123-1139.
64. Marraffini, L.; DeDent, A.C.; Schneewind, O. Sortases and the Art of Anchoring Proteins to the Envelopes of Gram-Positive Bacteria. *Microbiol Mol Biol Rev.* **2006**, 70(1), 192-221.
65. Schneewind, O.; Fowler, A.; Faull, K.F. Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. *Science* **1995**, 268(5207), 103-106.
66. Ilangovan, U.; Ton-That, H.; Iwahara, J.; Schneewind, O.; Clubb, R.T. Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* **2001**, 98(11), 6056-6061.
67. Piotukh, K.; Geltinger, B.; Heinrich, N.; Gerth, F.; Beyermann, M.; Freund, C.; Schwarzer, D. Directed evolution of sortase A mutants with altered substrate selectivity profiles. *J Am Chem Soc.* **2011**, 133(44), 17536-17539.
68. Chen, L.; Cohen, J.; Song, X. Improved variants of SrtA for site- specific conjugation on antibodies and proteins with high efficiency. *Sci Rep.* **2016**, 1-12.
69. Chen, I.; Dorr, B. M.; Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, 108, 11399–11404.
70. Schneewind, O.; Fowler, A.; Faull, K.F. Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. *Science* **1995**, 268(5207), 103.
71. Milton, R. C. D.; Milton, S. C. F.; Kent, S. B. H. Total Chemical Synthesis of a D-Enzyme - the Enantiomers of Hiv-1 Protease Show Demonstration of Reciprocal Chiral Substrate-Specificity. *Science* **1992**, 256, 1445–1448.
72. Collier, R. J.; Kandel, J. Structure and activity of diphtheria toxin. Thiol-dependent dissociation of fraction of toxin into enzymically active and inactive fragments. *J. Biol. Chem.* **1971**, 246, 1496.
73. Baghirova, S.; Hughes, B.G.; Hendzel, M.J.; Schulz, R. Sequential fractionation and isolation of subcellular proteins from tissue or cultured cells. *MethodsX* **2015**, 2, e440-e445.
74. Kisselev, A.F.; Goldberg, A.L. Proteasome inhibitors: from research tools to drug candidates. *Chem Biol.* **2001**, 8, 739-758.
75. Li-cor Biosciences. Western Blot Normalization: Challenges and Considerations for Quantitative Analysis.

2.5 Experimental Section

Materials. Peptides were synthesized using Fmoc-protected L- and D-amino acids, N,N,N',N'-tetramethyl-O- (1H-benzotriazol-1-yl) uranium hexafluorophosphate (HBTU), and 1-[bis-(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]-pyridinium 3-oxid hexafluorophosphate (HATU) purchased from Creosalus and ChemImpex. Dimethylformamide, piperidine, diisopropylethylamine, trifluoroacetic acid, and triisopropylsilane were purchased from VWR or Sigma-Aldrich. All cloning was accomplished using the QuikChange Lightning kit (Agilent) or HiFi DNA Taq Polymerase (LifeTechnologies) and pET SUMO Champion kit (LifeTechnologies). All proteins were expressed in BL21(DE3) from LifeTechnologies. All medium for tissue culture was from LifeTechnologies. For Western blots were used nitrocellulose membranes 0.2 μm (Bio-Rad), filters (BioRad), and Trans-Blot@Turbo™ Transfer Buffer (Bio-Rad), Odyssey blocking buffer - PBS (LI-COR), were used. We used the following primary and secondary antibodies: LF (Santa Cruz Biotechnology), Erk1/2 (Cell Signaling Technology), Rab5 (Cell Signaling Technology), goat anti-mouse IRdye800 (LI-COR), goat anti-rabbit IRdye680, and streptavidin IRdye800 (LI-COR). Unless specified otherwise, all other reagents were purchased from VWR, Sigma-Aldrich, or LifeTechnologies. All commercially available materials (Aldrich®, TCI®, Novabiochem®, Fluka®) were used without further purification.

2.5.1 Chemistry

Peptide synthesis and purification. Designed peptide sequences were synthesized under automated flow coupling/deprotection conditions following a standard Fmoc/tBu protocol for SPPS. In figure 1 is depicted the synthesizer. During a coupling reaction, the pumps draw reagents from the storage module, and mix the desired amino acid with diisopropylethylamine (DIEA), and an activating agent in the mixing module. This mixture flows through the activation module, an electrically heated tubular reactor, where it quickly heats to 90°C and forms an active ester. Within two seconds of activation, the activated amino acid flows through the coupling module, a packed bed of peptide synthesis resin maintained at 90°C, where amide bond formation is complete within seven seconds. Fmoc deprotection was carried out with 20% piperidine in DMF, and coupling reactions were performed with a mixture of the opportune Fmoc-amino acid solution 0.2 M, HATU 0.17 M as activating agent and 5 v/v% DIEA in DMF over 7s. Couplings with Arg were

accomplished using PyAOP as the activator and over 23s. In addition, regarding the synthesis of COOH-Terminal Peptides, the first coupling was performed on the manifold employing 2-Cl-trtCl resin (1.60 mmol/g), 4 mmol Fmoc-amino acid, 650 μ l DIEA in 5ml of DMF over 30min. the first deprotection was carried out with 20% Piperidine solution in DMF over 25 min. Thus, the synthesis was taken up using the automated synthesizer. Following chain assembly, global deprotection and cleavage from the resin was carried out with 94% TFA with added 1% triisopropylsilane (TIS), 2.5% 1,2-Ethanedithiol (EDT) and 2.5% H₂O. After cleavage, TFA was dried under N₂(g) and the crude peptides were triturated three times with cold ether then dissolved in 50:50 A:B and lyophilized. The products were purified by reversed-phase HPLC using a semi-preparative Agilent Zorbax C18 column at a flow rate of 4 ml min⁻¹ at 1-61% of aqueous TFA and CH₃CN containing 0.1% TFA over 80 min. The peptide was identified by MALDI-TOF MS.

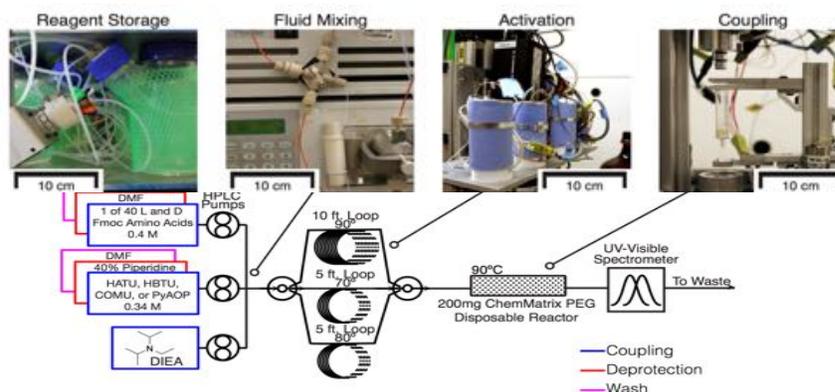


Figure 1. The Automated Flow Peptide Synthesizer consists of five modules

1: Purity >95%, tR 26.2 min (analytical HPLC, 1–61% ACN in H₂O (0.1% TFA) over 80 min, flow rate of 4.0 mL/min); molecular formula, C₆₃H₁₀₂N₂₄O₂₀; calculated mass, 1530.647; found, 765.8983 (M + 2H⁺)/2.

2: Purity >95%, tR 27.4 min (analytical HPLC, 1–61% ACN in H₂O (0.1% TFA) over 80 min, flow rate of 4.0 mL/min); molecular formula, C₆₃H₁₀₂N₂₄O₂₀; calculated mass, 1530.647; found, 765.8981 (M + 2H⁺)/2.

3: Purity >95%, tR 28.2 min (analytical HPLC, 1–61% ACN in H₂O (0.1% TFA) over 80 min, flow rate of 4.0 mL/min); molecular formula, C₆₃H₁₀₂N₂₄O₂₁; calculated mass, 1531.647; found, 766.3905 (M + 2H⁺)/2.

4: Purity >95%, tR 25.8 min (analytical HPLC, 1–61% ACN in H₂O (0.1% TFA) over 80 min, flow rate of 4.0 mL/min); molecular formula, C₆₃H₁₀₂N₂₄O₂₁; calculated mass, 1531.647; found, 765.8979 (M + 2H⁺)/2.

5: Purity >95%, tR 26.5 min (analytical HPLC, 1–61% ACN in H₂O (0.1% TFA) over 80 min, flow rate of 4.0 mL/min); molecular formula, C₆₃H₁₀₂N₂₄O₂₀; calculated mass, 1530.647; found, 766.3905 (M + 2H⁺)/2.

6: Purity >95%, tR 27.2 min (analytical HPLC, 1–61% ACN in H₂O (0.1% TFA) over 80 min, flow rate of 4.0 mL/min); molecular formula, C₆₃H₁₀₂N₂₄O₂₀; calculated mass, 1530.647; found, 765.8965 (M + 2H⁺)/2.

7: Purity >95%, tR 28.2 min (analytical HPLC, 1–61% ACN in H₂O (0.1% TFA) over 80 min, flow rate of 4.0 mL/min); molecular formula, C₆₃H₁₀₂N₂₄O₂₁; calculated mass, 1531.647; found, 765.8949 (M + 2H⁺)/2.

8: Purity >95%, tR 26.8 min (analytical HPLC, 1–61% ACN in H₂O (0.1% TFA) over 80 min, flow rate of 4.0 mL/min); molecular formula, C₆₃H₁₀₂N₂₄O₂₁; calculated mass, 1531.647; found, 766.2965 (M + 2H⁺)/2.

9: Purity >95%, tR 27.6 min (analytical HPLC, 1–61% ACN in H₂O (0.1% TFA) over 80 min, flow rate of 4.0 mL/min); molecular formula, C₆₃H₁₀₂N₂₄O₂₀; calculated mass, 1530.647; found, 766.3944 (M + 2H⁺)/2.

10: Purity >95%, tR 28.4 min (analytical HPLC, 1–61% ACN in H₂O (0.1% TFA) over 80 min, flow rate of 4.0 mL/min); molecular formula, C₆₃H₁₀₂N₂₄O₂₀; calculated mass, 1530.647; found, 765.8896 (M + 2H⁺)/2.

11: Purity >95%, tR 27.8 min (analytical HPLC, 1–61% ACN in H₂O (0.1% TFA) over 80 min, flow rate of 4.0 mL/min); molecular formula, C₆₃H₁₀₂N₂₄O₂₀; calculated mass, 1530.647; found, 765.9979 (M + 2H⁺)/2.

12: Purity >95%, tR 25.2 min (analytical HPLC, 1–61% ACN in H₂O (0.1% TFA) over 80 min, flow rate of 4.0 mL/min); molecular formula, C₆₃H₁₀₂N₂₄O₂₀; calculated mass, 1530.647; found, 765.8903 (M + 2H⁺)/2.

13: Purity >95%, tR 26.2 min (analytical HPLC, 1–61% ACN in H₂O (0.1% TFA) over 80 min, flow rate of 4.0 mL/min); molecular formula, C₆₃H₁₀₂N₂₄O₂₀; calculated mass, 1530.647; found, 766.3908 (M + 2H⁺)/2.

Sortase A Mediated Ligation of LF_N-X and LF_N-DTA-X Constructs. The constructs were synthesized using the one-pot SrtA*-mediated ligation method.

N-terminal small ubiquitin-like modifier (SUMO) was first cleaved from the protein (His₆-SUMO-LF_N-LPSTGG-His₆ or His₆-SUMO-LF_N-DTA-LPSTGG-His₆), using 1 μg SUMO protease per mg of protein substrate for 1 h at RT to give the native amino termini of the substrate (LF_N-LPSTGG-His₆ or LF_N-DTA-LPSTGG-His₆)

For each conjugate, the assembly was performed in presence of the C-terminal LPSTGG tag on the protein (LF_N or LF_N-DTA (50 μM) and the N-terminal G5 motif on the X peptide (1 mM). The reaction was catalyzed by 5 μM SrtA* in SrtA buffer (50 mM Tris pH 7.5, 150 mM NaCl and 10 mM CaCl) and incubated at room temperature for 15 min rocking. The mixture were rotated for an additional 15 min after the addition of Ni-NTA agarose beads, which bound the enzyme and the unreacted starting materials from the reaction mixtures. The samples were then spin filtered at 4 °C and buffer exchanged three times into 20 mM Tris pH 7.5 and 150 mM NaCl to remove the excess peptide. LC–MS was used to analyze the purity of each construct.

Sv1: Purity >95%, tR 9.3 min (MALDI-TOF MS, 1–61% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 4.0 mL/min); calculated mass, 32965.98; found, 32965.17±0.4

Sv2: Purity >95%, tR 9.2 min (MALDI-TOF MS, 1–61% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 4.0 mL/min); calculated mass, 32965.98; found, 32965.02±0.4

Sv3: Purity >95%, tR 9.3 min (MALDI-TOF MS, 1–61% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 4.0 mL/min); calculated mass, 32965.98; found, 32965.15±0.4

Sv4: Purity >95%, tR 9.5 min (MALDI-TOF MS, 1–61% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 4.0 mL/min); calculated mass, 32965.98; found, 32965.2±0.4

Sv5: Purity >95%, tR 9.4 min (MALDI-TOF MS, 1–61% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 4.0 mL/min); calculated mass, 32965.98; found, 32965.4±0.4

Sv6: Purity >95%, tR 9.3 min (MALDI-TOF MS, 1–61% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 4.0 mL/min); calculated mass, 32965.98; found, 32965.19±0.4

Sv7: Purity >95%, tR 9.2 min (MALDI-TOF MS, 1–61% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 4.0 mL/min); calculated mass, 32965.98; found, 32965.15±0.4

Sv8: Purity >95%, tR 9.5 min (MALDI-TOF MS, 1–61% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 4.0 mL/min); calculated mass, 32965.98; found, 32965.5±0.4

Sv9: Purity >95%, tR 9.35 min (MALDI-TOF MS, 1–61% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 4.0 mL/min); calculated mass, 32965.98; found, 32965.17±0.4

Sv10: Purity >95%, tR 9.3 min (MALDI-TOF MS, 1–61% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 4.0 mL/min); calculated mass, 32965.98; found, 32965.18±0.4

Sv11: Purity >95%, tR 9.3 min (MALDI-TOF MS, 1–61% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 4.0 mL/min); calculated mass, 32965.98; found, 32965.09±0.4

Sv12: Purity >95%, tR 9.4 min (MALDI-TOF MS, 1–61% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 4.0 mL/min); calculated mass, 32965.98; found, 32965.11±0.4

Sv13: Purity >95%, tR 9.5 min (MALDI-TOF MS, 1–61% ACN in H₂O (0.1% TFA) over 15 min, flow rate of 4.0 mL/min); calculated mass, 32965.98; found, 32965.08±0.4

2.5.2 Biological Studies

Construction of plasmids for recombinant proteins and transfection.

The gene for LF_N and LF_N-DTA kindly provided by MIT, Department of Biological Engineering. For separation purposes, a 6xHis tag was added to both the terminus of the proteins. Site-directed mutagenesis was carried out using the QuikChange II site-directed mutagenesis kits (Agilent)

according to manufacturer's directions. The LF_N-DTA_{mut} E148S was PCR-based site directed mutated using the forward primer 5'- ctgagggctcttcttctgtttcttacatcaacaactgggaacagge -3' and the reverse primer 5'- gcctgttcccagttgtgatgtaagaacagaagaagaccctcag-3'. Sample reaction was prepared using 5 µl of 10× reaction buffer 50 ng of dsDNA template, 125 ng of oligonucleotide primer #1, 125 ng of oligonucleotide primer #2, 1 µl of dNTP mix and ddH₂O to a final volume of 50 µl. Then 1 µl of *PfuUltra* HF DNA polymerase (2.5 U/µl) was added. The following thermocycling conditions were used: 95 °C for 30 s; 18 cycles of (95 °C, 30 s; 55 °C, 1 min; 68 °C, 1 min/kb of plasmid length). Following temperature cycling, the reaction mixture was placed on ice for 2 minutes to cool the reaction to ≤37°C. After annealing and digestion with 1 µl of the *Dpn* I restriction enzyme (10 U/µl), 2 µl of the double-stranded DNA ligation product was transformed into XL1-Blue Supercompetent cells and plated on 30 µg/mL kanamycin plates and incubated overnight at 37 °C. Colonies were grown in LB media containing 30 µg/mL kanamycin. The plasmid DNA was isolated using the Qiaprep spin miniprep kit (Qiagen, Netherlands).

Protein expression and purification. His₆-SUMO-LF_N-LPSTGG-H₆, His₆-SUMO-LF_N-DTA-LPSTGG-H₆, His₆-SUMO-LF_N-DTA_{mut}-LPSTGG-H₆, SortA*-His₆ and WT anthrax protective antigen (PA) were expressed in *E. coli* BL21 (DE3) cells. The cells were grown at 37 °C to an ~OD₆₀₀ 0.6-1.0. The proteins were induced with 0.4 mM IPTG overnight at 30°C. After induction, the cells were pelleted. Approximately 40 g of cell pellet was resuspended in 100 ml of 20 mM Tris pH 7.5 and 150 mM NaCl with protease inhibitor cocktail (Roche), DNaseI (Roche), and lysozyme. The cells were sonicated three times for 20 seconds on ice then spun down for 30 minutes at 35,000 g and 4° C. After sonication, the suspension was centrifuged at 17,000 rpm for 40 minutes. The lysate was purified over three 5 ml HisTrap FF Ni-NTA columns (GE Healthcare, UK) pre-equilibrated with 100 mL of 20 mM Tris-HCl pH 8.5, 150 mM NaCl, at pH 8.5 Each protein was loaded onto the column then washed with the equilibration buffer and then with 100 mL of 40 mM imidazole in 20 mM Tris-HCl pH 8.5, 500 mM NaCl. The protein was eluted from the column using 500 mM imidazole in 20 mM Tris-HCl pH 8.5, 500 mM NaCl. The eluted protein was buffer exchanged into 20 mM Tris- HCl pH 8.5, 150 mM NaCl using a HiPrep 26/10 Desalting column (GE Healthcare, UK). After desalting the proteins, the native N-termini were obtained by cleaving the SUMO protein fusion. For SUMO leavage, 1 µg SUMO protease was added per mg protein for 1 hour at room temperature. WT PA was overexpressed in the periplasm of *E. coli* BL21 (DE3) cells and purified by anion exchange chromatography.

Cell line maintenance. Chinese hamster ovary (CHO-K1) cells were grown in F-12K medium supplemented with 10% (v/v) fetal bovine serum and 1× penicillin–streptomycin in an incubator set at 37 °C and 5% CO₂.

Protein Synthesis Inhibition Assay with LF_N-DTA-X Constructs. CHO- K1 cells were plated at densities of 2 to 2.5 × 10⁴ cells per well in 96-well plates and incubated one day prior to the assay. The following day, cells were treated with medium containing 10-fold serial dilutions of the LF-DTA-X construct (starting from 100 nM); then PA was added to each well for a final concentration of 20 nM. The plates were incubated for 6 h at 37 °C and 5% CO₂. Therefore, the cells were washed twice with PBS. Leucine-free F- 12K containing H-leucine (1 μCi mL⁻¹, PerkinElmer) was added and incubated for an additional 1 h at 37 °C and 5% CO₂. After the incubation, the cells were washed again twice with PBS and suspended in 100μl of scintillation fluid. The amount of H-Leu present was calculated by scintillation counting. For each sample, the scintillation counts were normalized against cells treated with PA only. The data were fitted with a sigmoidal Boltzmann fit using OriginLab software. Each data point corresponds to the average of measurements performed in triplicate.

Translocation of LF_N-DTA-X and LF_N-X Constructs. 200,000 CHO- K1 cells were seeded per well on 12-well plates 16 h prior to the treatment. Cells were treated with 100 nM LF-DTA-X construct in the presence of 20 nM PA in serum-containing F- 12K for 6 h at 37 °C and 5% CO₂. In select experiments, with the aim of inhibiting the proteasome activity, cells were preincubated with 20 μM lactacystin for 30 minutes at 37 °C and 5% CO₂ and then subsequently treated with the LF-DTA-X constructs in the presence of PA, as described above. After translocation, the medium was removed and cells were detached enzymatically with 1× 0.25% trypsin-EDTA solution for 5 min at 37 °C and 5% CO₂. The cells were washed twice with PBS followed by centrifugation at 1200 rpm for 2 min at room temperature. In order to obtain the cytosolic fraction, cells were lysed with buffer consist of 50 μg mL digitonin in 75 mM NaCl, 1 mM Na₂HPO₄, 8 mM Na₂HPO₄, 250 mM sucrose, and protease inhibitor cocktail (Roche) for 10 min on ice and then spun down at 4 °C for 10 min.

Western Blot Analysis with LF_N-DTA-X and LF_N-X Constructs. The extracted lysates were analyzed by Western blot. Samples were prepared in Laemmli sample buffer and were electrophoresed in outer chamber filled with MES running buffer (50 mM MES, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.3) at 80 V for 10 mins and then at 165 until the dye front reached the end of the gel, using polyacrylamide gels (4-12%) of 1.0 thickness. Nitrocellulose membrane

0.2µm and filters were soaked in 60% 5 × Trans-Blot®Turbo™ Transfer Buffer, 20% Millipore water and 20% ethanol (reagent grade ~85% or molecular biology grade ~95-98% purity). Protein gels were blotted at 25 V for 30 mins using a Trans-Blot®Turbo™ Transfer apparatus. The membrane was blocked with Odyssey blocking buffer - PBS for 1h at RT and with a gentle agitation and then probed with the appropriate primary antibody or with Streptavidin (each one diluted 1:2000) in Tris-Buffered Saline with Tween-20 (TBST - 50 mM Tris-HCl, 150 mM NaCl, 0.05% Tween 20 (w/v)) + Odyssey blocking buffer – PBS (LI-COR) for 2h at RT. The blots were washed three times for 5 min with TBST, then immunostained with the appropriate secondary antibody (1:5000 dilution) in TBST, and imaged using an Odyssey infrared imaging system (LI-COR). The efficiency of lysis was analyzed by anti-Erk1/2 (cytosolic protein) and anti-Rab5 (early endosome) immunostaining. Data were obtained from experiment in triplicates and normalized by Erk 1/2 bands as a trustworthy and quantifiable loading control. The amount of lysate loaded was within the linear dynamic range to ensure precise, quantifiable, and normalized densitometric data.

