Chemically synthesized peptide libraries as a new source of BBB-shuttles. Use of Mass Spectrometry for peptide identification and quantification

Bernat Guixer

2014

Bernat Guixer

Chemically synthesized peptide libraries as a new source of BBB-shuttles. Use of Mass Spectrometry for peptide identification and quantification

Bernat Guixer

UNIVERSITAT DE BARCELONA

Programa de Química Orgànica

Tesi Doctoral

Chemically synthesized peptide libraries as a new source of BBB-shuttles. Use of Mass Spectrometry for peptide identification and quantification

Bernat Guixer

Dirigida i revisada per:

Dr. Ernest Giralt
(Universitat de Barcelona)

Dra. Meritxell Teixidó
(Institut de Recerca Biomèdica Barcelona)

Barcelona, 2014

Contents

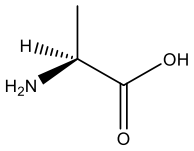
Abbreviations and annexes

Abbreviations

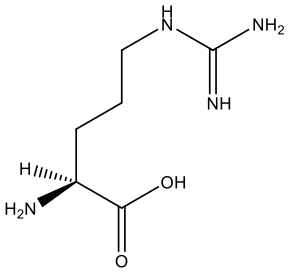
Aa	Amino acid
ACH	α-cyano-4-hydroxycinnamic acid
ADME	Absortion, distribution, metabolism and excretion
Amp	Ampicilin
amu	Atomic mass unit
ATP	Adenosine triphosphate
BBB	Blood-brain barrier
BBEC	Bovine brain endothelial cell
*t*Bu	*tert*-Butyl
Boc	*tert*-Butyloxycarbonyl
BSA	Bovine serum albumin
cAMP	Cyclic adenosine monophosphate
CD-31	Cluster of differentiation 31
CEF	Chicken embryo fibroblasts
COMU	1-Cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate
CNS	Central nervous system
CPPs	Cell penetrating peptides
DAPI	4',6-diamidino-2-phenylindole
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCM	Dichloromethane
DIEA	N,N-Diisopropylethylamine
DIPCDI	N,N'-Diisopropylcarbodiimide
DMEM	Dulbecco's modified eagle medium
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DMT	Dimethoxytrityl
DNA	Deoxyribonucleic acid
EB	Elution buffer
E. Coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
eq	Equivalent
ESI	Electrospray ionization
FA	Formic acid
Fmoc	9-Fluorenylmethoxycarbonyl
FT-ICR	Fourier transform ion cyclotron resonance
HPLC-MS	High-performance liquid chromatography coupled to a MS detector
HPLC-UV	High-performance liquid chromatography coupled to a UV detector
hTfR	Human transferrin receptor
HTS	High-throughput screening
IT	Ion trap
JAM's	Junction associated molecules
Kan	Kanamicine
KRB	Krebs-Ringer bicarbonate buffer
LB	Lysogeny broth
LC	Liquid chromatography
LDL	Low-density lipoprotein
LTQ	Linear trap quadrupole
LY	Lucifer Yellow
MALDI	Matrix assisted laser desorption/ionization
MDCK	Madin Darby canine kidney
MeCN	Acetonitrile

MeOH	Methanol
MS	Mass spectrometry
MS1	Precursor ion mass spectra
MS/MS	Tandem mass spectrometry
m/z	Mass-charge ratio
NMR	Nuclear magnetic resonance
NVU	Neuro-vascular unit
OBOC	One-bead one-compound
PAMPA	Parallel artificial membrane permeability assay
P_{app}	Apparent permeability
Pbf	2,2,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl
PBS	Phosphate buffer saline
PBST	Phosphate buffer saline, 0.3% Triton X-100
PCR	Polymerase chain reaction
P_e	Effective permeability
PEG	Polyethylene glycol
P-gp	P-glycoprotein
PIC	Protein inhibitor cocktail
ppm	Parts per milion
q-PCR	Quantitative polymerase chain reaction
QqQ	Triple quadrupole
Q-trap	Quadrupole trap
RBE4	Rat brain endothelial cell line
RO-24	4-(3-Butoxy-4-methoxybenzyl)-2-imidazolidinone
RP	Reverse phase
SDS	Sodium dodecyl sulfate
SIC	Selected ion chromatogram
SID	Stable-isotope dilution
SOC	Super optimal broth with catabolite repression
SPPS	Solid-phase peptide synthesis
SRM	Selected reaction monitoring
t	Time
TB	Terrific broth
TBMB	1,3,5-tris(bromomethylene)benzene
TEER	Transendothelial electrical resistance
Tet	Tetracicline
TFA	Trifluoroacetic acid
TOF	Time of flight
TfR	Transferrin receptor
TIC	Total ion chromatogram
TIS	Triisopropylsilane
Trt	Trityl
TU	Transducing unit
RT	Retention time
UWL	Unstirred water layer
ZO	Zonula occludens
%R	Percentage of membrane retention
%T	Percentage of transport

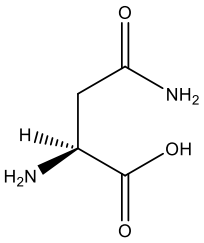
Annex I: Proteinogenic L-amino acids used



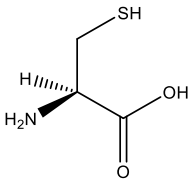
L-Alanine
Ala
A



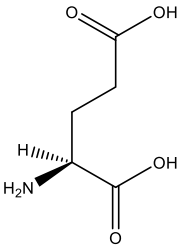
L-Arginine
Arg
R



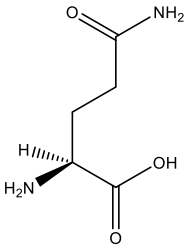
L-Asparagine
Asn
N



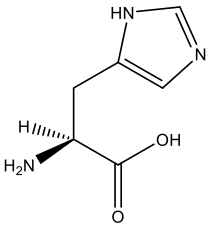
L-Cysteine
Cys
C



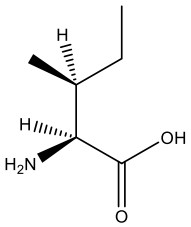
L-Glutamic acid
Glu
E



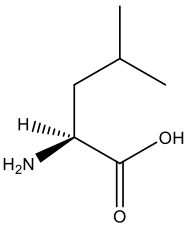
L-Glutamine
Gln
Q



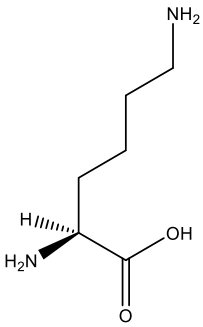
L-Histidine
His
H



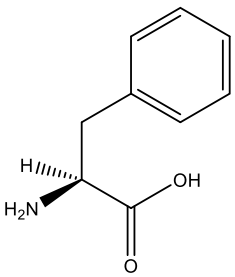
L-Isoleucine
Ile
I



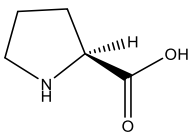
L-Leucine
Leu
L



L-Lysine
Lys
K

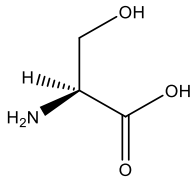


L-Phenylalanine
Phe
F

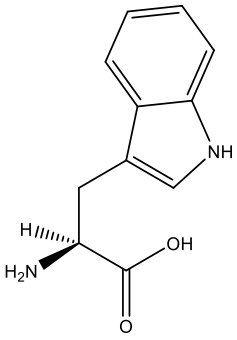


L-Proline
Pro
P

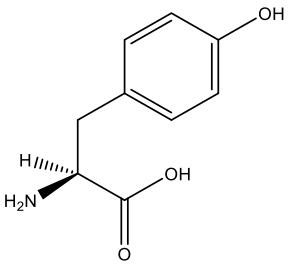
* Amino acids abbreviations follow the rules of the Comission on Biochemical Nomenclature of the IUPAC-IUB as specified in *Eur. J. Biochem.* 1984; 138: 9-37 and *Eur. J. Biochem.* 1993; 213: 2.



L-Serine
Ser
S

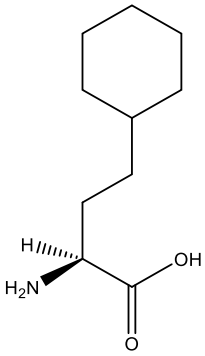


L-Tryptophan
Trp
W

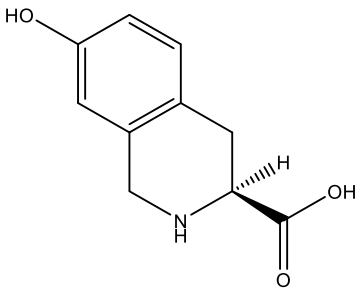


L-Tyrosine
Tyr
Y

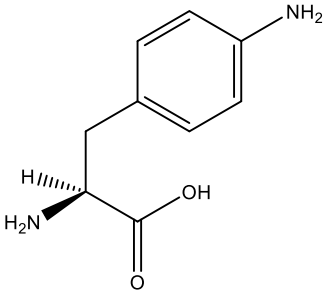
Annex II: Non-proteinogenic L-amino acids used



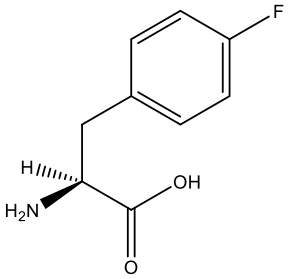
homocyclohexyl-L-alanine



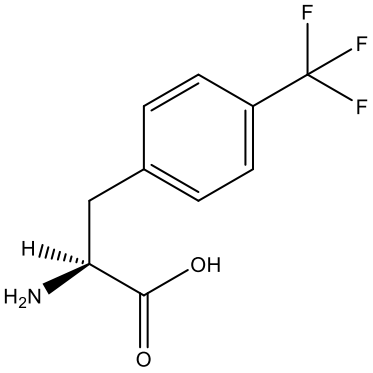
7-hydroxy-(S)-1,2,3,4-tetraydroisoquinoline-3-carboxylic acid



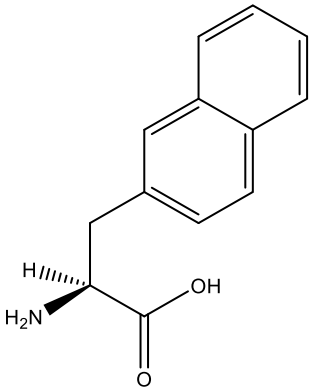
4-amino-L-phenylalanine



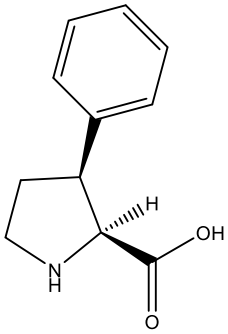
4-fluoro-L-phenylalanine



4-trifluoromethyl-L-phenylalanine

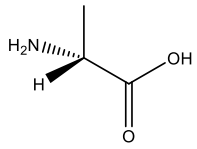


L-2-naphthylalanine

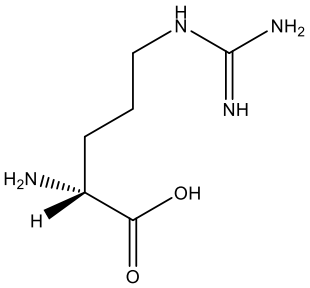


(2S, 3S)-3-phenylpirrolidine-2-carboxilic acid

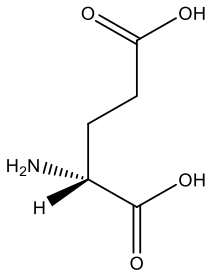
Annex III: D-amino acids used



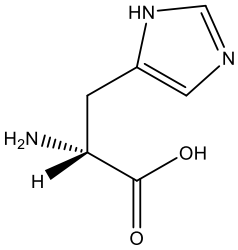
D-Alanine
a



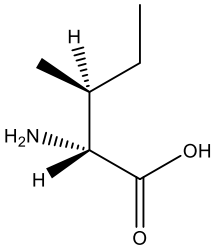
D-Arginine
r



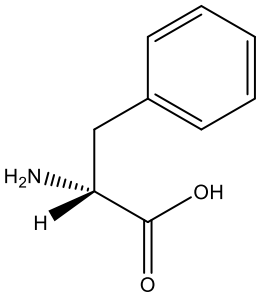
D-Glutamic acid
e



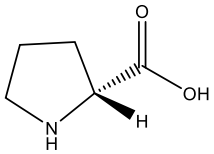
D-Histidine
h



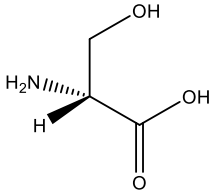
D-Isoleucine
i



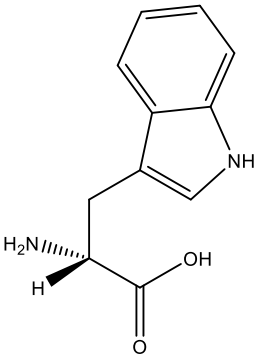
D-Phenylalanine
f



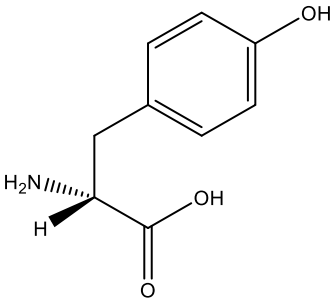
D-Proline
p



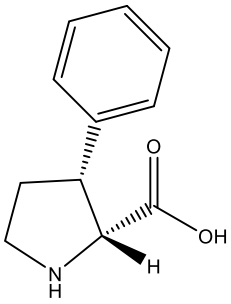
D-Serine
s



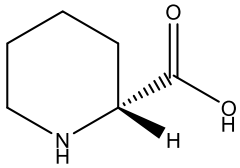
D-Tryptophan
w



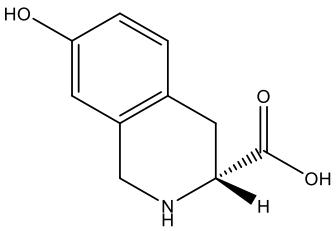
D-Tyrosine
y



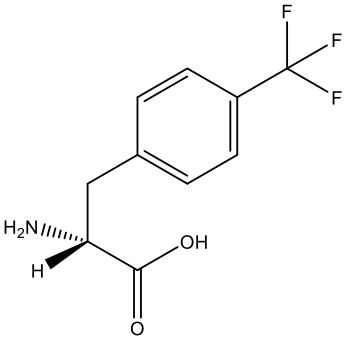
(2R, 3R)-3-phenylpirrolidine-2-carboxilic acid



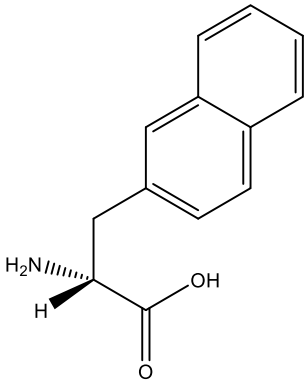
D-pipecolic acid



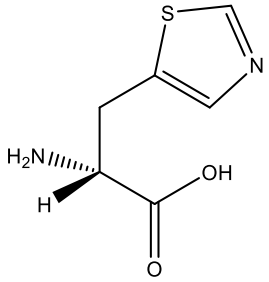
7-hydroxy-(R)-1.2.3.4-tetrahydroisoquinoline-3-carboxilic acid



4-trifluoromethyl-D-phenylalanine

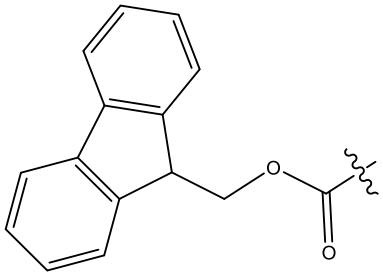


D-2-naphthyalanine

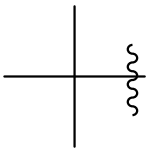


4-thiazoylalanine

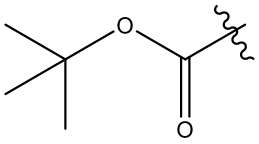
Annex IV: Protecting groups used



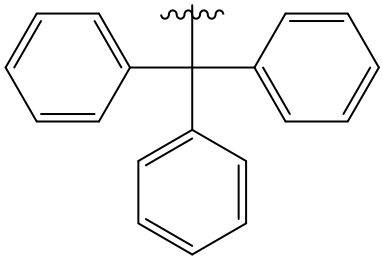
Fmoc



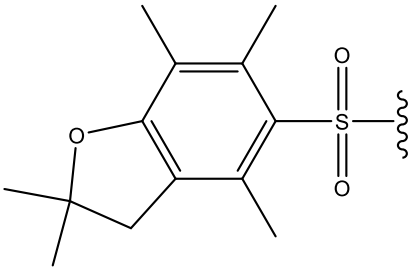
*t*Bu



Boc

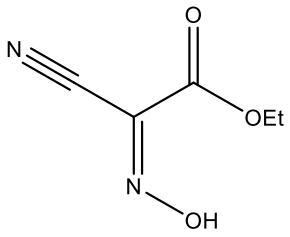


Trt

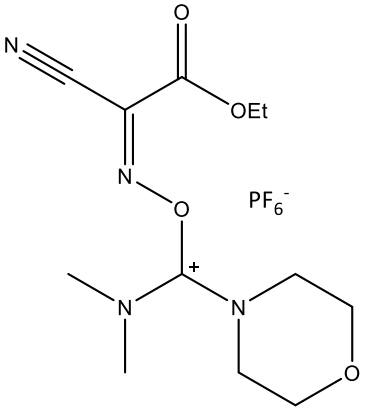


Pbf

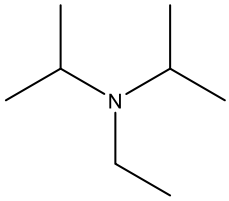
Annex V: Coupling reagents used



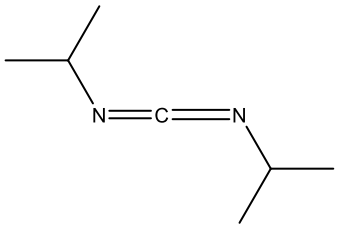
Oxima®



COMU

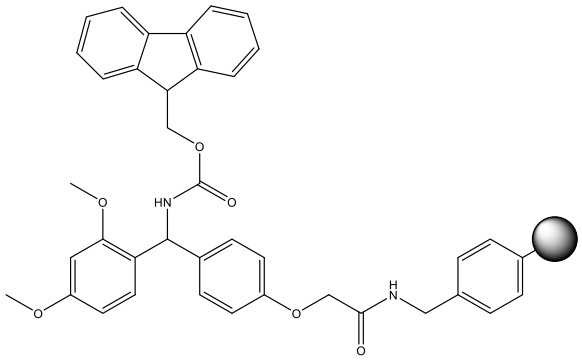


DIEA

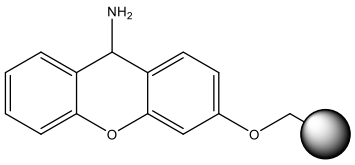


DIPCDI

Annex VI: Resins used



Rink amide AM resin



Sieber Resin

INTRODUCTION AND OBJECTIVES

Introduction

1. The blood-brain barrier

1.1. Social and clinical scope of central nervous system diseases

The spectrum of disorders affecting the brain is broad and has a high-prevalence in our modern society where the life expectancy has increased significantly. Those disorders are usually accompanied by short- and/or long-term impairments and disabilities, which result in emotional, social and financial burden for patients and their relatives.

Brain diseases are more common than is generally perceived. Studies such as *Cost of Disorders of the Brain in Europe 2010* by Jes Oleson[1] commissioned by the European Brain Council (EBC) alert politicians and policy makers of the urgent need for development of European Union policies for prevention of brain disorders and increase founding for promoting brain research in Europe to improve treatment and quality of life of patients suffering brain diseases.

Total European 2010 cost of brain disorders was €798 billion in 2010: 24% of direct health care costs in Europe; which is similar to the combined costs of cardiovascular disease, cancer and diabetes.[2] Direct expenditure such as health care cost and non-medical cost correspond to 37% and 23% of this amount, respectively; while 40% is associated to indirect costs such as loss of productivity when unable to work. On average, the European cost of disorders of the brain was 1550€ per capita on 2010.

In Catalonia, social initiatives supported by the Catalan government and its national TV such as TV program *La Marató de TV3* in its 1996, 2000, 2005, 2008, 2010 and 2013 editions,[3] play a fundamental role in normalizing and spreading the awareness and knowledge of brain diseases as well as helping in research programs founding. Diffusion of brain disorders also helps in breaking the stigma, that sometimes accompanies these mental and neurological disorders, which affect individuals in a way that many patients describe as worse as the illness itself.[4]

Despite enormous effort on developing drugs to tackle CNS diseases, the presence of the blood-brain barrier (BBB) severely restricts drug delivery to the sites where drugs must play specific effects. Actually, 98% of small molecules and 100% of large-molecule pharmaceuticals do not cross the BBB.[5] Furthermore, BBB dysfunction is implicated in some neurodegenerative disorders. Understanding the BBB features and nature is fundamental for the design of effective drugs and therapies for the treatment of CNS diseases.

Altogether, it is clear there is an urgent need for new drugs able to bypass the BBB and delivery to specific brain targets to tackle brain disorders.

1.2. Structure and function of the blood-brain barrier

The blood-brain barrier (BBB) is considered the most important barrier layer among CNS barriers.[6] Located in the brain capillaries, the BBB is formed by endothelial cells, which are importantly influenced by the neuro-vascular unit (NVU). NVU is composed of several cell types, which through complex interactions between them maintain the optimal function of the BBB and the brain. Initially described at the level of arterioles, NVU has now been applied to capillary level and it is described to strongly modulate the features and functions of brain capillary endothelial cells. Pericytes and astrocytes (embedding the microvessel wall by its endfeet) are crucial in barrier properties induction (physiology and integrity). Furthermore, astrocytes provide links to neurons. Microglia are CNS-resident immune cells and are also considered within NVU.

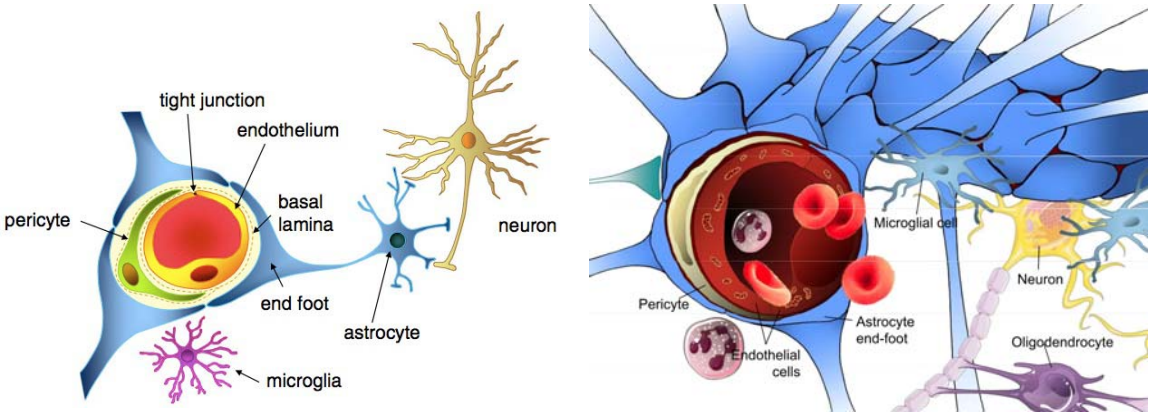


Figure 1. Schematic images of the BBB showing presence and interactions of endothelial cells with the neuro-vascular unit, which confer unique BBB features. From N. J. Abbott.[7]

Other barriers than BBB, such as blood-cerebrospinal fluid barrier, located in choroidal plexus epithelium, or ependymal, a non-vascularized arachnoid epithelium, constitute the whole set of barriers which seals the CNS. The BBB focus most attention among CNS barriers because it is considered the most important influx barrier. The BBB is located along 600 km of brain capillaries and has an approximate surface area of 20 m^2.

The BBB is crucial for molecular flux between blood and brain parenchyma. It protects the brain (and the CNS) from eventually harmful substances potentially being neurotoxic circulating in the bloodstream by strictly controlling the molecular passage through it. The ability to keep out toxins is essential to ensure neuronal longevity, preserve neural network connectivity and avoid innate immune response. Actually, there is evidence that the BBB evolved in such a way to favour precise neuronal signalling, thus providing and maintaining a homeostatic environment in the brain parenchyma.[8]

The barrier functions are dynamic and respond to a large spectrum of regulatory signals. The BBB is described to be a multifunctional barrier with physical, physiological, metabolic and immunological regulatory functions.[7]

The restrictive physical effect of the BBB is achieved by restricting the paracellular transport. Tight junctions are responsible to dramatically restrict diffusion of ions and other polar solutes as well as completely block passage of macromolecules by this route. Since ion flux is restricted, the electrical resistance of this barrier *in vivo* has been described as 1,800 $\Omega \cdot cm^2$. Tight junction formation involves interactions of molecular complexes such as *zonula occludens* (ZO), occludin and claudin proteins, as well as junction associated molecules (JAM's). Also adherent junctions driven by cadherin proteins play a crucial role.[9, 10] Another feature that limits the passage from blood to brain is its low vesicular transport and high metabolic activity.[11] Expression of a variety of extracellular and intracellular enzymes contributes on a higher level of protection by metabolizing a broad spectrum of molecules during transport events across this biological barrier.[12-14]

This multifunctional behaviour makes the BBB a very stringent hurdle for a vast number of molecules.

1.3. Transport through the blood-brain barrier

Despite the strict insulation of the CNS by the BBB, which limits the passage of unspecific molecules into the brain, there are a variety of mechanisms by which essential molecules and nutrients are supplied to the brain to ensure the correct function of this organ and also maintain the properties and functions of the BBB itself. These mechanisms can be divided into passive and active transport systems, which can be regarded as potential routes to deliver drugs into the brain.

Since paracellular flux is essentially hindered by tight junctions, passive diffusion is the main representative of passive transport. Passive diffusion strictly relies on the physicochemical properties of compounds, basically on molecular weight and lipophilicity (which is influenced by the capacity of molecules to make hydrogen bonds). Passive diffusion is a non-saturable and non-energy dependent mechanism. Gas molecules such as O_2 or CO_2 and low molecular weight (<400-600 amu) lipophilic molecules are specially suited to cross through this mechanism.

On the other hand, active transport is an energy-requiring and saturable mechanism that can be divided into: carrier-mediated transport, absorptive-mediated transcytosis, and receptor-mediated transcytosis.

Carrier-mediated transport is driven by highly specialized proteins that supply to the brain nutrients such as glucose, amino acids and purine bases. Occupation of the carrier determines the transport rate. Receptor-mediated transcytosis and absorptive-mediated transcytosis are both vesicle-based transporters. The former is described to be the mechanism of larger molecules to enter the brain such as transferrin, insulin, leptin and low-density lipoprotein, among others, through interaction with a membrane-receptor. Contrarily, the later is based on endocytosis upon binding of highly charged cationic compounds to the plasma membrane. However, this route is specially down-regulated to protect the brain from nonspecific exposure to polycationic compounds. The mechanisms by which those molecules cross via vesicular-mediated passage are still quite poorly understood.

Besides mentioned influx mechanisms, several efflux mechanisms are present at the BBB. P-glycoprotein (P-gp) is a transmembrane protein, which has high affinity for a variety of cationic and lipophilic compounds and subsequently restricts the net transport to the brain of many drugs. A variety of efflux pumps had been described. It is noteworthy that their presence significantly varies among species.[13, 15, 16]

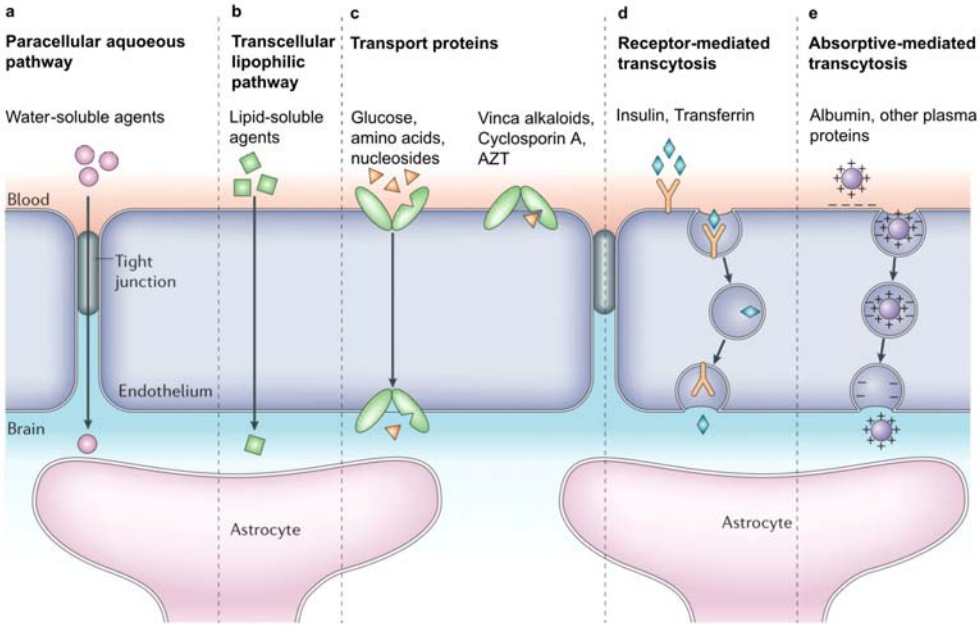


Figure 2. Molecular pathways across the BBB. Paracellular diffusion is highly restricted because of the presence of tight junctions between endothelial cells, whereas transcellular diffusion or passive diffusion is the main passive transport pathway, however is usually limited to small and lipophilic molecules and is effective if they are not substrates of the efflux pumps located on the BBB. Active transport are divided in carrier-mediated transport, which specifically mediates supply of nutrients; receptor-mediated transcytosis, a vesicle-based receptor mediated pathway; and absorptive-mediated transcytosis which is very limited in the BBB to avoid entrance of unspecific polycationic molecules. Efflux pumps are used as detoxifying tools for the brain and are described to hamper the net influx of many molecules. Modified from Abbot *et al*.[17]

1.4. Drug delivery to the brain

Despite enormous effort on addressing CNS diseases by pharmaceutical companies, the number and effectiveness of drugs released to the market are very limited, while the amount of resources spend are huge.

This is mainly due to the ineffective delivery of drugs targeting the CNS. Many approaches had been proposed to bypass the BBB: from invasive strategies such as intracerebral injection[18], convection-enhanced diffusion[19] and pseudo-invasive strategies as BBB temporal disruption[20] to non-invasive drug delivery. Invasive and pseudo-invasive strategies are very aggressive and can lead to severe secondary effects and sequels to patients. Thus, those strategies are discouraged, especially for chronic treatments, unless there is no other option.

On the other hand, since the brain it is highly vascularized, intravenous injection appears as an ideal route of administration. However, the presence of the BBB highly hampers the passage of molecules**.** To further exploit this approach, strategies such as modify the structure of the drug to increase its permeability had been postulated. On one hand, slightly modifying physicochemical properties to enhance drug passive diffusion. On the other hand, regarding to carried-mediated transport, only drugs closely mimicking endogenous carrier substrates will be able to be transported through the BBB, to do so, the structure of the drug aimed to be transported across the BBB is structurally modified with the objective to resemble the structure of any substrate molecule known to interact with any carrier-mediated transporter. However, these modifications may worsen drug effectiveness by ruining its therapeutic activity. In most cases it is hard to find a convenient balance between drug activity and BBB penetration. Another alternative is called BBB-shuttle strategy or Trojan horse strategy.[21, 22] This approach is based on linking a given drug unable to cross the BBB to a BBB-shuttle, which is a molecule able to cross the BBB. Ideally this construct BBB-shuttle-drug should cross together the BBB and facilitate the uptake of the drug

into the brain parenchyma. Linkage of both moieties can be performed either covalently[23, 24] or by non-covalent interactions,[25, 26] but must not interfere in the functionality of any of the two entities. This fact strongly governs the applicability of this approach. Ideally, a universal BBB-shuttles should be able to assist the passage of any molecule attached to it. Since passive diffusion relies on physicochemical properties, linkage of a drug (henceforth cargo) could dramatically affect the shuttle properties. Although successful and promising passive shuttles had been described,[27-29] present and future efforts focused on BBB-shuttles using any active transport mechanisms are mostly preferred to launch this shuttle strategy since, in theory, they should not be limited by cargo size. Targeting highly expressed BBB receptors such as transferrin, lactoferrin, insulin, leptin receptors, or the LDL-receptor related protein 1 and 2 with the aim to achieve drug transport across the BBB by means of receptor-mediated transcytosis might be the more effective strategy. Constructs developed include peptides, fusion proteins, dendrimers, liposomes and nanoparticles among others.[7]

The list below shows a compendium of companies and their systems for transport cargos across the BBB based on BBB-shuttles (data from Medtrack database).

- *ArmaGen technologies Inc* – the shuttles are Fc-IgG-based fusion proteins; this company closed a 17 million USD round in 2012, receiving investments from large pharma companies like Boehringer Ingelheim, Shire and Takeda.
- *CLL Pharma* - based on peptide vectors derived from natural mammalian antimicrobial peptides involved in an ancestral immune response system.
- *biOasis Technologies Inc* - based on p97, also known as melanotransferrin. This technology accepts only cargoes that bind to p97 or that can be conjugated to an anti-p97 antibody.
- *AngioChem* – based on peptide vectors that interact with LRP-1 receptor; recently announced a 300 Milion USD collaboration with GSK to tackle Lysosomal storage diseases.
- *Biogen Idec* - the shuttles are Fc-IgG-based fusion proteins.
- *Corden Pharma Switzerland LLC* – technology is a special formulation that increases drug penetration, possibly through passive transport.
- *Raptor Pharmaceuticals Corp* – shuttles are peptide-based and interact with one specific receptor (active transport).
- *Roche's Genentech Inc* – developed a bispecific antibody. One antigen-binding site specific for TfR and the other for β-site of amyloid precursor protein (APP)-cleaving enzyme.

1.5. Modelling the blood-brain barrier

Therapeutic potential of drugs targeting the CNS is not only related on their activity but also on their ability to reach the target in sufficient amounts, which requires bypassing the BBB. There is a pressing need to address BBB drug permeability early on the drug discovery process.

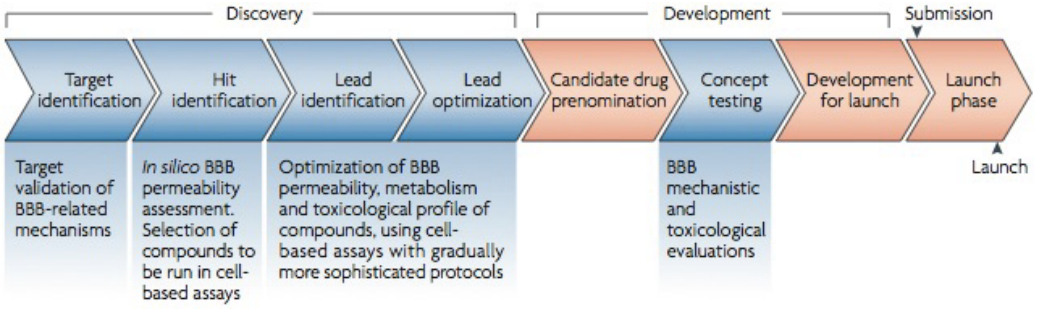


Figure 3. Phases on early stages of drug development where BBB models are especially useful. From Cecchelli *et al.*[30]

BBB modelling enables to study the effect of studied therapeutics at the level of the cerebral capillaries. This information is valuable to discard molecules not fulfilling BBB permeability requirements and also allowing rational drug design.

Nonetheless, modelling the BBB is not an easy task, which is indeed not still completely accomplished. Several approaches had been developed to assess drug BBB permeability.

1.5.1. *In vivo* BBB models

Usually, *in vivo* models are regarded as the best option since BBB is in its physiological environment. *In vivo* data is mostly obtained by *in situ* brain perfusion and brain microdialysis technique, which determines the concentration of unbound drug in the brain interstitial fluid. However, these techniques are not suitable for massive screening of compounds.[31]

However, although *in vivo* data is always more relevant in clinical research, it is extremely difficult to get data from humans. Significant differences had been reported in the brain uptake ratios between human and rodents (the first choice and most available *in vivo* model), and even in higher primates (which are considered very close to humans). This BBB specie-dependent performance is related to specific and variable presence and/or functionality of drug transporters (both influx and efflux) and enzymes at the BBB.[31]

Therefore, developing *in vitro* models closely reflecting (if possible human) *in vivo* situations is of most interest for reliable data and effective drug development.

1.5.2. *In vitro* cell-based BBB models

It is obvious that *in vitro* models cannot fully model all BBB features, but cell-based BBB assays are powerful and flexible experimental tools to study transport and dynamic functions of the BBB, that could not be achieved in any other way. In addition, it allows the massive screening of compounds. *In vitro* usefulness relies on mimicking *in vivo* situation for certain phenotypes.

First *in vitro* model was based on isolated brain capillaries and bring valuable data on cerebral endothelial receptors, transporters, and signalling mechanisms. [32, 33] This technique showed significant limitations since mainly assesses transport from the abluminal to the luminal side and suffers from low capacity.

Primary or low passage brain capillary endothelial cells had been established as the most useful and promising way to model the *in vitro* BBB. Models based on a wide variety of species had been reported. However, bovine and porcine models had attracted more attention as a source of brain endothelial cells because of brain size and attainability. However, substantial interest has been directed to rodent BBB assays because allows *in vitro*/*in vivo* correlations at a reasonable cost compared to other species.[34-36]

Obtaining primary cultures is not an easy task and many factors (some of them difficult to take under control) may affect the brain endothelial cell quality at the end of the process. Advances in cellular biology allow harvesting high quality endothelial cells that can be subcultured and stored in liquid nitrogen.[37-39] However, pure endothelial cells isolation is still not possible due to contamination with small amounts of other NVU cells.[40] Primary or low-passage cell cultures interestingly retain many morphological and biochemical properties of the *in vivo* BBB such as endothelial antigens, polarized expression of enzymes, transporters, receptors and tight junctions. However, it has been demonstrated that glial cells dynamically interact with the BBB regulating key properties of the BBB *in vivo*, such as mediating the expression of transporters and/or receptors or inducing tighten tight junctions between endothelial cells.[38, 41]

Co-culture of brain endothelial cells with primary astrocytes or glial cells had been reported to increase the *in vitro* BBB resemblance to the *in vivo* scenario. For transport experiments, culture of brain endothelial cells monolayers is usually performed in transwells, which support the growing of the endothilal cells on their polymer based membrane. Co-culture of glial cells had been described either on the bottom of wells containing those transwells, thus avoiding direct contact of endothelial and glial cells, or with glial cells seeded up-side down on the lower part of the transwell polymeric membrane.[38, 39, 42-44] Another way to generate glial cell influence on the endothelium monolayer is the use of media that has been harvested from growing astrocyte cultures, which is rich in biochemicals secreted by cultured astrocytes.[45-47]

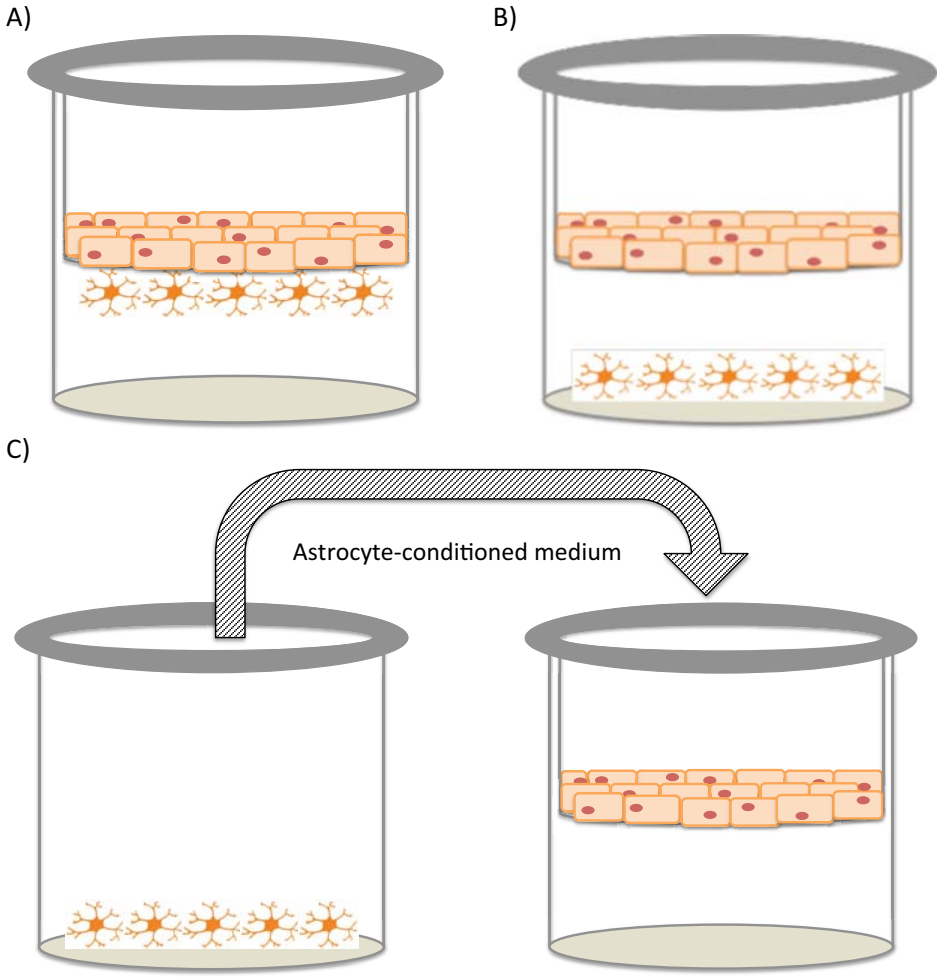


Figure 4. Schematic representation of described techniques for co-culture of glial cells and endothelial cells. While endothelial cells always laid over the transwell polymer membrane, astrocytes can be seeded up-side down on the lower part of the transwell polymeric membrane (A), on the bottom of the wells (B) or in a separate well (C), which produces astrocyte-conditioned medium for subsequent culturing of endothelial cells.

Adjuvant addition to enhance some *in vivo* BBB features is another approach that had been reported to improve *in vitro* BBB models. In this regard, cyclic adenosine monophosphate (cAMP) has been described as a modulating agent on paracellular permeability thus confirming the importance of phosphorylation and dephosphorylation processes to regulate tight junction strength or tightness.[48-50] cAMP analogue 8-(4-chlorophenylthio)-cAMP and 4-(3-butoxy-4-methoxybenzyl)-2-imidazolidinone (RO-24), a selective inhibitor of cAMP-phosphodiesterase had been successfully used for this purpose.[44] Also hydrocortisone-supplemented medium had been reported to enhance tightness of the barrier.[51] However, some authors argue those adjuvants lead to altered barriers, which may not mimic certain BBB features other than tightness, such as enzymatic or metabolic activity.

Fully assessment of *in vitro* BBB model quality is extremely challenging. Since cell-based models should feature several parameters of the BBB, a variety of controls should be performed to assess the mimicking degree of the model. Lacking precise understanding of interrelated factors affecting the BBB makes this even more difficult. Factors affecting *in vitro* BBB quality range from cell origin, quality after extraction, cell passage, adjuvant presence, among several others. Since those experiments and model establishment require significant degree of technical training on techniques used, this is another factor that should not be overlooked.

Optimization, characterization and validation of a model are crucial for trusting model performance. One of the most used parameters to assess BBB maturity is the so-called transendothelial electrical resistance (TEER) measurement. TEER relies on tight junction formation. The tighter the cells junctions, the higher the electrical resistance between apical and baso-lateral compartments on the transwell. This measurement can be used to monitor maturity of the cell layer to better choose the best membrane conditions to perform an assay. TEER measurement does not affect future performance of *in vitro* barriers because it is non-destructive. However, some authors argue this method disturb the cell development and enhances contamination chances and sometimes may lead to unreliable readout results. Other approach to control tight junctions formations is to assay permeability of compounds with known permeability across the BBB such as Lucifer Yellow (LY) or sucrose. However, other parameters than tight junctions should be monitored. Assaying the permeability of compounds described to cross de BBB by specific transporters/receptors would demonstrate the presence of those protein complexes on the cell membranes that trigger specific transcellular events. However, assaying compounds to assess *in vitro* BBB quality is not sustainable on each batch, since some wells should be assigned to this purpose. Initially, *in vitro* model establishment requires several controls (also over the cell passages) to finally have high confidence on performance and reproducibility of the model used, to fully trust data acquired on *in vitro* BBB model assays.

In some cases, positive and negative permeability controls can be assayed during compound evaluation in order to ensure the correct BBB performance of the model during the assay. This allows to known if there is a proper sealing between cultured cells and determine if the assayed compounds could eventually have a certain degree of cytotoxicity on the cell barrier (abnormal increase on permeability of the negative controls co-incubated with the compounds under study). However, it is important to highlight that those controls might somehow interact with the sample biasing results.

Immunochemistry analysis is a valuable tool to monitor the presence of certain proteins in the membrane. For example, proteins involved on tight jucntions formation such as ZO-1 or claudin, efflux pumps like the P-glycoprotein efflux pump (P-gp) or specific receptors such as the transferrin receptor or the low-density lipoprotein receptor.

The main problem on BBB *in vitro* model bibliography is comparing data among distinct models and studies. This fact is especially noteworthy when analyses are performed in distinct laboratories as reveal comparison of various studies.[52, 53]

Use of immortalized brain endothelial cell lines is been reported to display insufficient barrier properties to compete with the above-mentioned *in vitro* models, which mimic much closely the *in vivo* BBB. Neither rat RBE4 nor human immortalized cell lines (SV-HCEC or hCMEC/D3) lead to satisfactory results even adding cAMP or hydrocortisone.[54-56]

Non-cerebral cell lines had been also postulated as potential BBB model-building assays. Preferred cell lines to this purpose had been epithelial cell lines from the intestine (Caco-2), kidney (MDCK) and a clone of human bladder carcinoma (ECV304).[57] Caco-2, is primarily used to predict drug intestine absorbance, but since shows high TEER values and an elevated expression of efflux pumps it has been broadly used also as a tool to predict BBB permeability of compound on early stages of drug discovery. Similarly, MDCK cell line enables to achieve tight and

reproducible barrier properties. However, as also happens in Caco-2, cells show different morphology and gene expression than brain endothelial cells, thus lacking numerous characteristics of BBB.[58, 59]

Requirements for a good BBB model
Expression of tight junctions
Low permeation to sucrose/LY or high electric resistance
Selective and asymmetric permeability to physiologically relevant K^+ and Na^+ ions
Selective permeability to molecules
Expression of drug metabolizing enzymes
Functional expression of efflux pumps
Endothelial cells polarization
Ease culturing and low cost
Expression of tight junctions

Table 1. Most important requirements for a successful BBB model.

Human models

Until now, human models had not been regarded as a feasible approach for screening due to ethical reasons. Despite some *in vitro* human models had been established, low yield of cells, difficulty to have access to human brain tissue (autopsy, surgery or fetal tissue) and doubts on aged tissue coming from autopsy or surgery did not result as a widely used model.[60]

It has been reported the successful induction of pluripotent stem cells (derived from adult somatic cells) to generate neural stem cells, which can eventually differentiate into neurons and astrocytes.[61-63]

Similarly, stem cells differentiation to brain endothelial cells is knocking at the door and most likely models introducing this technology will deeply revolutionize the field in the coming years.

1.5.3. Further refinements on BBB models

Triple cultures with NVU cells had been described to widely enhance influence over brain endothelial cells. Usually a synergistic effect is expected when using this combination of cell lines. However, careful must be taken since adding a new cell line means increase exponentially the complexity of interactions, which might lead to non-beneficial effects.[60, 64]

Dynamic models had been also described to increase BBB conditions resemblance. All BBB models described above are considered static models. In those static models, cells forming the membrane are not exposed to the stress generated by the blood flow on the luminal side of endothelial cells. Dynamic models apply such stress by pumping cellular medium through three-dimensional tube structures where the cells are seed. Quasi-physiological growing of cells on that conditions led to morphology of endothelial cells resemble *in situ* phenotype and develop higher tightness (up to 1200 $\Omega \cdot cm^2$) thus closely mimicking the *in vivo* scenario. Dynamic models grown on self-polymerizing extracellular matrix protein scaffolds and on artificial capillary-like structures had been developed with promising results. However, those models still present some drawbacks and limitations such as they are not designed for high-throughput studies, require high technical skills, are time consuming and might require loading large amounts of cells for inoculation or mastering on self-polymerizing ECM matrixes depending on the nature of the dynamic approach. With the advances on microfluidics, this field is called to provide high-quality data for lead molecules studies since it is also compatible with human cells.[65]

1.5.4. Applications of BBB models

When closely mimicking the *in vivo* BBB those models should be able to determine the transport capacity of compounds to across the barrier. Cell-based BBB models take into account transcellular and paracellular diffusion pathways, active transport, metabolism and even non-specific interactions with cellular proteins or lipids. Remarkably, reliable models not only lead to valuable information on molecular transport abilities but also on transport mechanism if combined with *in vivo* models and on ADME (absorption, distribution, metabolism and excretion) characteristics. Use of transporter inhibitors, transporter substrates to create competition assays or performing assays at low temperature can lead to valuable information in terms of mechanistic understanding. Future developments on BBB models promise to provide an improved platform for BBB permeation studies.

Toxicological evaluation is also possible on BBB models as well as study of ADME characteristics of these toxic compounds and how they affect (and by which mechanism) BBB features. Although toxicological evaluation is of great interest on early stages of CNS drug discovery programs, actual model formats have not enough throughput for the analysis of large number of compounds.

Finally, controlled dysfunction of BBB models can mimic damaged BBB endothelium such as found in many CNS pathologies: Alzheimer disease, multiple sclerosis, encephalopathy caused by infections, gliomas and metastatic brain tumors, stroke and inflammatory pain. Strategies to restore brain homeostasis are of great therapeutic interest.[30]

1.5.5. Other *in vitro* models: PAMPA assay

Parallel artificial membrane permeability assay (PAMPA) is an *in vitro* transport assay based on non-biological artificial phospholipid membranes. Developed by Kansy *et al*.[66] was initially set up for oral absorption prediction. PAMPA has been widely used in the pharmaceutical industry as a high throughput permeability assay to assess the permeability properties of molecules as soon as possible in drug discovery.[67-69]

PAMPA is a robust high throughput, rapid, easy to manage and inexpensive technique. However, PAMPA is only valid for permeability studies were passive diffusion is involved, since is a phospholipid-based technology, therefore is not useful for the prediction of permeability by active transport.

By modifying the phospholipid composition of artificial membranes, the assay can be used to predict oral absorption,[70] human skin permeation[71] and BBB permeation,[68] among others. Di *et al.*[68] introduced a lipid formulation, commercially available, practical and reproducible that provides reliable predictions. Avanti Polar Lipids, Inc. obtains/harvest this lipid material from porcine brains and this guarantees a higher similarity on lipid composition to brain endothelial cells of the BBB. However, endothelial lipid composition changes depending on age, species and many other factors difficult to stabilize. Thus, a high throughput PAMPA assay will never be an exact mimic of the brain endothelial cell membrane, also because of the addition of dodecane diluent and the non-bilayer lipid structure of the PAMPA artificial membrane. Lipids artificial membrane has an equivalent thickness of 300 lipid bilayers.

PAMPA assay is usually performed on sandwich 96 well plates. Phospholipids lay over polycarbonate filters and a commercial buffer solution (pION) is used as a solvent. Cosolvents such as 1-propanol, ethanol, acetonitrile or DMSO may be required depending on solubility properties of tested molecules. Lower or donor compartment initially contains all sample and upper or acceptor compartment, initially containing solely buffer gather molecules crossing the phospholipid membrane. To avoid unstirred water layers (UWL)[69, 72] a magnetic seed is introduced in the donor compartment and it is stirred by a commercially available mechanical apparatus (Gut-Box™).

Despite this incapacity to mimic other properties of the BBB than passive diffusion, PAMPA is a great model to assess methodologies developed in the present thesis at very reasonable costs and easy handling to confirm or dismiss molecular passive diffusion on the BBB.

2. Combinatorial chemistry

Drug discovery is a complex, time consuming and very expensive path. Historically, the main source of new drugs has been natural products, isolated from plants, animals or fermentation sources.[73-75] Combinatorial chemistry emerged as a new source of potential drugs by creating libraries of organic molecules for massive screening and lead optimisation (by improving hit molecules). It opened a new horizon on reducing the time and costs associated with producing effective new drugs, especially on preclinical drug discovery stages.

High-throughput screening and combinatorial chemistry are two concepts that are closely linked. High-throughput screening of large chemical libraries of compounds is a proven way to identify novel chemical entities that target a biological system of interest. Its ability to quickly assay the biological or biochemical activity of a large number of chemicals leads pharmaceutical industries to use it straightforward. Mainly used for drug discovery process it has been applied in a variety of fields.[76-85]

With the advent of combinatorial chemistry those fields became extremely popular since allowed testing huge amounts of molecules in a very short time at a much lower expense. However, the complete process of drug discovery through this method was not trivial due to assay design and eventual identification of molecules of interest.

Despite combinatorial chemistry initially promised to rapidly deliver new candidates and drugs, early in the 2000s it was clear that after huge investments and enormous expectations created, specially in pharmaceutical industries, combinatorial chemistry partially failed to deliver on its promises.[83, 86, 87] This was probably due to a biased and very optimistic understanding on how combinatorial chemistry and high-throughput screening could contribute on drug discovery. It was clear then, that by increasing screened compounds; library synthesis does not automatically increase the chances of discovering candidates. However, careful understanding of those techniques could report on great achievements on targeted screenings. For example, scaffold library design is of paramount importance to orient the library compounds in a biologically relevant way for a particular target, thus conditioning all subsequent process.

Generally speaking, combinatorial chemistry deals with synthesizing a large number of distinct molecules in a short period of time. This is possible since analogue synthesis is usually performed using the same reaction conditions, a reduced amount of reaction vessels and a spectrum of monomers used as building blocks. Many compounds can be synthesized at a time while a few could be synthesized by simple methodology. Despite this enormous advantage, this very same characteristic may impair library screening and particularly leads to very complex identification scenarios to determine molecules of interest. Identification has been addressed with a variety of methods, none of them definitive or extensive for all situations.

Method of library synthesis is also crucial on the designing of screening method because have a strong influence not only on the techniques used for final identification of molecules, but also on the assay performance and its set up. Some authors report distinct crossing rates of drugs across *in vitro* BBB cell-based models when testing those molecules individually or as a cocktail.[88] They report drug-drug interactions at the level of transporters or metabolic pathways misleading transport results when testing cocktails. However, mixtures produced in some library synthesis workflows lead to molecules showing a broad variability but sharing similar scaffolds (such as those described in 2.2). It could be reasoned that members of those libraries will compete in

many cases for similar receptors or transport mechanisms. Then, in this scenario testing molecular cocktails enhance competition among candidates and read-out results only comprise the most promising molecules.

Combinatorial synthesis was first applied to peptides and oligonucleotides and then was expanded to proteins, synthetic oligomers, small molecules and oligosaccharides. Depending on the library desired distinct preparation methods are required. Peptide libraries can be both chemically synthesized or biologically displayed. Whereas small molecules or oligomers are usually prepared by synthetic approaches, protein libraries are mostly prepared by biological methods.[89]

2.1. First chemical libraries

Combinatorial techniques have their very first origin in Merrifield's work on solid-phase peptide synthesis (SPPS) back in 1963.[90, 91] SPPS opened a new scope on synthetic peptides that speed up in the mid-1980s with an increasing demand for synthetic peptides ranging a varied length and amino acid compositions. Independently, distinct groups developed strategies such as synthesis on pins, on tea bags or on cellulose support that paved the way for serious advances in synthetic combinatorial methods in the 1990's, the so-called *combinatorial chemistry gold era*.[92] Those approaches enabled the synthesis of hundreds of individual compounds for use in bioassays in a fraction of time and cost than earlier methods.

Pins: this synthetic strategy takes advantage of a distribution of amino-functionalized polyethylene pins used as a peptide carriers arranged in 12 columns and 8 rows fitting a 96-well plate. Coupling reactions take place on those plates, so that 96 distinct reactions can take place depending on the amino acid used on each well. Alternatively, washing and deprotection steps take place in suitable reactors since all pins share the same chemistry and thus deprotection conditions. Although this strategy allows low amounts of peptides synthesised in pins (300 nmol peptide/pin), it was successfully used for antibody epitope binding studies attached to the pins.[93] For cleavage a labile linker is required.[94]

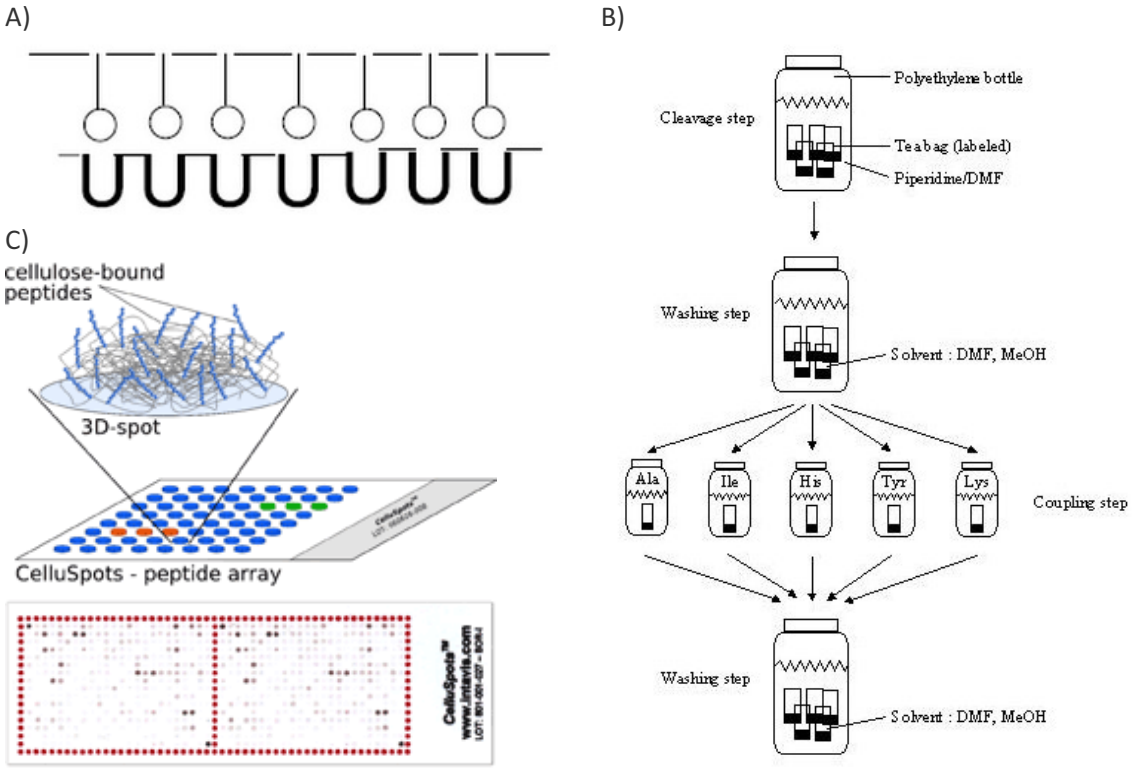


Figure 5. First strategies that enabled synthesis of hundreds of peptides. A) Pins strategy basis. B) Example of synthetic workflow used for tea-bag synthesis. C) Cellulose-based synthesis of peptide arrays.

Tea bags: on this strategy, Houghten *et al.* proposed sealing the synthetic resin in a polypropylene mesh (similar shape than a tea bag).[95, 96] Then, one bag is required for each peptide synthesis. Tea bags are sorted depending on which coupling reaction is required and distributed in reactors with specific amino acid on it. Thoroughly washing has to be performed and deprotection can be carried out joining all bags together in a same reactor. This synthetic cycle is repeated to grow the complete peptide. While this methodology allows synthesizing significant amounts of peptides it requires intense labour.

Cellulose: modifying free hydroxyl groups of cellulose with a TFA-labile anchor an automatic peptide synthesis was proposed performed on cellulose discs stacked in columns. The synthesis took place in a continuous flow synthesized and one peptide was synthesized on each column.[97]

Despite the enormous increase on synthesis of individual compounds that allowed those techniques it became evident that was still not enough to fulfil the desired variability spectrum of compounds for testing in therapeutic bioassays.

2.2. Chemical libraries based on mixtures

2.2.1. Reagent mixtures

Direct evolution of pins synthesis by Geysen *et al.* resulted in the first mixed combinatorial libraries.[98] Instead of individual peptides they synthesized peptide mixtures on pins by coupling mixtures of amino acids instead of individual. Theoretically, this very same concept could be equally applied on tea bags (or simply pools of resin beads) or cellulose papers.

A fundamental problem of this strategy is the variable coupling efficiency of individual amino acids that may lead in the end to a bias on the variability of the library due to distinct concentrations of peptides. In addition, coupling efficiency also depends on the amino acid to which the coupling its being performed. Therefore, the design and use of *equireactive* reagent mixtures requires a thorough knowledge of the mechanism and kinetics involved in coupling reactions.[99] Although not ideal, equimolar amounts of amino acids in an overall ratio of 1:1 to free coupling positions could be also considered, though reaction times might be long.

2.2.2. Resin mixtures

Furka *et al.* described the "split synthesis" methodology and then applied to synthesize libraries of 27 tetrapeptides and 180 pentapeptides.[100, 101] In parallel, so did Lam *et al*.[102] and Houghten *et al*.[103] which investigations share the very same concept applied in distinct purposes. Lam described and took advantage of a uniqueness of this synthesis method for library screening. Because each bead only reacts with one amino acid on each coupling step they reasoned that all peptides bounded in a single bead share the same sequence, leading to one-bead one-compound (OBOC) libraries concept. Then, by testing the libraries with peptides still attached to the resin (orthogonal cleavable linkers were developed[104]) they could physically separate beads that performed positively in an assay to eventually cleave and determine the peptide (or any other molecule depending on library nature) structure.

For synthesis, resin is evenly split in reactors before coupling. One single amino acid is coupled on each reactor. Although this strategy takes more time than using amino acid mixtures, equal concentrations of all individual peptides can be ensured at the end of the synthesis since each coupling reaction can go through completion with no competition of other amino acids. Resin from all reactors is mixed, washed and the *N*-terminal protecting group is removed. Then, this cycle is repeated as many times as required for the library design.

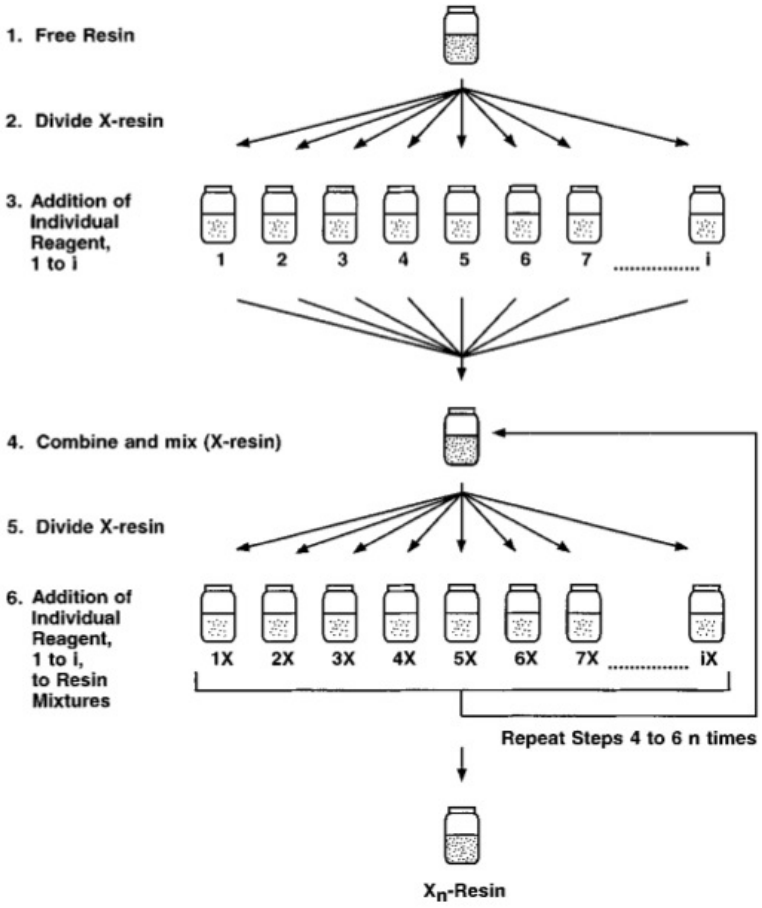


Figure 6. Schematic workflow of *mix and split* synthesis. From Houghten *et al.*[105]

2.3. High-throughput parallel synthesis

Several techniques were proposed to optimize and further develop parallel synthesis. The main objective was to bypass disadvantage features of combinatorial libraries such as deconvolution of mixture-based combinatorial libraries and decoding (structural determination) of active compounds, while keeping the advantageous characteristics (availability of large amount of compounds).

The synthesis was usually performed in a comprehensive rational-designed array. Robotic implementations largely increased the synthesis and assay feasibility of enormous quantities of compounds. Although array synthesis was initially usually used for lead optimization rather than for finding new lead compounds (confining the amount of compounds to synthesize), new techniques developed last decade makes this approach very appealing and widely used for drug discovery process both in pharmaceutical industry and academia.

The synthesis of many single compounds in milligram quantities introduces a new set of challenges from the analytical point of view. The main advantage of this synthesis method is that avoid analysis of very similar and redundant compounds together. Amounts to be quantified are sufficient for most used modern analytical methods such as HPLC, MS or NMR. Then, as it is for synthesis, throughput of those analytical techniques is one of the milestones to achieve rapid detection and eventually quantification of tested molecules.[106]

2.4. Compound identification strategies

Structural identification of library components in mixed synthesis libraries is nowadays the bottleneck of the process to find leads by combinatorial chemistry. This very challenging

identification step leads to a wide variety of approaches that intended to identify compounds of interests with sophisticated and imaginative tactics. Those approaches can be divided in:

2.4.1. Direct analysis: Microanalysis

Direct analysis is always the first choice for identification of the compounds of a synthetic library. Three main approaches depending on the library nature can be employed to address this purpose.

Peptide libraries with free *N*-terminus can be sequenced by Edman degradation. Similarly, oligonucleotide libraries can be analysed by DNA sequencing techniques. Finally, mass spectrometry has an extraordinary potential in this field being nowadays the gold standard for analysis since basically embraces libraries of all natures. For this reason and because it plays a principal role in the present thesis a complete section of the introduction is devoted to mass spectrometry.

2.4.2. Deconvolution

Deconvolution approaches rely on the synthesis and test of sublibraries; fractionated or simplified libraries to deduct the active compound within the library. Most impact deconvolution approaches has been iterative deconvolution,[98, 103] positional scanning,[107, 108] orthogonal deconvolution[109] and recursive deconvolution.[110]

2.4.3. Encoding/Decoding

Since no system of analysis has been established for identifying members of a mixed combinatorial library, various encoding strategies have been developed to make this analysis possible. Encoding means the introduction of tags or labels, which translate or express all structural information in a way that can be quickly and simply read with any specific technique. Careful must be taken on those encoding strategies since tag may interfere in the biological assays.

Here we review in a non-systematic way the most impact proposed techniques. Only some of them are discussed in a bit more detail due to their significant development from their proposal in the late-90s to date. Compound classes that allow fast and easy trace analysis are mainly oligonucleotides (PCR amplification and subsequent sequencing), peptides or amino acids (Edman degradation or mass spectrometry), haloaromatic compounds (gas chromatograpy) and secondary amines (HPLC after derivatization). Usage of any of those possibilities relies on nature of synthesized library and availability of analytical techniques.

Direct encoding is the simplest option in case library is composed by linear oligomers. An *n*-mer ligand sequence (for screening) is represented by an *n*-mer tag sequence, which encodes it structural information. This approach was used for example to encode peptide libraries with non-natural amino acids.[111]

Similarly, other tags such as amino acid or peptide tags were described translating non-natural amino acids with amino acid or peptide tags.[112, 113]

Luo *et al.* proposed a colour-encoding method based on staining with dyes resin beads of OBOC libraries. This strategy allows staining distinct sublibraries with a variety of colours. Rapid and easy identification of *in vitro* tests lead to detection of preferred sublibraries for further screening. This methodology can be applied to alanine scan to facilitate identification.[114]

Another approach to identify peptides sequences is the so-called ladder approach in which after each coupling and deprotection step small percentage of peptides in the resin are capped with

acetic acid. At the end of the synthesis, a ladder of distinctly sized peptides corresponding to the synthetic route is present on each bead facilitating sequence elucidation by mass spectrometry. This methodology requires assay libraries on beads and finally cleave peptide of each bead separately and MS analysis.[115, 116]

Binary encoding approach using molecular tags, gives information about the coupling events depending on the presence or absence of tags. Haloaromatic compounds[117, 118] or secondary amines[119] are especially well suited for this strategy.

Isotopically labelled encoding tags approach introduced to the synthesis isotope mixtures of an amino acid creating a peptide fingerprint on their mass spectrometry spectrum that serves as a bar code for peptide identification.[120]

Oligonucleotide tags

Despite the importance of the aforementioned techniques for the advance of the field and its implementation in specific situations, the gold standard approach that lasted and evolved is the DNA-encoded libraries approach. This concept was first introduced by S. Brenner and R. Lerner[121] in 1992 and revival in the 2000s in contrast to other combinatorial chemistry strategies.[122]

Diverse approaches to synthesize DNA-encoded chemical libraries had been proposed most of them initially based on split methodology. One of the main concerns and difficulties for this approach was the use of orthogonal and compatible synthetic conditions for both displayed synthesis in polymers such as a bifunctional serine anchor. The peptides were synthesized at the amino group of the serine (using an spacer to avoid steric hindrance when growing both oligomer chains) using Fmoc strategy and the encoding oligonucleotides were synthesized with the phosphoramidite method at the side chain of the serine (DMT, dimethoxytrityl protecting group).[123]

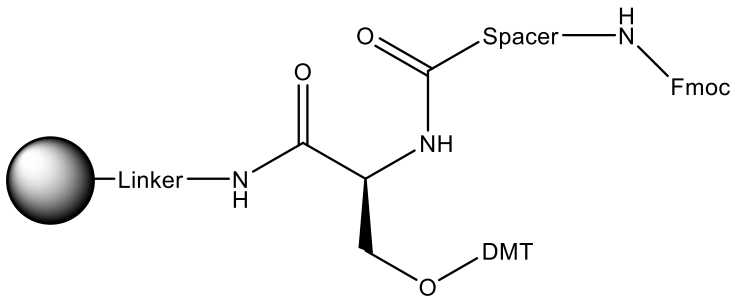


Figure 7. Direct encoding of a peptide library by oligonucleotides according to Nielsen *et al.*[123] A bifunctional serine anchor attached to the solid support by a linker enables parallel synthesis of two oligomers: peptide oligomers through Fmoc/*t*Bu strategy and oligonucleotides with phosporamidite method. Spacer significantly improves the synthesis avoiding steric hindrance of the growing oligomers.

Modern approaches to DNA-encoded chemical libraries omit the polymer beads used to perform the synthesis and have the chemical compounds directly conjugated to DNA. Later on, Nuevolution and Praecis companies patented an enzymatic ligation strategy for DNA code assembly, which allow sequential chemical synthesis and DNA-tagging steps.[124, 125]

DNA-template chemistry represents another avenue to the construction of DNA-encoded libraries. The proximity conferred by the hybridizing DNA strands increases the effective molarity of the reagents and allows the chemical reactions to occur at very low concentrations. This synthetic pathway was introduced by D. R. Liu and very much exploited from date.[126-129] Small organic compounds can be coupled to biotinylated DNA fragments (donor) and transferred to suitable complemetnary DNA strands (acceptor). Donor strand is subsequently cleaved and removed using avidin-coated beads.

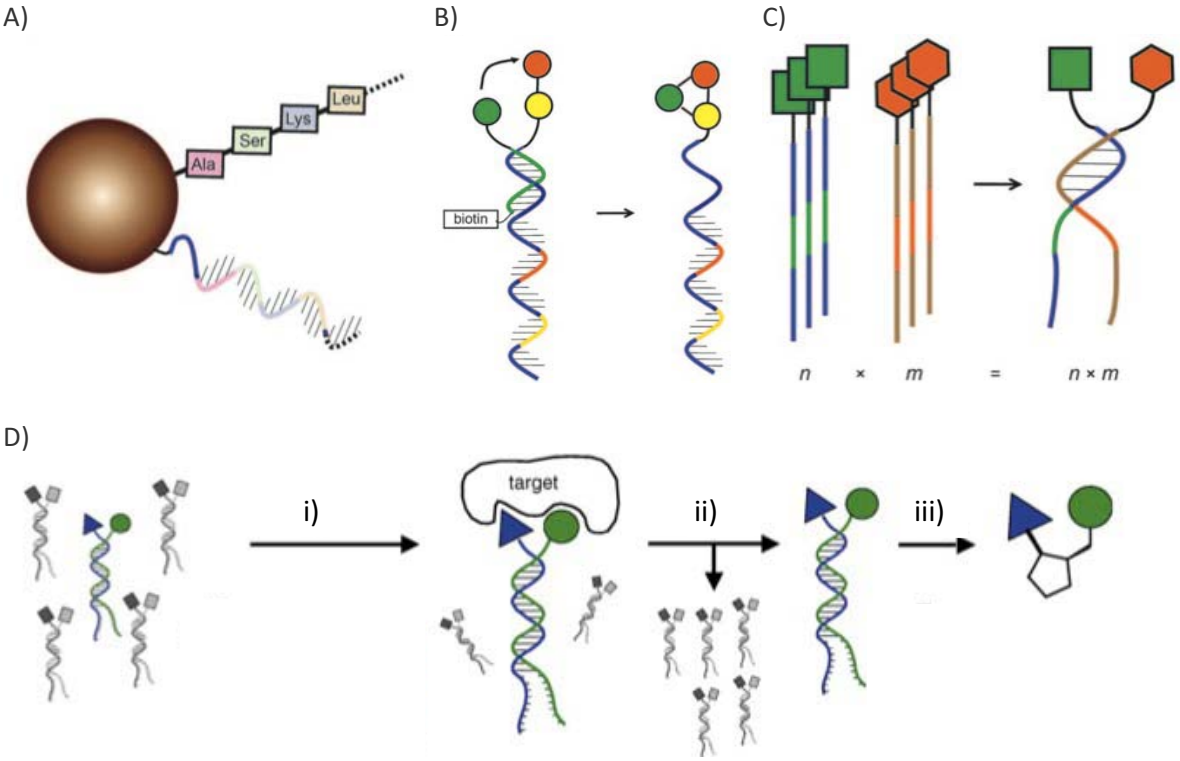


Figure 8. A) Alternate peptide and DNA synthesis on beads. B) Template-assisted sequential addition of chemical groups from reagent oligonucleotides onto performed encoding template oligonucleotides. C) ESAC libraries; two DNA-encoded sublibraries are hybridized to form a dual pharmacophore library. D) i- Generated ESAC library and assay against a target protein. ii- Washing for removal of unbound molecules. iii- Validated binding partners are conjugated with a suitable scaffold yielding small-drugs. Figure adapted from J. Scheuermann and D. Neri.[122, 130]

Alternatively, library formats presenting two separate molecular entities were described. Large DNA-encoded chemical libraries can be constructed by combinatorial self-assembly of relatively small complementary DNA libraries. Two sublibraries with 1000 members each yield 10^6-member encoded self-assembly chemical library (ESAC library) in duplet format.[131] Relative flexibility of DNA fragments carrying neighbour moieties allow them to adapt onto the target protein surfaces. Eventually, validated and positive binding partners are conjugated to suitable scaffolds yielding small molecules holding explored activity.[132-135]

DNA-encoded chemical libraries are now being actively used in hit and lead discovery for both academic and industrial companies with successful results.[136] Not only for *de novo* discovery of small-molecule binders of target proteins but also for the systematic exploration of binding space of lead compounds.[137]

3. Mass spectrometry and proteomics: a crossroad

In the last decade, mass spectrometry definitively became the best partner for proteomics. Advances on MS equipment allowed intensely refinement of proteomic techniques. In a way, proteomics needs drive MS improvement to fulfil its challenges. However, ideal identification and quantification of complex protein mixtures is still not fully achieved.

Accurate quantification of specific set of peptides or proteins across multiple samples is required to acquire reliable data to postulate conclusions or establish models for the studied processes or systems.

3.1. Instrumentation and approaches for analysis of complex mixtures

For exhaustive analysis, peptides in a mixture are not introduced altogether to the mass spectrometer. Instead, high-performance liquid chromatography (HPLC) is usually coupled to the mass spectrometer. Peptides are eluted from these columns using a solvent gradient of increasing organic portion to fractionate complex samples to simpler ones that can be better adressed by the MS coupled on-line. As HPLC is strictly used to simplify the sample there is no need to achieve a non-overlapping peaks chromatogram since MS will further differentiate peptides depending on their m/z ratio. Since signal intensity in the mass spectrum depends on analyte concentration (speaking in global sample terms) microscale capillary columns (50-150 µm of inner diameter) are used to deal with small elution volumes, thus more concentrated.

3.1.1. Molecular ionization on mass spectrometry

Any molecule analysed by mass spectrometry has to be in the gas phase. Several ionization techniques had been applied to this purpose. However, analysis of proteins and peptides require a soft ionisation technique to avoid the fragmentation and maintaining the fundamental structure of these molecules at that stage.

Several ionization techniques had been described, but here we strictly focus to matrix-assisted laser desorption/ionization (MALDI) and electrospray ionisation (ESI), which are especially suitable for this purpose.

MALDI consist on ionization of the sample (analyte) by irradiation with a focused laser beam of an appropriate wavelength, which has been mixed and dried with a large excess of ultraviolet-absorbing matrix (usually a low-molecular weight aromatic acid). On this process matrix molecules sublime and transfer the embedded non-volatile analyte molecules into the gas phase. Singly protonate analyte ions are formed after numerous ion-molecule collisions. Finally, those ions are accelerated and directed to the mass analyser by applying a series of electric potentials. MALDI does not allow direct on-line coupling to HPLC, but automatic robots to deposit LC fractions had been designed. MALDI mass spectrometers are usually capable of peptide fragmentation for further peptide analysis.

Electrospray ionization (ESI) is mediated by the action of a strong electric potential.[138] The liquid of the sample flowing through a needle is vaporized and the molecules are subsequently ionized. Commonly coupled to a LC, ESI ionize the analytes by driving the sample flow through a narrow metal needle, which held a high electrical potential with respect to the entrance of the mass spectrometer. The liquid containing the sample (that exits HPLC column) is electrostatically dispersed. This generates highly charged dropleds due to an excess of protons. For the action of an inert gas, the solvent evaporates, which decrease the size and increase the charge density of the droplets. Desolvated ions are generated by desorption of analyte ions from the droplet surface due to high electrical fields until formation of droplets containing single analyte ions.

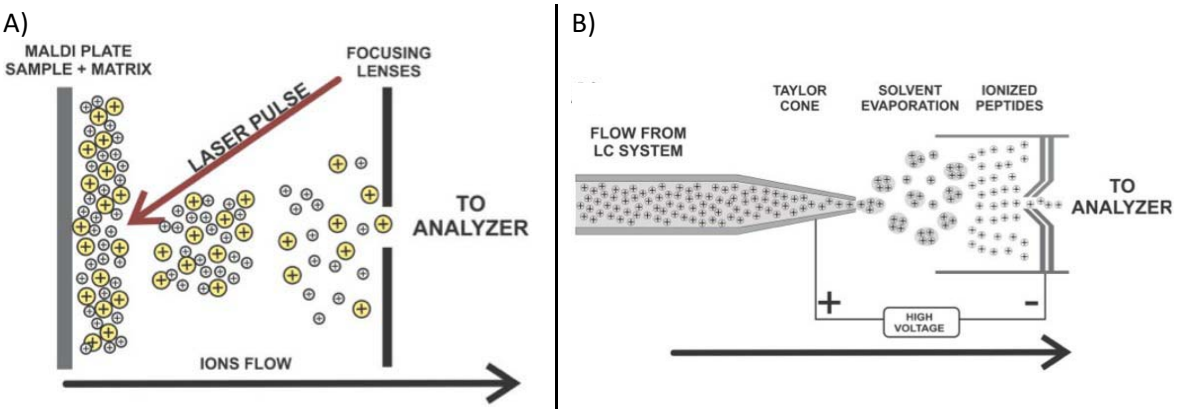


Figure 9. A) The ionization process by matrix-assisted laser desorption/ionization (MALDI). The laser pulse strikes the MALDI plate, in which the sample (yellow) is co-crystallized with the matrix (grey). Both sample and matrix molecules can reach the detector. B) Simplified scheme of electrospray ionization (ESI). The positively charged droplets and desolvated peptides are represented by signal +. Voltage applied pushes droplets away from the column. On the way solvent evaporates and ionized peptides directed to the analyser.

Detection of ionized species (molecular or peptide ions) yields to the so-called MS1 spectra and draws a picture of the composition of the sample in terms of m/z. Since ionization strongly depends on molecular nature, it cannot be ensured that MS1 strictly represent all analytes in the sample in its accurate proportions. Thereafter, MS is not a quantitative technique *per se*. Peptides exhibit a variety of physicochemical properties that lead to large difference and variable performance in response to mass spectrometry. Subsequently, lots of efforts in the field had been devoted to develop diverse workflows to achieve either relative or absolute quantification results. This topic will be further discussed later on this thesis.

3.1.2. Fragmentation and sequencing

Despite mass accuracy and MS1 spectra contribute on sample characterisation, other spectrometric techniques are required to further study each compound to fully elucidate its structure. Tandem mass spectrometry (MS/MS) involves the activation of a known precursor ion formed in the ion source and the mass analysis of its fragmentation products.

The ion activation or fragmentation method is crucial to the experiment because eventually defines which molecular fragmentation ions result. Several ion activation methods are described such as collision-induced dissociation (CID),[139, 140] electron-capture dissociation (ECD),[141] electron-transfer dissociation (ETD),[142, 143] negative electron-transfer dissociation (NETD),[144] higher-energy C-trap dissociation (HCD),[145] surface induced dissociation (SID),[146] photodisociation,[147] among others. It is not the aim of this introduction to review this field. One can find in-depth revisions in the field.[148]

When a peptide is fragmented by imparting energy onto the molecule, several cleavage sites can be produced depending on the method used. Fragmentation mainly occurs on peptide backbone bonds. Usually, Roepstorff-Fohlmann-Biemann nomenclature is used (Figure 10.A).[149, 150]

Tandem MS yield to spectra with peaks representing all derived ions produced after fragmentation of a selected molecular/peptide ion, called precursor ion or parent ion. Each peptide fragment (product ion or daughter ion) differs from its neighbours by a distinct composition. Therefore since all peaks represent fragmented species of the same peptide it is therefore possible to determine the amino acid sequence by considering the mass difference between peaks in a series (Figure 10.B).

However, information provided by MS/MS is most often not complete. This happens because peptide or molecular fragments are not produced or because they might not be visible for the detector (most likely because they do not carry any charge). Then, identification of peptide

sequences is not trivial. Furthermore, isobaric amino acids or isobaric combinations of amino acids or fragments sharing very similar molecular weight (not distinguishable considering the error of the measures) may lead to spectra interpretation confusions and incorrect assignments. Moreover, in complex mixtures precursor ions may not be fully isolated to enter the fragmentation cell if there are other ions with similar molecular weigh and identical retention time in the MS1 spectrum. Then fragmentation spectra can incorporate signals not corresponding to the targeted specie enormously hampering the assignment. Great effort it is being devoted for narrowing m/z selection windows to isolate precursor ions. Best instruments achieve around 0.5-1 Da windows.

Several *de novo* sequencing computer algorithms had been designed but are still not fully reliable for proteomic approaches, which inevitably deal with complex mixtures of peptides.

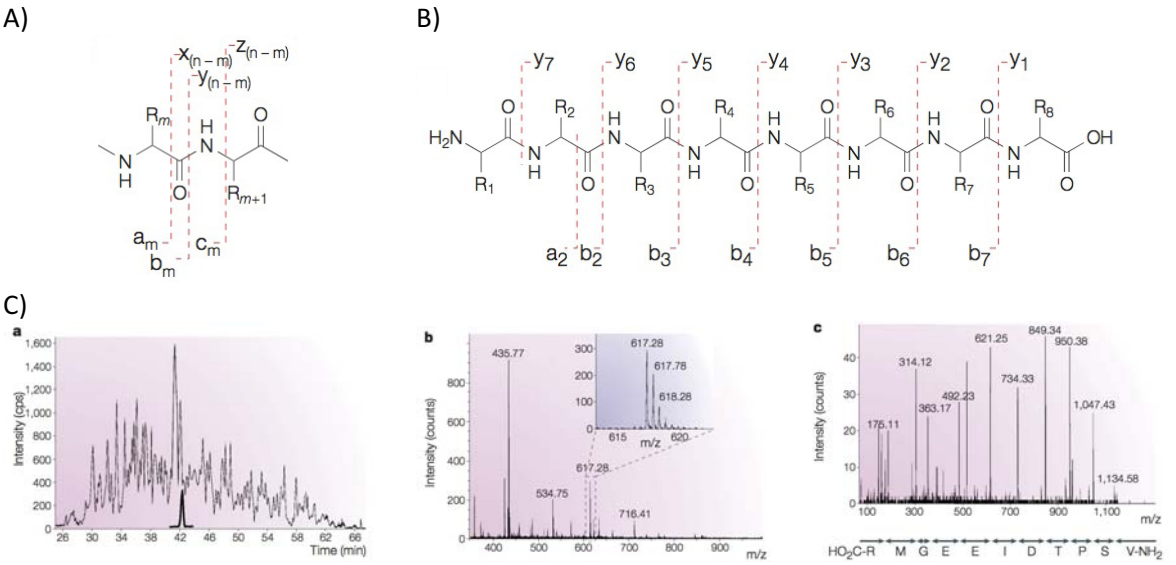


Figure 10. A) Theoretical backbone fragmentation pattern depending on fragmentation technique; Roepstorff-Fohlmann-Biemann nomenclature. B) Typical backbone fragmentation of a peptide on CID fragmentation. C) Sequence followed for MS/MS identification. Selecting a given fraction in the TIC of HPLC (a) to perform a MS1 spectra containing all peptide ions in the sample (b). Fragmentation of a selected m/z results on MS/MS spectra (c); differences on neighbouring peaks leads to amino acid sequence determination. Modified from Steen, H. *et al*.[151]

3.1.3. LTQ-Orbitrap instrument

Mass spectrometers usually consist on the following parts: ion source, mass analyser and detector. Mass analysers store and separate ions based on the mass-to-charge (m/z) ratios. Mass analysers can be categorized depending on how they separate ions. Trapping mass analysers such as ion trap (IT), Orbitrap, and ion cyclotron resonance (ICR) separate ions based on their m/z resonance frequency. On the other hand, the so-called ion-beam mass analysers such as quadrupole, which is based on m/z stability or time-of-flight (TOF) on ion flight time. Hybrid mass spectrometers have been built to enhance performance by stressing advantageous features of combined analysers.

It is not the aim of this introduction to fully review neither mass spectrometry instrumentation nor mass analysers. Excellent reviews and textbooks had been published on those topics.[152-155]

However, LTQ-Orbitrap deserves a special attention since has been critical for the development of the present thesis. LTQ-Orbitrap is a hybrid instrument. Orbitrap uses orbital trapping of ions in its static electrostatic fields in which ions orbit around a central electrode and oscillate in axial directions.

The Orbitrap mass analyser features high resolution (up to 150000), high mass accuracy (2-5 ppm), a mass-to charge range of 6000, and a dynamic range greater than 10^3. When coupled to LTQ ion

trap, the hybrid instrument takes advantages of best Orbitrap features (high resolution and mass accuracy) and LTQ (speed and sensitivity). Then, LTQ-Orbitrap can be operated in parallel: while Orbitrap acquires MS full scans, LTQ carries out fragmentation reactions. These features makes this equipment specially suited for analyse complex protein mixtures.[156-158] Such measurements achieve extremely high resolution rivalling that of FT-ICR (Fourier-transform ion cyclotron resonance) instruments. This feature allows accurate mass acquisition with Orbitrap equipment.

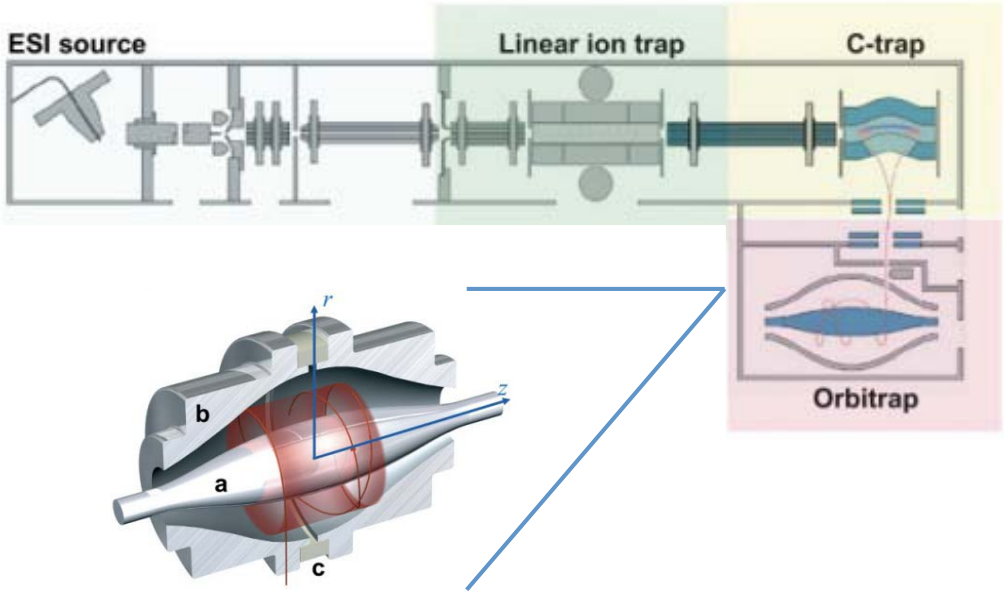


Figure 11. Scheme of a hybrid linear ion trap/orbitrap MS (LTQ-Orbitrap). The front part is an LTQ linear ion trap mass spectrometer capable of detecting MS and MS/MS spectra. In the C-trap, ions are accumulated. Ions are then inserted into the orbitrap and finally their signal detected. At bottom-left a model of the Orbitrap mass analyser: the ions suffer an electrostatic attraction towards the central electrode (a) that is compensated by a centrifugal force that arises from the initial tangential velocity of ions. Electrostatic fields force ions to move in complex spiral patterns. Axial component of these oscillations can be detected as an image current thanks to the two halves of the electrode encapsulating the orbitrap. These outer electrode (b) is split in two by a ceramic ring (c). Adapted from M. Scigelova *et al.*[159]

3.2. Sequencing peptides, the first step to sequence whole proteins

3.2.1. Shotgun approach

Proteomic studies are commonly performed by shotgun approach.[154, 160, 161] The also called bottom-up proteomics approach aims to identify proteins in a mixture by enzymatically digesting the sample and analysing the subsequent formed peptides. Since *de novo* sequencing approaches are still not mature enough, other tactics had been considered.

Since a peptide-fragmentation spectrum might not contain enough information to unambiguously elucidate the complete amino acid sequence the new approach consisted on matching the information of fragmentation spectrum with database information of the screened sample. This database-matching approach facilitates the assignment because only an infinitesimal fraction of all possible peptide amino acid sequences occur in nature. Although a peptide-fragmentation spectrum would not be complete it might have enough information to match uniquely to a peptide sequence in the database. This matching is done on the basis of the observed and expected fragment ions. There are several algorithms used to this purpose such as SEQUEST,[162] Mascot,[163] MassMatrix,[164] OMSSA,[165] and X!Tandem,[166] among others. An obvious limitation of this approach is that can only be applied to samples coming from organisms that have had their genome sequenced, so that all possible peptides are known.

In this approach only a subset of the enzymatically degraded peptides present in the sample analysed by mass spectrometry are selected by the mass spectrometer in a process called data-

dependant precursor selection. However, this process is not optimal for a systematic quantification of these proteins because this precursor selection has a stochastic nature.

Nowadays, shotgun approach has still not achieved the quantitative proteomic analyses to the extent transcriptome did with microarray technology.[167] Main shortcomings of shotgun proteomics are limited sensitivity and poor reproducibility of target selection, which frustrates a faithful identification of protein sets in similar samples.

3.2.2. Targeted proteomics

Alternatively, MS approaches based on selected reaction monitoring have emerged to precisely and quantitatively analyse complex biological samples. This technique was introduced in the late 1970s when first triple quadrupole mass spectrometers were developed.[168, 169] Originally it was applied to the measurement of small molecules (metabolites and drugs),[170] but was not until last decade that was applied to proteomics. [171]

Usually coupled to liquid chromatography for better efficiency, selected reaction monitoring (SRM) takes advantage of the unique capabilities of triple quadrupole (QqQ) MS for quantitative analysis. First and third quadrupoles (Q_1 and Q_3) are used as filters to specifically select specific m/z values corresponding to the peptide ion (precursor ion) and a specific fragment ion of the peptide respectively, whereas the second quadrupole (q_2) act as collision cell. Since QqQ is coupled to LC this technology allows monitoring several transitions (precursor and fragment ion) over time, yielding a set of chromatographic traces with the retention time and signal intensities for each monitored transition. High selectivity results are achieved from this process since two mass selection windows (precursor ion and fragment ion) are applied which restrict the unspecific ions gate-crashing into the analyser.

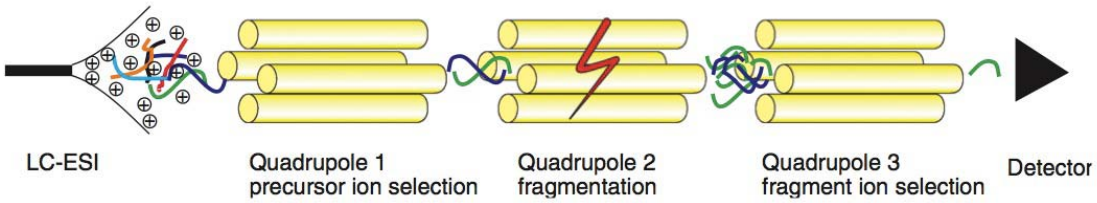


Figure 12. Scheme of a SRM analysis on QqQ MS. Several analytes are coeluting from the chromatographic system. The first quadrupole (Q_1) filter out most coeluting ions by their m/z ration. However, two ions (green and blue) are selected since they share the same m/z. Second quadrupole (q_2) fragments the analytes. The m/z selection of the third quadrupole (Q_3) filters out a specific fragmentation ion corresponding to fragment of the green precursor ion. A transition results from a selected combination of Q_1 and Q_3 filters. From Lange *et al*.[171]

Due to its non-scanning nature, this operation mode offers an increased sensitivity up to one or two orders of magnitude compared to conventional scanning techniques. Furthermore, it results in a linear response over a wide dynamic range up to five orders of magnitude. This allows detecting low-abundance analytes in high complex mixtures and it is especially interesting to achieve systematic quantitative studies.[172]

Since SRM on a QqQ MS exhibits outstanding selectivity, sensitivity and dynamic range over other MS techniques it has been applied to address major proteomic challenges.

Contrarily to shotgun proteomic studies, SRM targets a restricted collection of pre-set peptides. Moreover, only a limited amount of transitions corresponding to each selected peptides are monitored. Then, an SRM-based proteomic experiment strongly relies on its initial design.[172] Careful must be taken on selecting unique peptides (proteotypic peptides[173]) that optimally represent the targeted protein set, the selection of the best set (most intense) of SRM transitions for each selected peptide and optimizations of other parameters such as collision energy. Then in SRM experiments previous information is required to define these transitions.

3.3. MS as a quantification tool: SRM quantification

The final goal of an SRM experiment is the quantification of a set of targeted proteins (or any targeted molecule). Depending on the nature of the experiment and the information the user is willing to obtain, rather relative or absolute quantification could be determined.

Label-free quantification approach can be based on the signal intensities of specific SRM transitions by direct evaluation of signal intensities of peptides contained in the sample.[174] Despite its lower precision compared to label-based approaches it is straightforward and it might be useful. On the other hand, relative quantification of the same analyte in distinct samples can also be achieved by LC-MS analysis under the very same conditions. However, satisfactory relative quantification by this approach is not trivial due to variations in signal intensities depending on sample, time of analysis and other factors that cannot be precisely controlled such as fluctuations in ionization efficiency or matrix effects.[171]

The addition of isotopically labelled internal standards avoids the experimental and instrumental variability. Moreover, early isotopic label introduction in the workflow is required to decrease technical variability. Quantification is then based on relative intensities of the analyte signals compared with that of the isotopically labelled internal standard. Most labelling approaches applicable for quantitative shotgun proteomics such as ICAT,[175] SILAC[176] or ICPL[177] among others[178] can fit SRM-based quantitative experiments.

An intermediate solution between the label-based and label-free approaches is the labelled reference peptide (LRP) method in which a single labelled peptide is used as a reference standard for all other measured endogenous peptides.[179]

The gold standard for absolute quantification is stable-isotope dilution (SID). It was first applied to peptides in 1983[180] and later on, on protein digests.[181] Any mass spectrometry technique can take advantage on SID, but only triple quadrupole instruments fully exploit the MS/MS potential in SID-SRM. The most used approach uses isotopically labelled reference peptides that are chemically identical to the light native peptides (commercially known as AQUA peptides).[182] Since intrinsic issues of this methodology such as incomplete digestion or artifactual modifications may lead to slight errors on analytical precision (up to 5%),[172] absolute quantification is sometimes referred to as *precise relative quantification*.[183]

In SID approach isotopically labelled reference peptides must be spiked into the sample. As the amount of those spiked peptides is precisely determined, the absolute amount of the targeted peptide or protein can be determined from the relative intensity of the light/heavy transitions. Peptides or proteins should be spiked in amounts to produce relative intense signals allowing high precision quantification. Excessive amounts of internal standard might saturate the detector and ruin the experiment. It should be noted that stable isotope-labelled peptides might have a purity of 98-99% that might bias quantification. For this reason, previous quantification experiment, isotopically labelled peptides should be injected alone analysing transitions for both heavy and light peptides. Depending on isotopic labelling nature those errors might be even larger. Alternatively to simply spike peptides, another methods spikes in full-length isotope-labelled proteins into the sample.[184]

For further discussion on absolute quantification in proteomics and method comparison we refer to those studies.[179, 185]

4. Phage display among display technologies

Display technologies are based on the ability to physically link the polypeptide produced by a library clone to its corresponding genotype. This allows to recover the DNA encoding the clone selected based on the desired polypeptide phenotype. Display techniques comprise phage display, cell display, ribosome display, mRNA display and DNA display. They have been widely used for high-throughput molecule selection.

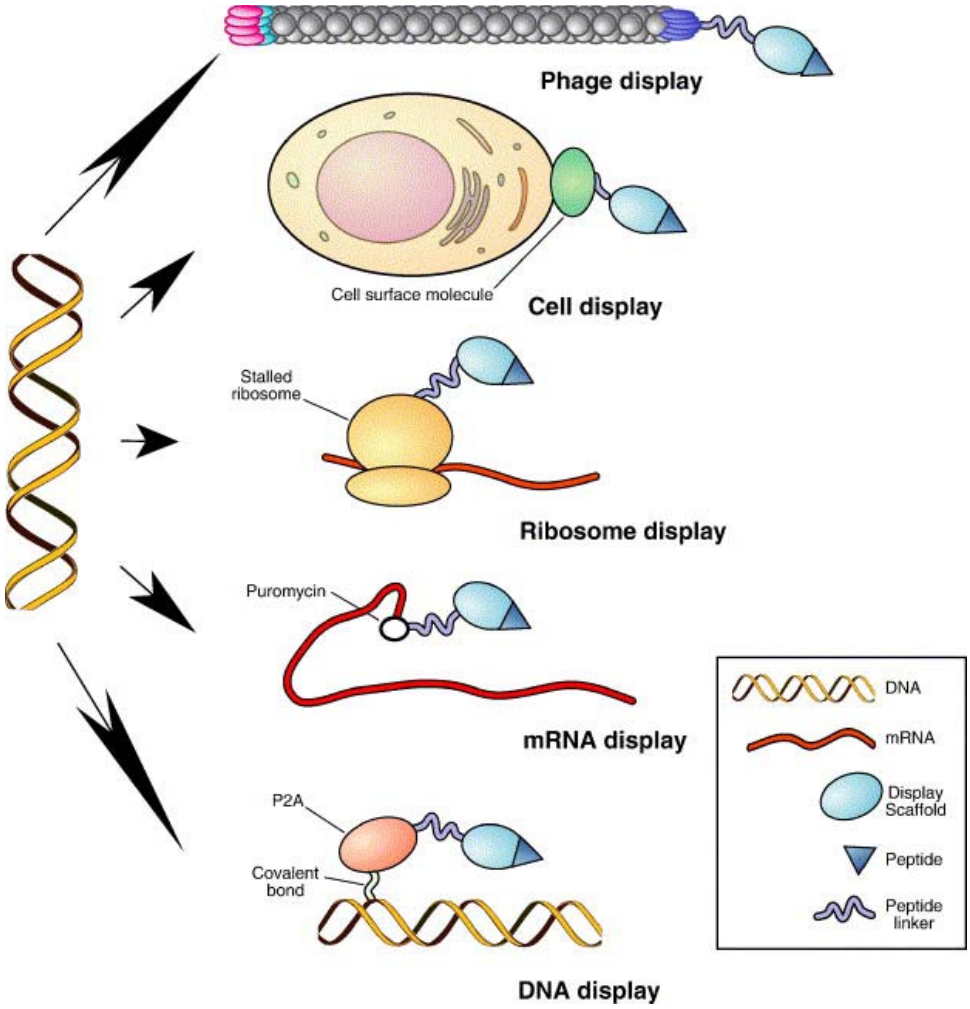


Figure 13. Scheme of available display technologies. All display platforms are based on the ability to physically link the polypeptide produced by a library clone to its corresponding genotype. This allows one to recover the DNA encoding the clone selected based on the desired polypeptide phenotype, such as binding to the target. In phage/virus display, linkage between the gene and the encoded polypeptide is achieved by expression of the polypeptide as a fusion with a coat protein from DNA packaged in the same particle. In cell display, linkage between the gene and the encoded polypeptide is achieved by expression of the polypeptide as a fusion with a cell surface molecule from DNA, which the cell receives in order to be transformed. In ribosome display, linkage between the gene and the encoded polypeptide is achieved by stabilization of complexes between the ribosome, mRNA and the encoded polypeptide upon termination of elongation with a permissive marker, such as chloramphenicol or low temperature. In mRNA display, linkage between the gene and the encoded polypeptide is achieved by a puromycin molecule covalently bonding the mRNA 3' and the translated polypeptide upon the ribosome stalling at the junction of mRNA and an engineered single-stranded DNA linker. In covalent DNA display, linkage between the gene and the encoded polypeptide is achieved by covalent bond that forms between the DNA-binding protein P2A (produced as a fusion with polypeptide) with the DNA encoding the fusion. Figure is not depicted in scale. From Sergeeva, A. *et al.*[186]

Among all of those, phage display technique is by far the most used high-throughput screening tool for the evaluation of enormous amounts of peptides.

4.1. Filamentous phage biology

The filamentous bacteriophages (*Genius Inovirus*) are a group of viruses that infect prokaryotic cells. Bacteriophages (also simply called phages) contain a circular single stranded DNA genome encapsulated in a long protein capsid. With the advent of genetic engineering in the late 1970s, phage-based vectors were among the first cloning vehicles.[187-189]

Ff class of filamentous phage (f1, fd, and M13) use the tip of F conjugative pilus as a receptor and thus are specific for *E. coli* containing the F plasmid. DNA of these three phages are 98% homologous; consequently the protein sequences of the gene products are practically the same.[190]

The Ff phages do not kill the host during their infection and replication cycle. The single-stranded viral DNA is replicated via a double stranded intermediate by a mixture of bacterial and phage encoded components. The result is a newly synthesized viral single-stranded DNA. The capsid proteins are all synthesized as integral membrane proteins that remain in the membrane until they are assembled around the DNA. Bacteria tolerate this process and continue to grow and divide with a generation time approximately 50% longer than that of uninfected bacteria.

4.2. Ff bacteriophage

4.2.1. Phage structure

The Ff phage particle is approximately 6.5 nm in diameter and 930 nm in length. The mass of the particle is $16.3 \cdot 10^6$ Da. 87% of phage particle are proteins. 11 distinct proteins had been described to be present in phage in different amounts. Function, location and exact interactions among all of them are still not completely understood. The most abundant protein is the so-called pVIII, which conforms the whole length of the capsid cylinder. Approximately 2700 pVIII protein conforms a, in some extent, flexible capsid that encapsulates the phage genome.

To form the capsid, pVIII protein monomers are arranged in an overlapping shingle-type array with a symmetry defined by a five-fold rotational axis with a two-fold screw axis of pitch 3.2 nm. The axis of the helical pVIII monomer is tilted 20º to the long axis of the particle.[191-194]

In one end of the virion there are approximately 5 copies of pVII and pIX, whereas the other end contains 5 molecules of pIII and pVI.

pVII and pIX are small hydrophobic proteins. 5 copies of each confer the first assembled phage region. It is suggested that only pIX is exposed whereas pVII is buried.[195] On the other end, pIII is been described to have three domains (N1, N2 and CT). N1 contains 68 amino acids and it is the *N*-terminal domain, which is required during infection. A glycine-rich region links N1 domain with N2, made up by residues 87-217 of the protein. This domain is responsible for binding to the F pilus.[196] Those two domains are exposed on the surface of the phage and removal of them by protease action lead to non-infectious phage.[197, 198] Finally, CT domain is made up of the 150 amino-terminal residues. CT is supposed to interact with pVI protein and it is essential for forming a stable phage particle.[199-201]

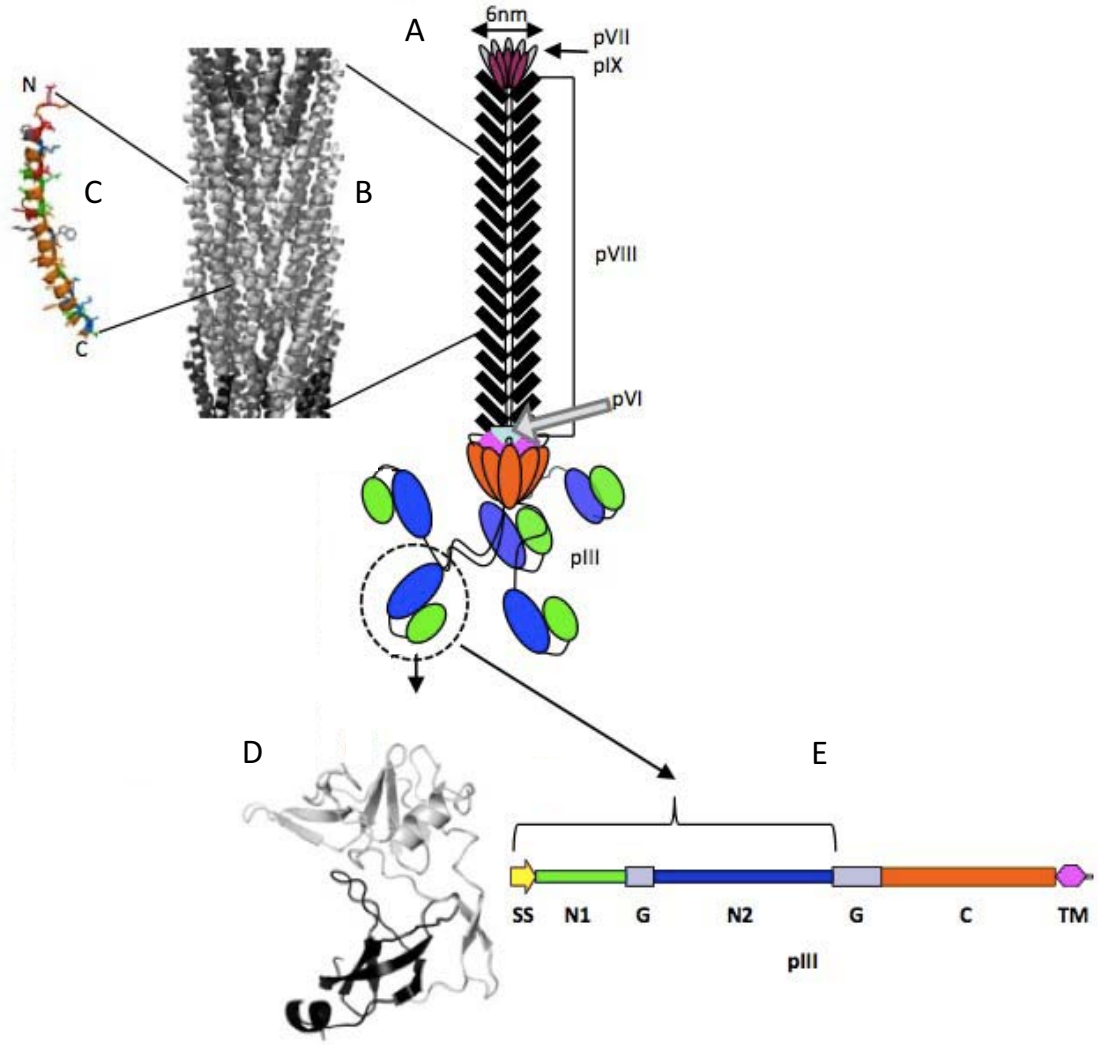


Figure 14. Schematic representation of a phage virion (A) with multiple copies of pVIII form the capside while only 5 copies of pIII protein are present in one end of the viron. Ribbon representation of the pVIII arrangement within the filamentous phage capsid forming shingle-like helices array (B). Model of a pVIII monomer (C) with hydrophobic residues (green), hydrophilic residues (orange), positively charged residues (blue) and negatively charged residues (red). Ribbon representation of pIII protein (D) with N1 (dark grey) and N2 (light gray) domains. Domain organization of pIII preprotein (E): SS, signal sequence, N1, N2, CT domains of pIII; G, glycine-rich linkers; TM, transmembrane helix. Modified from Rakonjac, J. *et al*.[202]

4.2.2. Phage genome considerations

The genome is a single-stranded circular DNA that comprises about 6400 nucleotides. It is located inside the pVIII protein capsid. The genomes of Ff phage have been completely sequenced.[203] All of them encode 11 genes producing the above-mentioned 11 proteins.

The genes are grouped in the genome according to their function in the life cycle of the phage. First group (*genes II*, *V* and *X*) encodes the proteins required for the replication of the phage genome. Second group encodes the capsid proteins (pVII, pIX, pVIII, pIII and pVI). The last group encodes three proteins that are involved in the membrane-associated assembly of the bacteriophage (pI, pXI and pIV).

Phage genome is highly efficient since there are very few regions that do not code for protein. Cassettes encoding antibiotic resistance are generally inserted in the intergenic region or in the space between the end of *gene VIII* and the beginning of *gene III*.

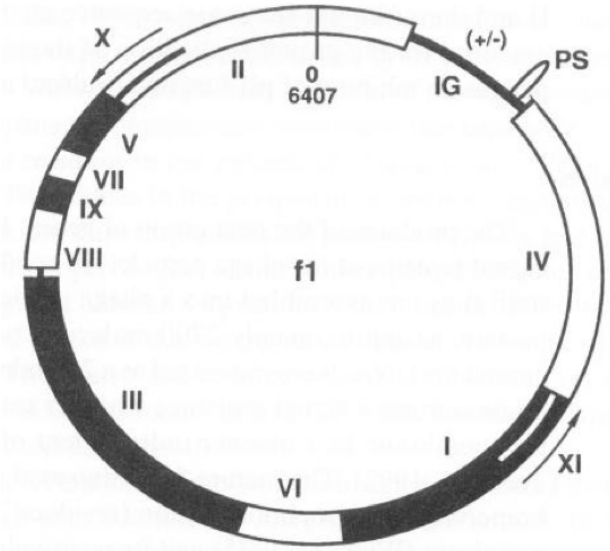


Figure 15. Genome representation of f1 bacteriophage. Modified from Barbas III, C. F.[190]

Gene	Function	Protein amino acids	Protein MW
II	DNA replication	410	46,137
X	DNA replication	111	12,672
V	Binding ssDNA	87	9,682
VIII	Major capsid protein	50	5,235
III	Minor capsid protein	406	42,522
VI	Minor capsid protein	112	12,342
VII	Minor capsid protein	33	3,599
IX	Minor capsid protein	32	3,650
I	Assembly	348	39,502
IV	Assembly	405	43,476
XI	Assembly	108	12,424

Table 2. Function of proteins encoded by each gene of f1 bacteriophage, taken as an example. Number of amino acid residues forming each protein and protein molecular weight are shown. From Barbas III, C. F.[190]

4.3. The phage life cycle

The phage life cycle is divided in various processes such as infection, replication and protein synthesis, and finally assembly.

4.3.1. Infection

Infection involves several steps requiring interactions with the F conjugative pilus and some bacterial cytoplasmic membrane Tol proteins (Q, R and A).[204]

Infection starts by the binding of the F pilus tip to the N2 domain of pIII phage protein. The subsequent steps involved in phage infection are unclear. Capside proteins are probably disassembled into the cytoplasmic membrane as the phage DNA is translocated into the cytoplasm for replication.[205, 206]

4.3.2. Replication and protein synthesis

Once the viral (+) strand DNA enters the cytoplasm, the complementary (-) strand is synthesized by bacterial enzymes, resulting in a covalently closed, double-stranded DNA, called replicative form DNA (RF). (-) strand of this RF is the template for transcription, and the resulting nRNAs are

translated into all of the phage proteins. pII and pX generated proteins mediate the synthesis of a new viral strand via a "rolling-cycle" mode of replication that uses bacterial enzymes to generate a pool of progeny double-stranded RF molecules to increase phage protein production. RF DNA synthesis continues until the amount of pV protein reaches a critical concentration in which dimerizes and binds to viral single-stranded DNA and prevent its conversion to RF DNA.

Only pII, pX and pV proteins, involved in DNA replication, are present in the host cytoplasm. All other phage proteins are inserted into the cytoplasmic or outer membrane while being synthesized.

4.3.3. Protein phage assembly

The assembly of Ff phage is a membrane-associated event. The substrate for the assembly process is pV-DNA complex. pV-DNA interactions are disrupted and capsid proteins are assembled around the DNA. The integrity of the membrane is maintained during this extrusion process allowing bacteria to grow, although in a slower rate.

Assembly process is divided into initiation, elongation and termination steps. Those details are far out of the scope of this introduction and thus not discussed here.[201, 207]

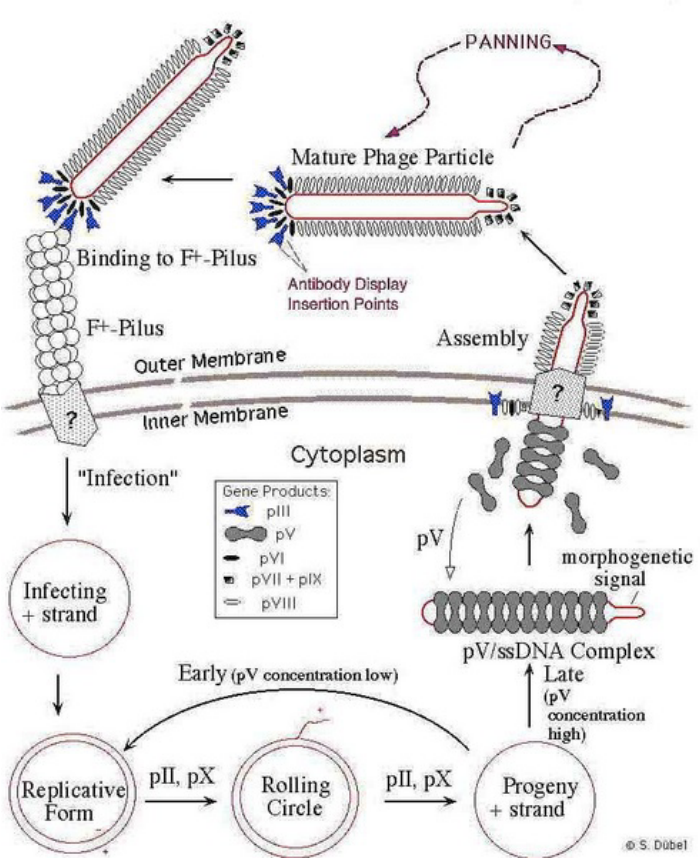


Figure 16. Phage life cycle scheme showing infection, replication and protein synthesis and assembly in detail. From S. Dübel's webpage.[208]

4.4. Display of peptides and proteins on phage particles

In 1985, J. Smith[209] reported how to fuse a foreign peptide or protein into the periplasmic portion of capsid proteins by phage genome engineering, eventually being packaged into a phage particle. Attempts to display peptides and antibodies to capsid proteins had been reported. From antibodies fusion into pVII and pIX[210] to peptides fused into pVI,[211] pIII or pVIII. However, by far the most commonly used virion proteins for phage display are pVIII and pIII.[212]

4.4.1. pVIII or major coat protein

pVIII protein is by far the most abundant protein within phage. Initially, only short peptide sequences (6-8 residues long) could be displayed on each copy of pVIII in a virion,[213-215] otherwise hampering packaging of the particles. However, display of larger peptides on pVIII has been described using phagemid vectors, resulting in hybrid virions with a varied range of pVIII engineered proteins and pVIII wild type proteins.[216-218] Most fused proteins are displayed on *N*-terminus, although *C*-terminus fused proteins had been described.[219] Peptides displayed on pVIII protein might be susceptible to cleavage by proteases as a result of their exposed position.[220]

4.4.2. pIII or minor coat protein

Five pIII protein copies are located in one edge of the virion. pIII is the protein of choice for most phage display fusions due to its tolerance for large insertions. However, since pIII protein plays an important role on infectivity, large insertions may dramatically reduce infectivity of phage virions. This can also be overcome by using hybrid constructs containing both protein fused pIII and wild-type pIII proteins.

While display constructs in pVIII enables to insert small peptides in huge quantities, pIII constructs allow larger inserts in only 5 positions (only 5 pIII proteins are present in each virion). All these possiblilities describing shorter or larger inserts and hybrids with variable amounts of displayed fused peptides in distinct capsid proteins allow a flexible playground to explore activity, affinity and specificity of displayed molecules.

4.5. Phage display vectors

A variety of vectors to introduce fused peptides within phage capsid proteins have been described. Vector selection relies on the coat protein used for display (usually pIII or pVIII), on whether peptides displayed are fused to all copies of pIII or pVIII or to only some of them and if recombinant fusion is encoded on the phage genome or on a phagemid.[221]

Phage vectors - Phage types

Phage display systems can be classified according to the arrangement of the coat protein genes.[212, 221, 222]

Type 3 vector displays the encoded foreign peptide in all 5 copies of pIII protein. There is a single chromosome (genome) bearing a single gene III. If foreign DNA is inserted to this single gene III, all expressed pIII proteins will include the fused inserted protein.

Similarly, Type 8 (and also, but more rarely Type 6) display foreign peptides on every single copy of pVIII (or pVI). Only short foreign peptides can be displayed on every copy of pVIII. Otherwise those foreign peptides can dramatically modify physical and biological properties and assembly problems can arise making the phage non-viable.[215, 223]

Type 88 vector, phage genome bears two genes VIII, encoding two distinct types of pVIII molecule: one recombinant bearing the foreign insert, and the other wild-type. The resulting phage particle is an hybrid containing both wild-type pVIII protein and recombinant one. This vector allows inserting larger fused proteins, in lesser amounts. Type 33 vector also contains two genes III, and thus could encode both wild-type and recombinant pIII protein.

Finally, Smith described the so-called vectors Type 8+8 and Type 3+3. Those type vectors differ from 88 and 33 in that the two genes VIII or genes III are on separate genomes. The wild-type version is on a phage (helper phage) while the recombinant version is on a phagemid.

For large phage library synthesis, most phage display vectors are introduced as naked DNA into *E. coli* cells by electroporation.[224]

4.6. Affinity selection

Biopanning is an affinity selection technique, which selects for peptides that bind to a given target.

4.6.1. *In vitro* affinity selection

Phage display was initially invented to allow the affinity selection of protein fragments encoded by a corresponding cDNA fragment by Smith (1985).[209] In 1988,[225] Parmley and Smith presented an improved selection procedure, which they called "biopanning".

Generally speaking, a biopanning protocol involves immobilization of ligand of interest on a solid support. Then, the phage display library is applied in solution and phage variants showing affinity bind to the immobilized ligand. Several washing rounds are performed to eliminate the adherent non-binders. Finally, what remains bound is eluted in such a way it do not spoil infectivity of virions. Then, those *positive* phage are amplified (by re-infecting bacteria and let them replicate), precipitated and titred prior to the next round of selection.

Affinity selection protocols can vary depending on data one seeks and the screening molecules and models. When choosing a panning experiment one should decide whether to maximize phage capture or affinity discrimination. This decision strongly influences the phage library design to better fit the biopanning protocol. Moreover, while *in vitro* affinity selection experiments can be tuned in a way to favour desired biopanning information, *in vivo* affinity selection models are not tuneable.

The highest yields of a binding phage are obtained by multivalent display coupled with a high density of screening molecule adsorbed to a solid surface. In contrast, discrimination between tight and weak-binding phage (affinity discrimination or stringency) is favoured by low densities of the target receptor and by monovalent display of the foreign peptide. This can be achieved both by solid-phase or in-solution affinity selection. Stringency leads inevitably to low yields.

In some cases, stringency requirements can be modified throughout panning rounds. Generally, in the first round of selection the goal is to capture as many positive variants as possible while removing non-specific binders. As enrichment increases over panning rounds, the stringency of selection should be increased with a variety of techniques to select the best high-affinity candidates.[226-229]

Usually, at least three panning rounds are required in order to amplify the binding variants and eliminate the background of non-specific binders. However, if several variants of binders are present in the library, the high-affinity variants will outcompete those with low affinity. This could be a problem in case we want to identify a landscape of numerous binding variants in a phage display library. Then, not many panning rounds should be applied in case all variants have to be present.[230]

4.6.2. *In vivo* affinity selection

Protocols for biopanning rounds *in vivo* slightly differ from those *in vitro*. Phage administration is performed by injecting anesthetized mice intravenously (right or left lateral tail vein), with phage peptide library in DMEM or PBS. Then, a wash out step to remove all unbound phage to reduce phage background in highly vascularized tissue[231] is performed by whole-body perfusion with PBS after mice are again deeply anesthetized. Then, phage are recovered by harvesting, weighting and grinding the selected organs or tissues. *E. coli* bacteria are added to tissue homogenate and

plated to determine phage sequences (by PCR sequencing). Finally, only selected phage colonies or all present colonies can be amplified, precipitated and tittered prior to next panning round.[232]

In general, three panning rounds of *in vivo* selection it is described to be sufficient to select for organ- and tumour-specific homing phage.

4.7. Phage library applications

Since G.P. Smith[209] described phage display technology in 1985, it has been extensively used for a wide variety of applications. Sometimes, creative modifications on standard protocols or constructs have been described to address specificities of each approach.[212]

First described applications were mainly related with affinity selection of polypeptides that specifically target a variety of receptors. The diversity of explored targets was and has been so broad, from conventional receptors like antibodies and hormone receptors to non-proteinaceous ligands like carbohydrates ligands or even plastic surfaces that do not have any natural ligand at all.[233]

Thus, first approaches screened and mapped epitopes; the small determinant on the surface of a ligand with which the receptor makes close, geometrically and chemically specific contact.[234] The technology allows searching for mimotopes. Mimotopes are small peptides that specifically bind a receptor's binding site (mimicking the epitope on the natural ligand) without matching the natural epitope at the amino acid sequence level.[98]

Since phage display technology is used to find ligands for a receptor, the same concept can be applied the other way around. Ligands can be used as a "probe" to identify new receptors binding them. As an example, eighteen SH3 domains were identified in such a way, nine of which were previously not known.[235, 236]

Most affinity selection targeted receptors are framed in drug discovery programs, and peptides selected are potential leads to new drugs. Those selected peptides might act as receptor agonist, antagonists or somehow modulated the receptor's biological effect. Phage display playground it has been extensively explored depending on specific drug discovery project requirements.

Phage display can serve as a tool for disease diagnosis and vaccine development when an antibody is used for affinity selection. Peptide ligands selected are called antigenic mimics of the corresponding natural epitope. Then, antigenic mimics can serve as a specific probes for antibodies that are diagnostic for a disease such as HIV. Therefore, antigenic mimics that are also immunogenic mimics can be useful as components of synthetic vaccines.[212, 237, 238]

The same affinity selection process used to explore protein-protein interactions have been reasoned to study DNA binding proteins.[239, 240]

Displaying peptides in pVIII phage proteins conforming the cylindrical-shaped capsid can be used as a platform to study three-dimensional conformations defined by its immediate neighbourhood. Three-dimensional conformations displayed on phage may strongly differ from the conformation displayed on wild-type particles aggregates, thus exploring other so-called "organic landscapes".[215, 223]

The observation that certain peptides preferentially adsorb to polymeric materials has driven some projects to screen random peptide libraries against a range of organic and inorganic materials. Thus, peptides with affinity to inorganic materials can be used to functionalize materials or trigger nucleation and stimulate growth of nanocrystals.[241]

Khoo *et al.* described a synthetic peptide construct based on tripeptide motif KHK found by panning phage display peptide libraries against a titanium alloy. This synthetic peptide contains repeats of KHK motif linked by diglycine linkers. When PEGylated, the peptide spontaneously

forms a hydrophilic film on the metal surface preventing nonspecific protein adsorption and bacterial colonization on orthopedic implantation devices.[242]

Phage library-derived peptides have also been described as peptide promoting cell adhesion to inorganic materials,[243] and also for the development of protein tools for controlled assembly of nanostructures from solutions.[244]

A. M. Belcher and co-workers described the use of genetically engineered filamentous phage as scaffolds for a variety of applications ranging from phage self-assembly into microelectrodes opening the door towards fabrication of high-performance nanostructured batteries,[245-247] efficient electron collection in photovoltaic devices,[248-250] *in vitro* tumour cell imaging[251] and *in vivo* imaging of prostate cancer,[252, 253] among others.

Since peptide-protein or protein-protein interactions are strongly influenced by conformation of both molecules, when studying interactions between targeted and ligand molecules, this parameter will be of great interest to find the best candidates. Thus, in some cases it can be beneficial to stabilize the target molecule to a particular conformation. Strategies using mutants or small molecule ligands had been described.[254, 255]

Contributions of phage display on drug delivery

Phage display technology has gained a prominent position on drug delivery research field. Identification of reliable disease markers may enable the design of targeted therapies as well as improve imaging or diagnosis tools. Characterization of molecular diversity at the disease cell surface level is the basis for the development of targeted therapies.[256, 257]

Phage display has an enormous potential to scan for peptides interacting with specific organs or tissues *in vivo*. *In vivo* phage display select for peptides displayed on phage particles that home to receptors differentially accessible in the vasculature of specific organs. *In vivo* phage display selection has two closely related main applications. On one hand, identification of ligands targeting specific vascular beds. On the other hand, by mapping molecular profiles of blood vessels in specific diseases allows identifying disease markers or variations on their expression on specific tissues.

Pioneer work by R. Pasqualini identified organ-selective targeting based on *in vivo* screening of random peptide sequences.[258] Described peptides showed up to 13-fold selectivity for brain and kidney blood vessels. Furthermore, synthetic peptide specifically inhibited localization of the homologous phage into the organs.

Phage display peptide libraries have been screened for ligands selectively homing to a variety of tissues in mice: brain and kidney,[258] lung, skin, pancreas, intestine, uterus, adrenal gland and retina,[231] lymph nodes,[259] muscle,[260] prostate,[261] breast,[262] placenta,[263] and white adipose tissue[264] among others. Animal models have also been used to identify tumour-vascular molecular addresses. A pancreatic cancer model in mice was used to monitor tumour progression from benign to metastatic state with phage displayed peptides.[265, 266] Other animals models had been developed other than mice such as rat or insects.

Limitations on applying selected candidates by direct extrapolation between most accessible animal models and humans have been reported due to differences in expression of many vascular targets.[261] *In vivo* phage display data derived from animal models must be carefully validated before performing clinical studies with human beings.

Arap & Pasqualini and co-workers reasoned *in vivo* random phage display selection in human patients to start monitoring peptides with significant affinities to the human vasculature.[267] After injecting a random phage display peptide CX_7C library to a cancer patient, they analysed peptide distribution in bone marrow, prostate, white fat, skeletal muscle and skin. They performed a high-

throughput analysis on peptide motifs recovered from those organs. They first conclude that among 47,160 screened tripeptide motifs, distribution of circulating combinatorial peptides in human tissues was non-random. Furthermore, they were able to identify 25 organ-homing tripeptide motifs. From those, eleven were found to be enriched in a single organ while the others were enriched in multiple organs. This work represent a start point towards mapping the human vasculature and established a frame for systematic evaluation of vascular ligand-receptor interactions.

To identify candidate protein-ligands mimicked by found peptides they searched on databases. High-throughput analysis revealed some cell surface proteins to be validated as potential receptors for selected sequences. As a proof of concept, Cardó-Vila *et al*.[268] studied the *in vivo*-homing motif CGRRAGGSC. Despite showing certain ligand-specific attributes mimicking interleukin 11 (IL-11), selected sequence was not within previously characterized interacting sites of IL-11Rα. Expression of IL-11Rα was reported to be increased in prostate cancer and in associated vasculature, specifically on prostate cancer metastasis being a candidate target in the progression of human prostate cancer.[269] By in-tandem binding assays, site-directed mutagenesis and NMR spectroscopy the authors shown that peptide mimics a receptor-binding site within IL-11 and that the binding of the selected peptide to IL-11Rα is functional inducing cell proliferation through STAT3 activation (signal transduction and activation transcription).

Since a single selection round might not be sufficient for homing-peptides enrichment in-patient, a pseudo-panning approach was envisaged. Synchronous serial *in vivo* selection was used to identify and validate tissue-specific molecular addresses in cancer patients.[270] This methodology consists on re-injecting the amplified recovered phage from the organ of interest in three-to-four rounds, utilizing one subject per round. This protocol has been limited by the need to perform individual multi-round screens for motifs homing to each tissue studied.

Ethical concerns in end-of-life cancer patient research had been addressed and discussed elsewhere.[271-273]

4.8. Phage library synthesis considerations

4.8.1. Oligonucleotide libraries

The success of phage display screening basically relies on the quality of the phage library, in terms of diversity. Since the inserted oligonucleotide library encodes displayed peptides, phage display library variability strongly depends on the quality of the initial DNA-library inserted to the phage genome.

Phage display libraries can range a variety of displayed shapes. Petrenko and Smith summarize some of described libraries displayed on phage.[212] 20^N is the number of different peptide sequences that can be obtained by randomizing the 20 natural amino acids in *N* residue positions.

A simple correlation between the number of random amino acids and the theoretical diversity of such peptide library is shown in Table 3. Experimentally, there are two main factors that limit this maximum number of obtained phage clones. On one hand, the amount of phage-encoding DNA molecules that can be generated *in vitro*. On the other hand, the efficiency of their introduction by electroporation into *E. coli* cells (usually, only 1% of DNA is efficiently transferred into bacterial cells).

Random positions	Total sequences
7	$1.28 \cdot 10^{9}$
8	$2.56 \cdot 10^{10}$
9	$5.12 \cdot 10^{11}$
10	$1.02 \cdot 10^{13}$
11	$2.05 \cdot 10^{14}$
12	$4.10 \cdot 10^{15}$

Table 3. Total number of sequences within phage libraries comprising 20 proteinogenic amino acids depending on randomized position.

By this means, it can be evaluated by the starting amount of DNA vector and the random residues of the targeted library how many clones can be covered. Lindener *et al*.[274] estimate that about 10 µg of an engineered vector are required to obtain a library with 10^{9} individuals clones, each present in 1000 copies. With this amount of DNA vector only 7-mer peptide libraries cover all clones, while only 1% or 0.001% of the sequence space is covered on random 9-mer peptide or 12-mer libraries, respectively.

4.8.2. Synthesis of library inserts

Although oligonucleotide library used for phage display library synthesis was commercially available, here we discuss some general approaches usually used on oligonucleotide library chemical synthesis.

Solid-phase synthesis[275] is the method of choice for the vast majority of approaches since it allows to meet most requirements to be used as peptide-encoding libraries such as defined chain length, correct sequences vicinal to the random section, appropriate placement of primer and restriction sites and high purity. Automated solid-phase DNA synthesis with phosphoramidite building blocks achieves coupling yields up to 99.5% and synthesis of 100 bases long oligonucleotides are feasible.

Equimolar mixtures of the four activated nucleotides are usually used to randomize oligonucleotide sequence. However, since 3 out of 64 trinucleotide codons are stop codons, not all three positions of codons are fully randomized. Instead, a synthesis pattern such as *NNK* or *NNF* is used, where *N* corresponds to any oligonucleotide (thymine, adenine, guanine or cytosine), while *K* corresponds to only guanine and thymine, and *F* to guanine and cytosine. This approach lead to the elimination of 2 of 3 stop codons by representing only 32 triplet codes that keeps the same amino acid proportion of the full 64 codon table. The remaining stop codon (TAG) can be suppressed by a *sup*E *E. coli* strain used for phage propagation.[276]

TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys
TTC		TCC		TAC		TGC	
TTA	Leu	TCA		TAA	**STOP**	TGA	**STOP**
TTG		TCG		TAG	**STOP**	TGG	Trp
CTT		CCT	Pro	CAT	His	CGT	Arg
CTG		CCC		CAC		CGC	
CTA		CCA		CAA	Gln	CGA	
CTG		CCG		CAG		CGG	
ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser
ATG		ACG		AAC		AGC	
ATA		ACA		AAA	Lys	AGA	Arg
ATG	Met	ACG		AAG		AGG	
GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly
GTC		GCC		GAC		GGC	
GTA		GCA		GAA	Glu	GGA	
GTG		GCG		GAG		GGG	

Table 4. Codons present in a *NNF* oligonucleotide library (highlighted in green), where *N* corresponds to any oligonucleotide (thymine, adenine, guanine or cytosine), while *F* to guanine and cytosine.

Other imaginative approaches had been described to finely tune the phage library design by randomizing with a limited number of amino acids some positions within the peptide displayed by phage.[277, 278] Thus, a limited variety of amino acids can be achieved by restricting oligonucleotides presence in specific codon positions as shown below.

Codon (X-X-X)	Amino acids
G/T-A/C-T	Tyr, Ala, Asp, Ser
A/T-A/C-T	Tyr, Thr, Asn, Ser
C/T-A/C-T	Tyr, Pro, His, Ser
A/G-G/C-A	Thr, Arg, Gly, Ala
G/C-A/G-C	His, Arg, Asp, Gly
A/G-G/C-T	Gly, Ala, Thr, Ser

Table 5. Selection of codon sets which limit randomization to four amino acids.

Alternatively, the trinucleotide synthetic approach assembles trinucleotide building blocks previously synthesized in solution. Since oligonucleotide phosphoramidites show large differences in reactivity, non-equimolar but *equireactive* mixtures of trinucleotides or mix and split method are required.[279-282] This methodology exhibits significant advantages: first, it allows to do not include stop codons. It allows choosing one single codon for each amino acid leading to strict and proportional amino acid presence and easily allows to tune the distribution of amino acids since each codon is regarded as a building block.[279, 283]

Codon	Amino acid	Codon	Amino acid
Ala	GCT	Leu	CTG
Arg	CGT	Lys	AAA
Asn	AAC	Met	ATG
Asp	GAC	Phe	TTC
Cys	TGC	Pro	CCG
Gln	CAG	Ser	TCT
Glu	GAA	Thr	ACT
Gly	GGT	Trp	TGG
His	CAT	Tyr	TAC
Ile	ATC	Val	GTT

Table 6. List of codons recommended for the trinucleotide approach by Kayushin *et al*.[279]

4.8.3. Variability of phage peptide libraries

After electroporation and growth, produced viral material is analysed by DNA sequencing. Although look for a picture of the whole family would be of great interest to validate the quality of the library this cannot be achieved by phage sequencing of the complete library. Indeed, this would be extremely expensive and time consuming and would waste the library. Several approaches have been described to assess the quality of the library. Most of them consist on sequencing a small, but significant, amount of phage within the library to evaluate this reduced library sample. Peptide sequence repetitions and presence of amino acids in each residue position can be checked to monitor library quality. Some tools as RELIC database have been designed to improve statistical analysis of this process.[284-287]

Continued iterative reinfection of *E. coli* with phage library it is not an option to obtain large amplifications of the library. Library diversity is compromised generation over generation due to non-linear propagation rates of individual clones (peptides are displayed on phage pIII protein, which is responsible of phage infectivity). Thus, in only two library-amplification steps exceptional propagation rate clones (i.e. 1.5 propagation factor rate) would seriously bias the library diversity.

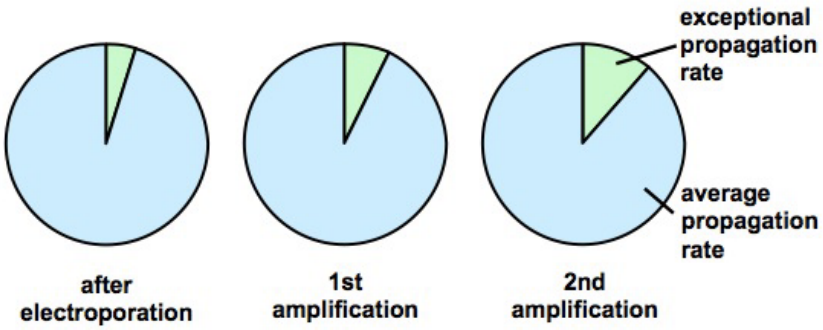


Figure 17. Schematic and approximate illustration on variability bias of phage library after continuous library amplification. Green portion corresponds to phage with exceptional propagation rate whereas blue portion corresponds to phage with standard propagation rate.

This intrinsic library feature, leads to make some considerations on screening affinity procedures like biopanning. Figure 18 shows in a qualitatively manner how affinity selection protocols can select best binding candidates from the library.

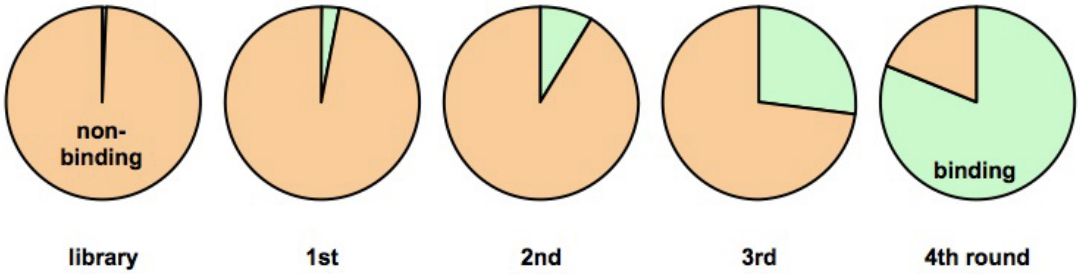


Figura 18. Qualitative illustration of enrichment of phage throughout affinity selection protocol (biopanning).

Since biopanning involves reinfection of *E. coli* by positive phage of each panning round, amplification of this selected phage can lead to slightly bias panned libraries with slight differences between positive clone library before and after amplification. Because biopanning could involve several rounds, these slight differences could become a dramatic bias on candidate selection.

Another concept must be considered to design this biopanning approach. Several clones present in a library can bind the target by unspecific interactions. Moreover, they will possess diverse propagation fitness.

In a binding assay, phage can be grouped as non-binding, unspecific binding and specific binding.

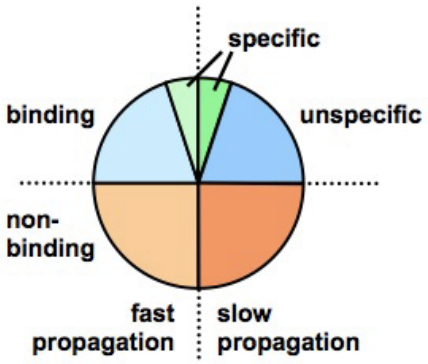


Figure 19. Figure showing a general schematic representation of phage behaviour in a biopanning assay in terms of phage propagation rate and binding nature. On one hand, phage can show fast or slow propagation rates. On the other hand, most phage do not bind to the tested target, others bind to the target unspecifically while others show specific binding.

Then, specific binders that are present in low number of copies having slow propagating rate may be in serious trouble to be selected. On the other hand, unspecific binders with high propagation competence can outgrow the specific binders in the following amplification step. Stringent competition among clones may not be sufficient to remove the unspecific clones.

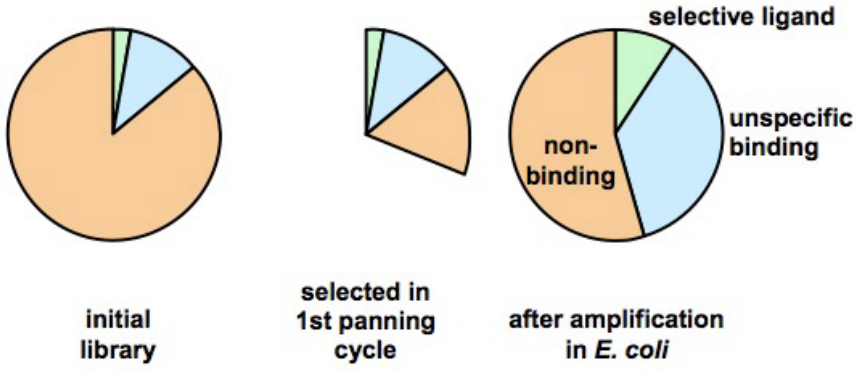


Figura 20. Schematic representation of how phage propagation rates can affect biopanning results. After one panning round (represented) unspecific binders with high propagation rates will compete specific binders with moderate propagation rates.

All this considerations must be taken into account on biopanning protocol designs.

4.9. Expanding screened chemical space by phage display

As mentioned above, several phage library constructs and scaffolds have been developed and reviewed.[212] Phage display generating tools allow a wide variety of libraries mimicking specific features of proteins by tailoring a variety of scaffolds such as disulphide bridges with cysteine constraining randomized peptide loops.[222] Disulfide bridges are one of the most common used methods to introduce peptide constrains to mimic natural compounds.[267, 288-291]

Another library design can lead to alpha helices and beta sheets by coding lipophilic and hydrophilic amino acids at specific positions of displayed peptides. Furthermore, by choosing codon sets one can establish acidic or basic positions within the displayed peptides.[292]

Heinis *et al*. described an approach that involves chemical modification of displayed peptides (Figure 21). They produced a bicyclic phage display library after chemical treatment of a phage library containing three cysteine residues at specific positions. They used 1,3,5-tris(bromomethylene)benzene (TBMB) to interconnect all cysteines and produce a phage library of bicyclic peptides.[293]

Chemical space screened by phage display can also be expanded as described by Meyer *et al*.,[294] looking for selective bivalent inhibitors of Protein Kinase A (Figure 22). They constructed a phage library displaying at *N*-terminus randomized hexapeptide flanked by cysteines forming a disulphide bond. Those displayed peptide library were linked by a displayed Fos domain to pIII phage protein. A synthetic Jun domain with a conjugated small molecule was prepared and added to the phage library. The synthesized small molecule was tethered to the phage library via non-covalent self-assembly of Jun and Fos domains. This construct guide displayed peptides to the proximity of enzyme's active site thus exploring bivalent inhibitor properties for increased affinity and enhanced selectivity for the targeted kinase. The bivalent construct can be eventually synthesized by covalent binding of both moieties.

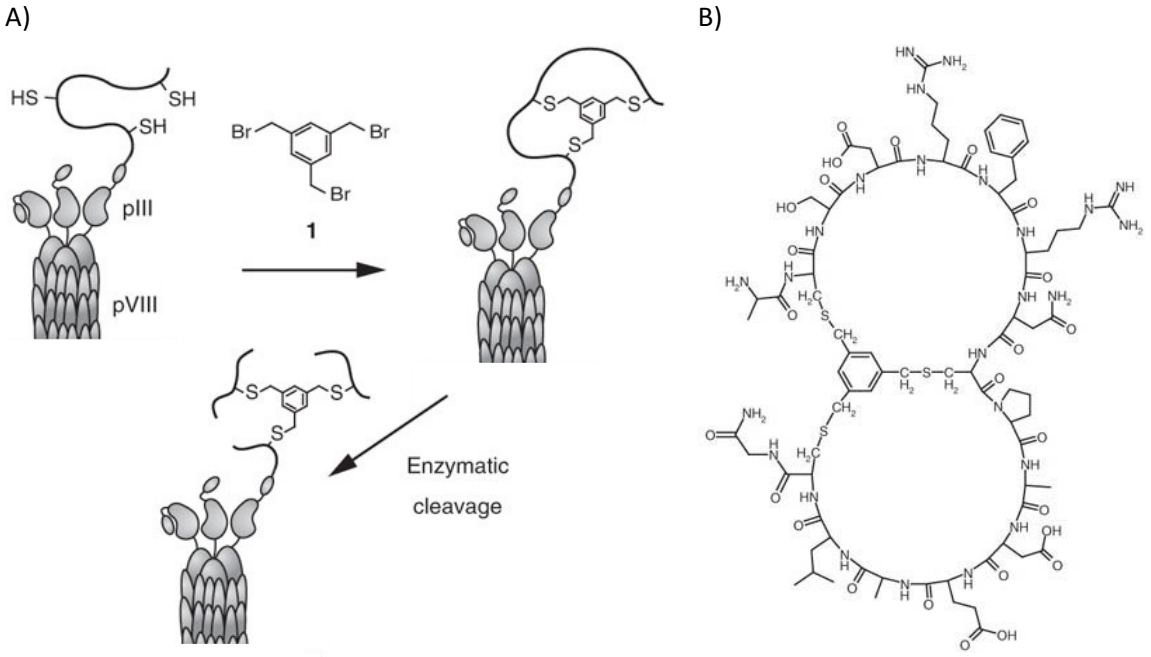


Figure 21. A) A phage-encoded peptide library containing three cysteine residues can be tethered to the trifunctional compound TBMB in a nucleophilic substitution reaction resulting in byciclic peptides displayed on phage. Authors suggest those peptides could be further modified by enzymatic reactions. B) Chemical structure of a macrocyclic plasma kallikrein inhibitor isolated with this library approach. Figure taken from Heinis, C. *et al*.[293]

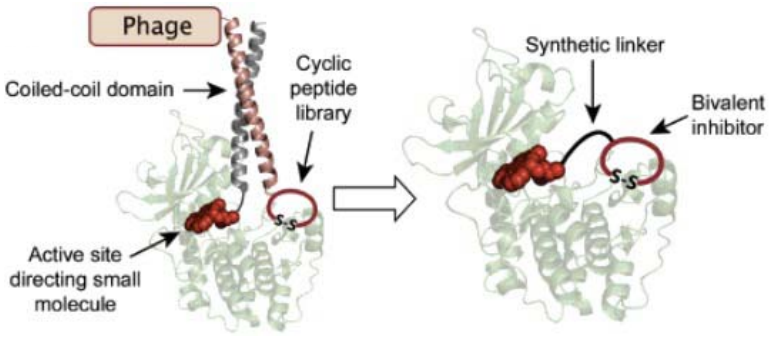


Figure 22. Bivalent inhibitor selection strategy described by Meyer. A non-covalent tethering of an active site directing small molecule (ligand) and a phage library by a coiled-coil heterodimer self-assembly for targeting PKA. Selected peptide can be subsequently conjugated to the small molecule providing a bivalent kinase inhibitor. Figure from Meyer, S.C. *et al*.[294]

Cwirla and co-workers envisaged a more complex library design.[295] They described a technology for attaching libraries of synthetic compounds to coat proteins of bacteriophage particles. They arrayed and grown in wells of microtiter plates the phage library. Each phage within the library uniquely encoded a specific insertion of a DNA tag in a non-expressed region of the genome. On the other hand, they prepare a library of synthetic compounds, and place each of them in one pool of the plate. Thereafter, they perform hydrazone attachment chemistry to attach peptides on phage (previously incubated with succinimidyl 4-formylbenzoate to introduce an aldehyde at the exposed primary amines present in the phage coat proteins). They constructed and tested a library of 980 analogue compounds of folic acid attached to T7 phage against KB carcinoma cells overexpressing folate receptors.

Dwyer *et al*. developed a method called biosynthetic phage display.[296] They displayed one part of the protein on pIII or pVIII of phage (DNA-encoded) while the other part was chemically synthesized and linked by native chemical ligation. This elegant technology allows randomizing some residues of displayed DNA-encoded protein whereas one could introduce non-natural

amino acids on the synthetic protein moiety. This methodology opens a new scope on the field of protein mutagenesis.

Since phage display as described by Smith only can display peptides containing proteinogenic amino acids, some efforts are been devoted to introduce non-natural amino acids to that technology to explore wider chemical space and produce protease resistant peptides. This work pioneered by Schultz relies on the presence in *E. coli* of an additional orthogonal tRNA/aminoacyl-tRNA synthase pair capable of incorporation various unnatural amino acids into proteins in response to unique nonsense codons.[297, 298] Several non-natural amino acid had been successfully inserted in phage display libraries.[299]

Other strategies to achieve D-peptide candidates by phage display consist on screen/test a regular phage display library against an enantiomer of the screened ligand synthesized *in vitro*.[300, 301] Thus, at the end of the selection process called *mirror image phage diplay*, the enantiomer of L-peptide displayed by the selected candidate can be synthesized since ligand screened will be found in L-enantiomer version in nature. L-peptide – D-ligand interaction is equivalent to the D-peptide – L-ligand interaction. This methodology is strictly linked to *in vitro* assay formats. Schumacher *et al*. successfully panned a phage library expressing L-amino acids peptides against a synthetic SH3 domain of c-Src entirely composed by D-amino acids.[302]

5. Outlook

From its birth, both combinatorial chemistry and high-throughput screening evolved trying to overcome very diverse difficulties found in the drug discovery process. Once synthesis of high-quality libraries comprising large amount of compounds was addressed, the bottleneck on drug discovery by use of combinatorial and high-throughput screening techniques shifted towards the identification of specific lead compounds.

Advances on technical engineering and robotics as well as in software for analysis, raised and perfected enormously HTS performance. Old HTS standards valued the use of automation to only increase throughput, while the effort now also focus to faithfully execute complex tasks with high precision. Modern HTS facilities employ automated screening systems composed of state-of-the-art liquid handling, plate readers, incubators and other instruments to support a wide variety of assays.[83] Considerable efforts to miniaturize all processes to significantly decrease experimental costs have been adopted by the so-called ultrahigh-throughput screening (UHTS) techniques, which are able to consistently work with nano-liter volumes.[303-305]

Interestingly, as was already discussed, phage display in particular and display technologies in general, presume from a very efficient and easy identification of lead compounds. Since displayed peptides are the phenotypic expression of a genotype, its identification is straightforward with amplification (via PCR) and by novel HTS sequencing techniques.[186] Despite remarkable improvements in last years to modify chemical space displayed, display technologies suffer from limited chemical space to be explored.

The goal to create large libraries of chemical compounds in a single pot, inevitably involves important limitations in the concentration of individual library members (down to nanomolar range) thus requiring extremely sensitive detection or coding methods.

In our opinion, two distinct approaches for chemically synthesized libraries are envisaged at that point to step further for improved tools.

First, DNA-encoding chemical libraries (both based on peptides or small molecules) are regarded as filling the gap between classical combinatorial chemistry and display technologies. As discussed above huge (up to 10^9-10^{12} members) DNA-encoded libraries can be constructed while identification, once evaluated, of such few amounts of every individual is thanks to ultra-sensitive DNA detection methodologies such as PCR.[306]

On the other hand, advances on mass spectrometry techniques launched, last decades, proteomics field into an unforeseen scene and could be applied for the identification of very complex mixtures from combinatorial chemistry.[161, 172]

The core of the present thesis is devoted to explore this last scenario by applying state-of-the-art mass spectrometry techniques to the screening of a mix and split peptide library in an *in vitro* cell-based BBB model.

Objectives

The main goal of the present thesis is to search for BBB-shuttles, peptides with the capacity to enter the brain, by using peptidic combinatorial libraries and mass spectrometry as research tools. To achieve this general aim, we defined the following objectives framing distinct approaches explored in the present thesis.

1. Design, synthesis and evaluation of an stereoisomeric peptide library based on Ac-(*N*-MePhe)$_4$-NH_2 to study chirality effect on transport passage across an *in vitro* model of the BBB only mimicking passive diffusion properties.

2. Design, synthesis and evaluation of a small peptide library −based on a phage display lead compound− to enhance stability, solubility and transport features across an *in vitro* cell-based BBB model.

3. Assessment of MS techniques and methodologies for transport quantification in *in vitro* assays.

4. Design, synthesis and *in vivo* evaluation of a peptide phage display library as a BBB-shuttles source, seeking for minimal peptidic motifs driving specific targeting interactions.

5. Design, synthesis and evaluation of a mix and split peptide library, *in vitro* and *in vivo*, as a source of new BBB-shuttles. Set up of a high-throughput screening tool for identification and quantification of lead compounds by mass spectrometry.

Bibliography

1. Gustavsson, A., Svensson, M., Jacobi, F., Allgulander, C., Alonso, J., Beghi, E., Dodel, R., Ekman, M., Faravelli, C., Fratiglioni, L., Gannon, B., Jones, D.H., Jennum, P., Jordanova, A., Jonsson, L., Karampampa, K., Knapp, M., Kobelt, G., Kurth, T., Lieb, R., Linde, M., Ljungcrantz, C., Maercker, A., Melin, B., Moscarelli, M., Musayev, A., Norwood, F., Preisig, M., Pugliatti, M., Rehm, J., Salvador-Carulla, L., Schlehofer, B., Simon, R., Steinhausen, H.C., Stovner, L.J., Vallat, J.M., Van den Bergh, P., van Os, J., Vos, P., Xu, W., Wittchen, H.U., Jonsson, B. & Olesen, J. Cost of disorders of the brain in Europe 2010. *Eur. Neuropsychopharm.* **2011**, 21, 718-779.
2. Olesen, J., Gustavsson, A., Svensson, M., Wittchen, H.U. & Jonsson, B. The economic cost of brain disorders in Europe. *Eur. J. Neurol.* **2012**, 19, 155-162.
3. Web site: http://www.tv3.cat/marato/totes_les_maratons., 1st December 2013.
4. Henderson, C. & Thornicroft, G. Stigma and discrimination in mental illness: Time to Change. *Lancet* **2009**, 373, 1928-1930.
5. Pardridge, W.M. The blood-brain barrier: bottleneck in brain drug development. *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics* **2005**, 3-14.
6. Pardridge, W.M. Brain Drug Targeting: The Future of Brain Drug Development. Cambridge University Press, Cambridge, UK; **2001**. pp. 1-12.
7. Abbott, N.J. Blood-brain barrier structure and function and the challenges for CNS drug delivery. *J. Inherit. Metab. Dis.* **2013**, 36, 437-449.
8. Abbott, N.J. Physiology and Pharmacology of the Blood-Brain Barrier. Springer, Heidelberg; **1992**. pp. 371-396.
9. Liu, W.Y., Wang, Z.B., Zhang, L.C., Wei, X. & Li, L. Tight junction in blood-brain barrier: an overview of structure, regulation, and regulator substances. *CNS Neurosci. Ther.* **2012**, 18, 609-615.
10. Abbott, N.J., Patabendige, A.A., Dolman, D.E., Yusof, S.R. & Begley, D.J. Structure and function of the blood-brain barrier. *Neurobiol. Dis.* **2010**, 37, 13-25.
11. de Boer, A.G. & Gaillard, P.J. Drug targeting to the brain. *Annu. Rev. Pharmacol. Toxicol.* **2007**, 47, 323-355.
12. Dauchy, S., Dutheil, F., Weaver, R.J., Chassoux, F., Daumas-Duport, C., Couraud, P.O., Scherrmann, J.M., De Waziers, I. & Decleves, X. ABC transporters, cytochromes P450 and their main transcription factors: expression at the human blood-brain barrier. *J. Neurochem.* **2008**, 107, 1518-1528.
13. Dutheil, F., Jacob, A., Dauchy, S., Beaune, P., Scherrmann, J.M., Decleves, X. & Loriot, M.A. ABC transporters and cytochromes P450 in the human central nervous system: influence on brain pharmacokinetics and contribution to neurodegenerative disorders. *Expert Opin. Drug Metab. Toxicol.* **2010**, 6, 1161-1174.
14. Ghosh, C., Puvenna, V., Gonzalez-Martinez, J., Janigro, D. & Marchi, N. Blood-brain barrier P450 enzymes and multidrug transporters in drug resistance: a synergistic role in neurological diseases. *Curr. Drug. Metab.* **2011**, 12, 742-749.
15. Demeule, M., Regina, A., Jodoin, J., Laplante, A., Dagenais, C., Berthelet, F., Moghrabi, A. & Beliveau, R. Drug transport to the brain: key roles for the efflux pump P-glycoprotein in the blood-brain barrier. *Vascul. Pharmacol.* **2002**, 38, 339-348.
16. Loscher, W. & Potschka, H. Blood-brain barrier active efflux transporters: ATP-binding cassette gene family. *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics* **2005**, 2, 86-98.
17. Abbott, N.J., Ronnback, L. & Hansson, E. Astrocyte-endothelial interactions at the blood-brain barrier. *Nat. Rev. Neurosci.* **2006**, 7, 41-53.
18. Krewson, C.E., Klarman, M.L. & Saltzman, W.M. Distribution of nerve growth factor following direct delivery to brain interstitium. *Brain Res.* **1995**, 680, 196-206.
19. Ai, Y., Markesbery, W., Zhang, Z., Grondin, R., Elseberry, D., Gerhardt, G.A. & Gash, D.M. Intraputamenal infusion of GDNF in aged rhesus monkeys: distribution and dopaminergic effects. *J. Comp. Neurol.* **2003**, 461, 250-261.
20. Zunkeler, B., Carson, R.E., Olson, J., Blasberg, R.G., DeVroom, H., Lutz, R.J., Saris, S.C., Wright, D.C., Kammerer, W., Patronas, N.J., Dedrick, R.L., Herscovitch, P. & Oldfield, E.H. Quantification and pharmacokinetics of blood-brain barrier disruption in humans. *J. Neurosurg.* **1996**, 85, 1056-1065.
21. Pardridge, W.M. Molecular Trojan horses for blood-brain barrier drug delivery. *Curr. Opin. Pharmacol.* **2006**, 6, 494-500.

22. Deeken, J.F. & Loscher, W. The blood-brain barrier and cancer: transporters, treatment, and Trojan horses. *Clin. Cancer Res.* **2007**, 13, 1663-1674.
23. Hu, K., Li, J., Shen, Y., Lu, W., Gao, X., Zhang, Q. & Jiang, X. Lactoferrin-conjugated PEG-PLA nanoparticles with improved brain delivery: in vitro and in vivo evaluations. *J. Control. Release* **2009**, 134, 55-61.
24. Temsamani, J., Bonnafous, C., Rousselle, C., Fraisse, Y., Clair, P., Granier, L.A., Rees, A.R., Kaczorek, M. & Scherrmann, J.M. Improved brain uptake and pharmacological activity profile of morphine-6-glucuronide using a peptide vector-mediated strategy. *J. Pharm. Exp. Ther.* **2005**, 313, 712-719.
25. Wu, D., Song, B.W., Vinters, H.V. & Pardridge, W.M. Pharmacokinetics and brain uptake of biotinylated basic fibroblast growth factor conjugated to a blood-brain barrier drug delivery system. *J. Drug Target.* **2002**, 10, 239-245.
26. Kumar, P., Wu, H., McBride, J.L., Jung, K.E., Kim, M.H., Davidson, B.L., Lee, S.K., Shankar, P. & Manjunath, N. Transvascular delivery of small interfering RNA to the central nervous system. *Nature* **2007**, 448, 39-43.
27. Teixido, M., Zurita, E., Malakoutikhah, M., Tarrago, T. & Giralt, E. Diketopiperazines as a tool for the study of transport across the blood-brain barrier (BBB) and their potential use as BBB-shuttles. *J. Am. Chem. Soc.* **2007**, 129, 11802-11813.
28. Malakoutikhah, M., Teixido, M. & Giralt, E. Toward an optimal blood-brain barrier shuttle by synthesis and evaluation of peptide libraries. *J. Med. Chem.* **2008**, 51, 4881-4889.
29. Malakoutikhah, M., Prades, R., Teixido, M. & Giralt, E. N-methyl phenylalanine-rich peptides as highly versatile blood-brain barrier shuttles. *J. Med. Chem.* **2010**, 53, 2354-2363.
30. Cecchelli, R., Berezowski, V., Lundquist, S., Culot, M., Renftel, M., Dehouck, M.P. & Fenart, L. Modelling of the blood-brain barrier in drug discovery and development. *Nat. Rev. Drug Discov.* **2007**, 6, 650-661.
31. Shawahna, R., Decleves, X. & Scherrmann, J.M. Hurdles with using in vitro models to predict human blood-brain barrier drug permeability: a special focus on transporters and metabolizing enzymes. *Curr. Drug. Metab.* **2013**, 14, 120-136.
32. Joo, F. The blood-brain barrier in vitro: Ten years of research on microvessels isolated from the brain. *Neurochem. Int.* **1985**, 7, 1-25.
33. Joo, F. The cerebral microvessels in culture, an update. *J. Neurochem.* **1992**, 58, 1-17.
34. Demeuse, P., Kerkhofs, A., Struys-Ponsar, C., Knoops, B., Remacle, C. & van den Bosch de Aguilar, P. Compartmentalized coculture of rat brain endothelial cells and astrocytes: a syngenic model to study the blood-brain barrier. *J. Neurosci. Meth.* **2002**, 121, 21-31.
35. Perriere, N., Yousif, S., Cazaubon, S., Chaverot, N., Bourasset, F., Cisternino, S., Decleves, X., Hori, S., Terasaki, T., Deli, M., Scherrmann, J.M., Temsamani, J., Roux, F. & Couraud, P.O. A functional in vitro model of rat blood-brain barrier for molecular analysis of efflux transporters. *Brain Res.* **2007**, 1150, 1-13.
36. Coisne, C., Dehouck, L., Faveeuw, C., Delplace, Y., Landry, C., Morissette, C., Fenart, L., Cecchelli, R., Tremblay, P. & Dehouck, B. Mouse syngenic in vitro blood-brain barrier model: a new tool to examine inflammatory events in cerebral endothelium. *Lab. Invest.* **2005**, 85, 734-746.
37. Meresse, S., Dehouck, M.P., Delorme, P., Bensaid, M., Tauber, J.P., Delbart, C., Fruchart, J.C. & Cecchelli, R. Bovine brain endothelial cells express tight junctions and monoamine oxidase activity in long-term culture. *J. Neurochem.* **1989**, 53, 1363-1371.
38. Dehouck, M.P., Meresse, S., Delorme, P., Fruchart, J.C. & Cecchelli, R. An easier, reproducible, and mass-production method to study the blood-brain barrier in vitro. *J. Neurochem.* **1990**, 54, 1798-1801.
39. Cecchelli, R., Dehouck, B., Descamps, L., Fenart, L., Buee-Scherrer, V.V., Duhem, C., Lundquist, S., Rentfel, M., Torpier, G. & Dehouck, M.P. In vitro model for evaluating drug transport across the blood-brain barrier. *Adv. Drug Deliv. Rev.* **1999**, 36, 165-178.
40. Yousif, S., Marie-Claire, C., Roux, F., Scherrmann, J.M. & Decleves, X. Expression of drug transporters at the blood-brain barrier using an optimized isolated rat brain microvessel strategy. *Brain Res.* **2007**, 1134, 1-11.
41. Dehouck, B., Dehouck, M.P., Fruchart, J.C. & Cecchelli, R. Upregulation of the low density lipoprotein receptor at the blood-brain barrier: intercommunications between brain capillary endothelial cells and astrocytes. *J. Cell Biol.* **1994**, 126, 465-473.
42. Wolburg, H., Neuhaus, J., Kniesel, U., Krauss, B., Schmid, E.M., Ocalan, M., Farrell, C. & Risau, W. Modulation of tight junction structure in blood-brain barrier endothelial cells. Effects of tissue culture, second messengers and cocultured astrocytes. *J. Cell Sci.* **1994**, 107, 1347-1357.

43. Rist, R.J., Romero, I.A., Chan, M.W., Couraud, P.O., Roux, F. & Abbott, N.J. F-actin cytoskeleton and sucrose permeability of immortalised rat brain microvascular endothelial cell monolayers: effects of cyclic AMP and astrocytic factors. *Brain Res.* **1997**, 768, 10-18.
44. Gaillard, P.J. & de Boer, A.G. 2B-Trans technology: targeted drug delivery across the blood-brain barrier. *Methods Mol. Biol.* **2008**, 437, 161-175.
45. Rubin, L.L., Hall, D.E., Porter, S., Barbu, K., Cannon, C., Horner, H.C., Janatpour, M., Liaw, C.W., Manning, K., Morales, J., Tanner, L.I., Tomaselli, K.J. & Bard, F. A cell culture model of the blood-brain barrier. *J. Cell Biol.* **1991**, 115, 1725-1735.
46. O'Donnell, M.E., Martinez, A. & Sun, D. Cerebral microvascular endothelial cell Na-K-Cl cotransport: regulation by astrocyte-conditioned medium. *Am. J. Physiol.* **1995**, 268, C747-754.
47. Vandenhaute, E., Sevin, E., Hallier-Vanuxeem, D., Dehouck, M.P. & Cecchelli, R. Case study: adapting in vitro blood-brain barrier models for use in early-stage drug discovery. *Drug discovery today* **2012**, 17, 285-290.
48. Deli, M.A., Dehouck, M.P., Abraham, C.S., Cecchelli, R. & Joo, F. Penetration of small molecular weight substances through cultured bovine brain capillary endothelial cell monolayers: the early effects of cyclic adenosine 3',5'-monophosphate. *Exp. Physiol.* **1995**, 80, 675-678.
49. Rubin, L.L. & Staddon, J.M. The cell biology of the blood-brain barrier. *Annu. Rev. Neurosci.* **1999**, 22, 11-28.
50. Sallee, J.L. & Burridge, K. Density-enhanced phosphatase 1 regulates phosphorylation of tight junction proteins and enhances barrier function of epithelial cells. *J. Biol. Chem.* **2009**, 284, 14997-15006.
51. Hoheisel, D., Nitz, T., Franke, H., Wegener, J., Hakvoort, A., Tilling, T. & Galla, H.J. Hydrocortisone reinforces the blood-brain barrier properties in a serum free cell culture system. *Biochem. Biophys. Res. Commun.* **1998**, 244, 312-316.
52. Franke, H., Galla, H.J. & Beuckmann, C.T. An improved low-permeability in vitro-model of the blood-brain barrier: transport studies on retinoids, sucrose, haloperidol, caffeine and mannitol. *Brain Res.* **1999**, 818, 65-71.
53. Zhang, Y., Li, C.S., Ye, Y., Johnson, K., Poe, J., Johnson, S., Bobrowski, W., Garrido, R. & Madhu, C. Porcine brain microvessel endothelial cells as an in vitro model to predict in vivo blood-brain barrier permeability. *Drug Metab. Dispos.* **2006**, 34, 1935-1943.
54. Roux, F. & Couraud, P.O. Rat brain endothelial cell lines for the study of blood-brain barrier permeability and transport functions. *Cell. Mol. Neurobiol.* **2005**, 25, 41-58.
55. Reichel, A., Abbott, N.J. & Begley, D.J. Evaluation of the RBE4 cell line to explore carrier-mediated drug delivery to the CNS via the L-system amino acid transporter at the blood-brain barrier. *J. Drug Target.* **2002**, 10, 277-283.
56. Weksler, B.B., Subileau, E.A., Perriere, N., Charneau, P., Holloway, K., Leveque, M., Tricoire-Leignel, H., Nicotra, A., Bourdoulous, S., Turowski, P., Male, D.K., Roux, F., Greenwood, J., Romero, I.A. & Couraud, P.O. Blood-brain barrier-specific properties of a human adult brain endothelial cell line. *FASEB J.* **2005**, 19, 1872-1874.
57. Scism, J.L., Laska, D.A., Horn, J.W., Gimple, J.L., Pratt, S.E., Shepard, R.L., Dantzig, A.H. & Wrighton, S.A. Evaluation of an in vitro coculture model for the blood-brain barrier: comparison of human umbilical vein endothelial cells (ECV304) and rat glioma cells (C6) from two commercial sources. *In Vitro Cell. Dev. Biol.-Anim.* **1999**, 35, 580-592.
58. Wang, Q., Rager, J.D., Weinstein, K., Kardos, P.S., Dobson, G.L., Li, J. & Hidalgo, I.J. Evaluation of the MDR-MDCK cell line as a permeability screen for the blood-brain barrier. *Int. J. Pharm.* **2005**, 288, 349-359.
59. Lundquist, S., Renftel, M., Brillault, J., Fenart, L., Cecchelli, R. & Dehouck, M.P. Prediction of drug transport through the blood-brain barrier in vivo: a comparison between two in vitro cell models. *Pharm. Res.* **2002**, 19, 976-981.
60. Deli, M.A., Abraham, C.S., Kataoka, Y. & Niwa, M. Permeability studies on in vitro blood-brain barrier models: physiology, pathology, and pharmacology. *Cell. Mol. Neurobiol.* **2005**, 25, 59-127.
61. Keep, R.F. Neural progenitor cells and blood-brain barrier modeling. *J. Neurochem.* **2011**, 119, 417-418.
62. Lippmann, E.S., Weidenfeller, C., Svendsen, C.N. & Shusta, E.V. Blood-brain barrier modeling with co-cultured neural progenitor cell-derived astrocytes and neurons. *J. Neurochem.* **2011**, 119, 507-520.
63. Onorati, M., Camnasio, S., Binetti, M., Jung, C.B., Moretti, A. & Cattaneo, E. Neuropotent self-renewing neural stem (NS) cells derived from mouse induced pluripotent stem (iPS) cells. *Mol. Cell. Neurosci.* **2010**, 43, 287-295.

64. Terasaki, T., Ohtsuki, S., Hori, S., Takanaga, H., Nakashima, E. & Hosoya, K. New approaches to in vitro models of blood-brain barrier drug transport. *Drug discovery today* **2003**, 8, 944-954.
65. Naik, P. & Cucullo, L. In vitro blood-brain barrier models: current and perspective technologies. *J. Pharm. Sci.* **2012**, 101, 1337-1354.
66. Kansy, M., Senner, F. & Gubernator, K. Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. *J. Med. Chem.* **1998**, 41, 1007-1010.
67. Mensch, J., Melis, A., Mackie, C., Verreck, G., Brewster, M.E. & Augustijns, P. Evaluation of various PAMPA models to identify the most discriminating method for the prediction of BBB permeability. *Eur. J. Pharm. Biopharm.* **2010**, 74, 495-502.
68. Di, L., Kerns, E.H., Fan, K., McConnell, O.J. & Carter, G.T. High throughput artificial membrane permeability assay for blood-brain barrier. *Eur. J. Med. Chem.* **2003**, 38, 223-232.
69. Kansy, M., Avdeef, A. & Fischer, H. Advances in screening for membrane permeability: high-resolution PAMPA for medicinal chemists. *Drug Discov. Today: Technologies* **2004**, 1, 349-355.
70. Bermejo, M., Avdeef, A., Ruiz, A., Nalda, R., Ruell, J.A., Tsinman, O., Gonzalez, I., Fernandez, C., Sanchez, G., Garrigues, T.M. & Merino, V. PAMPA-a drug absorption in vitro model 7. Comparing rat in situ, Caco-2, and PAMPA permeability of fluoroquinolones. *Eur. J. Pharm. Sci.* **2004**, 21, 429-441.
71. Ottaviani, G., Martel, S. & Carrupt, P.A. Parallel artificial membrane permeability assay: a new membrane for the fast prediction of passive human skin permeability. *J. Med. Chem.* **2006**, 49, 3948-3954.
72. Youdim, K.A., Avdeef, A. & Abbott, N.J. In vitro trans-monolayer permeability calculations: often forgotten assumptions. *Drug discovery today* **2003**, 8, 997-1003.
73. Newman, D.J. & Cragg, G.M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **2012**, 75, 311-335.
74. Cragg, G.M. & Newman, D.J. Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta* **2013**, 1830, 3670-3695.
75. Amedei, A. & D'Elios, M.M. New therapeutic approaches by using microorganism-derived compounds. *Curr. Med. Chem.* **2012**, 19, 3822-3840.
76. Bae, H., Chu, H., Edalat, F., Cha, J.M., Sant, S., Kashyap, A., Ahari, A.F., Kwon, C.H., Nichol, J.W., Manoucheri, S., Zamanian, B., Wang, Y. & Khademhosseini, A. Development of functional biomaterials with micro- and nanoscale technologies for tissue engineering and drug delivery applications. *J. Tissue Eng. Regen. Med.* **2012**, 1-14.
77. Spero, R.C., Vicci, L., Cribb, J., Bober, D., Swaminathan, V., O'Brien, E.T., Rogers, S.L. & Superfine, R. High throughput system for magnetic manipulation of cells, polymers, and biomaterials. *Rev. Sci. Instrum.* **2008**, 79, 083707.
78. Potyrailo, R., Rajan, K., Stoewe, K., Takeuchi, I., Chisholm, B. & Lam, H. Combinatorial and high-throughput screening of materials libraries: review of state of the art. *ACS Comb. Sci.* **2011**, 13, 579-633.
79. Inglese, J., Johnson, R.L., Simeonov, A., Xia, M., Zheng, W., Austin, C.P. & Auld, D.S. High-throughput screening assays for the identification of chemical probes. *Nat. Chem. Biol.* **2007**, 3, 466-479.
80. Hou, J., Liu, X., Shen, J., Zhao, G. & Wang, P.G. The impact of click chemistry in medicinal chemistry. *Expert Opin. Drug Discov.* **2012**, 7, 489-501.
81. Dolle, R.E., Le Bourdonnec, B., Goodman, A.J., Morales, G.A., Thomas, C.J. & Zhang, W. Comprehensive survey of chemical libraries for drug discovery and chemical biology: 2008. *J. Comb. Chem.* **2009**, 11, 739-790.
82. Dolle, R.E., Bourdonnec, B.L., Worm, K., Morales, G.A., Thomas, C.J. & Zhang, W. Comprehensive survey of chemical libraries for drug discovery and chemical biology: 2009. *J. Comb. Chem.* **2010**, 12, 765-806.
83. Kennedy, J.P., Williams, L., Bridges, T.M., Daniels, R.N., Weaver, D. & Lindsley, C.W. Application of combinatorial chemistry science on modern drug discovery. *J. Comb. Chem.* **2008**, 10, 345-354.
84. Kodadek, T. The rise, fall and reinvention of combinatorial chemistry. *Chem. Commun.* **2011**, 47, 9757-9763.
85. Lopez-Vallejo, F., Giulianotti, M.A., Houghten, R.A. & Medina-Franco, J.L. Expanding the medicinally relevant chemical space with compound libraries. *Drug Discov. Today* **2012**, 17, 718-726.
86. Macarron, R., Banks, M.N., Bojanic, D., Burns, D.J., Cirovic, D.A., Garyantes, T., Green, D.V., Hertzberg, R.P., Janzen, W.P., Paslay, J.W., Schopfer, U. & Sittampalam, G.S. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **2011**, 10, 188-195.

87. Mayr, L.M. & Bojanic, D. Novel trends in high-throughput screening. *Curr. Opin. Chem. Biol.* **2009**, 9, 580-588.
88. Berezowski, V., Landry, C., Lundquist, S., Dehouck, L., Cecchelli, R., Dehouck, M.P. & Fenart, L. Transport screening of drug cocktails through an in vitro blood-brain barrier: is it a good strategy for increasing the throughput of the discovery pipeline? *Pharm. Res.* **2004**, 21, 756-760.
89. Lam, K.S., Lebl, M. & Krchnak, V. The "One-Bead-One-Compound" Combinatorial Library Method. *Chem. Rev.* **1997**, 97, 411-448.
90. Merrifield, R.B. Peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **1963**, 85, 2149-2154.
91. Merrifield, R.B. Solid-phase synthesis. *Science* **1986**, 232, 341-347.
92. Pinilla, C., Appel, J.R., Borras, E. & Houghten, R.A. Advances in the use of synthetic combinatorial chemistry: mixture-based libraries. *Nat. Med.* **2003**, 9, 118-122.
93. Geysen, H.M., Meloen, R.H. & Barteling, S.J. Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amno acid. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, 81, 3998-4002.
94. Maeji, N.J., Bray, A.M. & Geysen, H.M. Multi-pin peptide synthesis strategy for T cell determinant analysis. *J. Immunol. Methods* **1990**, 134, 23-33.
95. Houghten, R.A. General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, 82, 5131-5135.
96. Beck-Sickinger, A.G., Durr, H. & Jung, G. Semiautomated T-bag peptide synthesis using 9-fluorenyl-methoxycarbonyl strategy and benzotriazol-1-yl-tetramethyl-uronium tetrafluoroborate activation. *Pept. Res.* **1991**, 4, 88-94.
97. Frank, R. & Döring, R. Simultaneous multiple peptide synthesis under continuous flow conditions on cellulose paper discs as segmental solid supports. *Tetrahedron* **1988**, 44, 6031-6040.
98. Geysen, H.M., Rodda, S.J. & Mason, T.J. A priori delineation of a peptide which mimics a discontinuous antigenic determinant. *Mol. Immunol.* **1986**, 23, 709-715.
99. Ivanetich, K.M. & Santi, D.V. Preparation of equimolar mixtures of peptides by adjustment of activated amino acid concentrations. *Meth. Enzymol.* **1996**, 267, 247-260.
100. Furka, A., Sebestyen, F., Asgedom, M. & Dibo, G. More peptides by less labour *Abstr. 10th Int. Symp. Med. Chem., Budapest, Hungary* **1988**, 288.
101. Furka, A., Sebestyen, F., Asgedom, M. & Dibo, G. General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Pept. Protein Res.* **1991**, 37, 487-493.
102. Lam, K.S., Salmon, S.E., Hersh, E.M., Hruby, V.J., Kazmierski, W.M. & Knapp, R.J. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **1991**, 354, 82-84.
103. Houghten, R.A., Pinilla, C., Blondelle, S.E., Appel, J.R., Dooley, C.T. & Cuervo, J.H. Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature* **1991**, 354, 84-86.
104. Lebl, M., Patek, M., Kocis, P., Krchnak, V., Hruby, V.J., Salmon, S.E. & Lam, K.S. Multiple release of equimolar amounts of peptides from a polymeric carrier using orthogonal linkage-cleavage chemistry. *Int. J. Pept. Protein Res.* **1993**, 41, 201-203.
105. Nefzi, A., Ostresh, J.M. & Houghten, R.A. The Current Status of Heterocyclic Combinatorial Libraries. *Chem. Rev.* **1997**, 97, 449-472.
106. Hughes, I. & Hunter, D. Techniques for analysis and purification in high-throughput chemistry. *Curr. Opin. Chem. Biol.* **2001**, 5, 243-247.
107. Jung, G. & Beck-Sickinger, A.G. Multiple peptide-synthesis methods and their applications. *Angew. Chem. Int. Ed.* **1992**, 367-383.
108. Pinilla, C., Appel, J.R., Blanc, P. & Houghten, R.A. Rapid identification of high affinity peptide ligands using positional scanning synthetic peptide combinatorial libraries. *BioTechniques* **1992**, 13, 901-905.
109. Déprez, B., Williard, X., Bourel, L., Coste, H., Hyafil, F. & Tartar, A. Orthogonal combinatorial chemical libraries. *J. Am. Chem. Soc.* **1995**, 117, 5405-5406.
110. Erb, E., Janda, K.D. & Brenner, S. Recursive deconvolution of combinatorial chemical libraries. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, 91, 11422-11426.
111. Beck-Sickinger, A. & Weber, P. Combinational Strategies in Biology and Chemistry. John Wiley & Sons, Chichester, Sussex; **2001**. pp. 82-93.
112. Kerr, J.M., Banville, S.C. & Zuckermann, R.N. Encoded combinatorial peptide libraries containing nonnatural amino-acids. *J. Am. Chem. Soc.* **1993**, 115, 2529-2531.

113. Nikolaiev, V., Stierandova, A., Krchnak, V., Seligmann, B., Lam, K.S., Salmon, S.E. & Lebl, M. Peptide-encoding for structure determination of nonsequenceable polymers within libraries synthesized and tested on solid-phase supports. *Pept. Res.* **1993**, 6, 161-170.
114. Luo, J., Zhang, H., Xiao, W., Kumaresan, P.R., Shi, C., Pan, C.X., Aina, O.H. & Lam, K.S. Rainbow beads: a color coding method to facilitate high-throughput screening and optimization of one-bead one-compound combinatorial libraries. *J. Comb. Chem.* **2008**, 10, 599-604.
115. Pastor, J.J., Lingard, I., Bhalay, G. & Bradley, M. Ion-extraction ladder sequencing from bead-based libraries. *J. Comb. Chem.* **2003**, 5, 85-90.
116. Kim, Y.G., Shin, D.S., Kim, E.M., Park, H.Y., Lee, C.S., Kim, J.H., Lee, B.S., Lee, Y.S. & Kim, B.G. High-throughput identification of substrate specificity for protein kinase by using an improved one-bead-one-compound library approach. *Angew. Chem. Int. Ed.* **2007**, 46, 5408-5411.
117. Ohlmeyer, M.H., Swanson, R.N., Dillard, L.W., Reader, J.C., Asouline, G., Kobayashi, R., Wigler, M. & Still, W.C. Complex synthetic chemical libraries indexed with molecular tags. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, 90, 10922-10926.
118. Nestler, H.P., Wennemers, H., Sherlock, R. & Dong, D.Y. Microautoradiographic identification of recepotr-ligand interation in bead-supported cominatorial libraries. *Bioorg. Med. Chem. Lett.* **1996**, 6, 1327-1330.
119. Ni, Z.J., Maclean, D., Holmes, C.P., Murphy, M.M., Ruhland, B., Jacobs, J.W., Gordon, E.M. & Gallop, M.A. Versatile approach to encoding combinatorial organic syntheses using chemically robust secondary amine tags. *J. Med. Chem.* **1996**, 39, 1601-1608.
120. Geysen, H.M., Wagner, C.D., Bodnar, W.M., Markworth, C.J., Parke, G.J., Schoenen, F.J., Wagner, D.S. & Kinder, D.S. Isotope or mass encoding of combinatorial libraries. *Chem. Biol.* **1996**, 3, 679-688.
121. Brenner, S. & Lerner, R.A. Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 5381-5383.
122. Scheuermann, J., Dumelin, C.E., Melkko, S. & Neri, D. DNA-encoded chemical libraries. *J. Biotechnol.* **2006**, 126, 568-581.
123. Nielsen, J., Brenner, S. & Janda, K.D. Synthetic methods for the implementation of encoded combinatorial chemistry. *J. Am. Chem. Soc.* **1993**, 115, 9812-9813.
124. Morgan, B., Hale, S., Africo-Muendel, C.C., Clark, M., Wagner, R., Israel, D.I., Gefter, M.L., Benjamin, D., Hansen, N., Kavarna, M.J., Creaser, S.P., Franklin, G.J., Centrella, P.A. & Acharya, R.A. Methods and building blocks for synthesis of combinato- rial libraries of mols. comprising functional moieties operatively linked to encoding oligonucleotides. **2004**, WO2005058479.
125. Freskgard, P.O., Franch, T., Goluliaev, A.H., Lundorf, M.D., Felding, J., Iksen, E.K., Holtmann, A., Jakobsen, S.N., Sams, C., Glad, S.S., Jensen, K.B. & Pedersen, H. Bifunctional substances and their use in prepa- ration and enzyme-based encoding of combinatorial libraries. **2004**, WO2004039825.
126. Kleiner, R.E., Dumelin, C.E., Tiu, G.C., Sakurai, K. & Liu, D.R. In vitro selection of a DNA-templated small-molecule library reveals a class of macrocyclic kinase inhibitors. *J. Am. Chem. Soc.* **2010**, 132, 11779-11791.
127. Gartner, Z.J., Tse, B.N., Grubina, R., Doyon, J.B., Snyder, T.M. & Liu, D.R. DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **2004**, 305, 1601-1605.
128. Gartner, Z.J. & Liu, D.R. The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules. *J. Am. Chem. Soc.* **2001**, 123, 6961-6963.
129. Tse, B.N., Snyder, T.M., Shen, Y. & Liu, D.R. Translation of DNA into a library of 13,000 synthetic small-molecule macrocycles suitable for in vitro selection. *J. Am. Chem. Soc.* **2008**, 130, 15611-15626.
130. Scheuermann, J. & Neri, D. DNA-encoded chemical libraries: a tool for drug discovery and for chemical biology. *Chembiochem* **2010**, 11, 931-937.
131. Melkko, S., Scheuermann, J., Dumelin, C.E. & Neri, D. Encoded self-assembling chemical libraries. *Nat. Biotechnol.* **2004**, 22, 568-574.
132. Melkko, S., Zhang, Y., Dumelin, C.E., Scheuermann, J. & Neri, D. Isolation of high-affinity trypsin inhibitors from a DNA-encoded chemical library. *Angew. Chem. Int. Ed.* **2007**, 46, 4671-4674.
133. Dumelin, C.E., Scheuermann, J., Melkko, S. & Neri, D. Selection of streptavidin binders from a DNA-encoded chemical library. *Bioconjugate Chem.* **2006**, 17, 366-370.
134. Scheuermann, J., Dumelin, C.E., Melkko, S., Zhang, Y., Mannocci, L., Jaggi, M., Sobek, J. & Neri, D. DNA-encoded chemical libraries for the discovery of MMP-3 inhibitors. *Bioconjugate Chem.* **2008**, 19, 778-785.

135. Melkko, S., Dumelin, C.E., Scheuermann, J. & Neri, D. On the magnitude of the chelate effect for the recognition of proteins by pharmacophores scaffolded by self-assembling oligonucleotides. *Chem. Biol.* **2006**, 13, 225-231.
136. Mannocci, L., Leimbacher, M., Wichert, M., Scheuermann, J. & Neri, D. 20 years of DNA-encoded chemical libraries. *Chem. Commun.* **2011**, 47, 12747-12753.
137. Mannocci, L., Melkko, S., Buller, F., Molnar, I., Bianke, J.P., Dumelin, C.E., Scheuermann, J. & Neri, D. Isolation of potent and specific trypsin inhibitors from a DNA-encoded chemical library. *Bioconjugate Chem.* **2010**, 21, 1836-1841.
138. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, 246, 64-71.
139. Wells, J.M. & McLuckey, S.A. Collision-induced dissociation (CID) of peptides and proteins. *Meth. Enzymol.* **2005**, 402, 148-185.
140. Laskin, J. & Futrell, J.H. Collisional activation of peptide ions in FT-ICR mass spectrometry. *Mass Spectrom. Rev.* **2003**, 22, 158-181.
141. Zubarev, R.A., Kelleher, N.L. & McLafferty, F.W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, 120, 3265-3266.
142. Mikesh, L.M., Ueberheide, B., Chi, A., Coon, J.J., Syka, J.E., Shabanowitz, J. & Hunt, D.F. The utility of ETD mass spectrometry in proteomic analysis. *Biochim. Biophys. Acta* **2006**, 1764, 1811-1822.
143. Syka, J.E., Coon, J.J., Schroeder, M.J., Shabanowitz, J. & Hunt, D.F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101, 9528-9533.
144. Coon, J.J., Shabanowitz, J., Hunt, D.F. & Syka, J.E. Electron transfer dissociation of peptide anions. *J. Am. Soc. Mass Spectrom.* **2005**, 16, 880-882.
145. Olsen, J.V., Macek, B., Lange, O., Makarov, A., Horning, S. & Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **2007**, 4, 709-712.
146. Grill, V., Shen, J., Evans, C. & Cooks, R.G. Collisions of ions with surfaces at chemically relevant energies: Instrumentation and phenomena. *Rev. Sci. Instrum.* **2001**, 72, 3149-3179.
147. Brodbelt, J.S. Shedding light on the frontier of photodissociation. *J. Am. Soc. Mass Spectrom.* **2011**, 22, 197-206.
148. Sleno, L. & Volmer, D.A. Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* **2004**, 39, 1091-1112.
149. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass. Spectrom.* **1984**, 11, 601.
150. Biemann, K. Mass spectrometry of peptides and proteins. *Annu. Rev. Biochem.* **1992**, 61, 977-1010.
151. Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **2004**, 5, 699-711.
152. Yates, J.R., 3rd Mass spectral analysis in proteomics. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, 33, 297-316.
153. Liu, T., Belov, M.E., Jaitly, N., Qian, W.J. & Smith, R.D. Accurate mass measurements in proteomics. *Chem. Rev.* **2007**, 107, 3621-3653.
154. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **2006**, 312, 212-217.
155. de Hoffmann, E. & Stroobant, V. Mass Spectrometry: Principles and Applications. Wiley, New York; **2007**. pp. 489.
156. Perry, R.H., Cooks, R.G. & Noll, R.J. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom. Rev.* **2008**, 27, 661-699.
157. Olsen, J.V., de Godoy, L.M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S. & Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, 4, 2010-2021.
158. Yates, J.R., Cociorva, D., Liao, L. & Zabrouskov, V. Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. *Anal. Chem.* **2006**, 78, 493-500.
159. Scigelova, M. & Makarov, A. Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* **2006**, 6, 16-21.
160. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, 422, 198-207.
161. Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.C. & Yates, J.R., 3rd Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* **2013**, 113, 2343-2394.
162. Eng, J.K., McCormack, A.L. & Yates III, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, 5, 976-989.

163. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, 20, 3551-3567.
164. Xu, H. & Freitas, M.A. MassMatrix: a database search program for rapid characterization of proteins and peptides from tandem mass spectrometry data. *Proteomics* **2009**, 9, 1548-1555.
165. Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L., Xu, M., Maynard, D.M., Yang, X., Shi, W. & Bryant, S.H. Open mass spectrometry search algorithm. *J. Proteome Res.* **2004**, 3, 958-964.
166. Craig, R., Cortens, J.P. & Beavis, R.C. Open source system for analyzing, validating, and storing protein identification data. *J. Proteome Res.* **2004**, 3, 1234-1242.
167. Katagiri, F. & Glazebrook, J. Overview of mRNA expression profiling using DNA microarrays. *Curr. Protoc. Mol. Biol.* **2009**, Chapter 22, Unit 22 24.
168. Baty, J.D. & Robinson, P.R. Single and multiple ion recording techniques for the analysis of diphenylhydantoin and its major metabolite in plasma. *Biomed. Mass. Spectrom.* **1977**, 4, 36-41.
169. Yost, R.A. & Enke, C.G. Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Anal. Chem.* **1979**, 51, 1251-1264.
170. Zweigenbaum, J. & Henion, J. Bioanalytical high-throughput selected reaction monitoring-LC/MS determination of selected estrogen receptor modulators in human plasma: 2000 samples/day. *Anal. Chem.* **2000**, 72, 2446-2454.
171. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **2008**, 4, 222.
172. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **2012**, 9, 555-566.
173. Mallick, P., Schirle, M., Chen, S.S., Flory, M.R., Lee, H., Martin, D., Ranish, J., Raught, B., Schmitt, R., Werner, T., Kuster, B. & Aebersold, R. Computational prediction of proteotypic peptides for quantitative proteomics. *Nat. Biotechnol.* **2007**, 25, 125-131.
174. Rinner, O., Mueller, L.N., Hubalek, M., Muller, M., Gstaiger, M. & Aebersold, R. An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. *Nat. Biotechnol.* **2007**, 25, 345-352.
175. Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. & Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, 17, 994-999.
176. Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. & Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, 1, 376-386.
177. Schmidt, A., Kellermann, J. & Lottspeich, F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **2005**, 5, 4-15.
178. Ong, S.E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **2005**, 1, 252-262.
179. Zhang, H., Liu, Q., Zimmerman, L.J., Ham, A.J., Slebos, R.J., Rahman, J., Kikuchi, T., Massion, P.P., Carbone, D.P., Billheimer, D. & Liebler, D.C. Methods for peptide and protein quantitation by liquid chromatography-multiple reaction monitoring mass spectrometry. *Mol. Cell. Proteomics* **2011**, 10, M110 006593.
180. Desiderio, D.M. & Kai, M. Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biomed. Mass. Spectrom.* **1983**, 10, 471-479.
181. Barr, J.R., Maggio, V.L., Patterson, D.G., Jr., Cooper, G.R., Henderson, L.O., Turner, W.E., Smith, S.J., Hannon, W.H., Needham, L.L. & Sampson, E.J. Isotope dilution--mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. *Clin. Chem.* **1996**, 42, 1676-1682.
182. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. & Gygi, S.P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 6940-6945.
183. Elschenbroich, S., Ignatchenko, V., Clarke, B., Kalloger, S.E., Boutros, P.C., Gramolini, A.O., Shaw, P., Jurisica, I. & Kislinger, T. In-depth proteomics of ovarian cancer ascites: combining shotgun proteomics and selected reaction monitoring mass spectrometry. *J. Proteome Res.* **2011**, 10, 2286-2299.
184. Dupuis, A., Hennekinne, J.A., Garin, J. & Brun, V. Protein Standard Absolute Quantification (PSAQ) for improved investigation of staphylococcal food poisoning outbreaks. *Proteomics* **2008**, 8, 4633-4636.
185. Mirzaei, H., McBee, J.K., Watts, J. & Aebersold, R. Comparative evaluation of current peptide production platforms used in absolute quantification in proteomics. *Mol. Cell. Proteomics* **2008**, 7, 813-823.

186. Sergeeva, A., Kolonin, M.G., Molldrem, J.J., Pasqualini, R. & Arap, W. Display technologies: application for the discovery of drug and gene delivery agents. *Adv. Drug Deliv. Rev.* **2006**, 58, 1622-1654.
187. Zacher, A.N., 3rd, Stock, C.A., Golden, J.W., 2nd & Smith, G.P. A new filamentous phage cloning vector: fd-tet. *Gene* **1980**, 9, 127-140.
188. Hines, J.C. & Ray, D.S. Construction and characterization of new coliphage M13 cloning vectors. *Gene* **1980**, 11, 207-218.
189. Sargent, T.D., Wu, J.R., Sala-Trepat, J.M., Wallace, R.B., Reyes, A.A. & Bonner, J. The rat serum albumin gene: analysis of cloned sequences. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, 76, 3256-3260.
190. Barbas III, C.F., Burton, D.R., Scott, J.K. & Silverman, G.J. Phage Display: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York; **2004**. pp. 1.1-2.19.
191. Marvin, D.A., Hale, R.D., Nave, C. & Helmer-Citterich, M. Molecular models and structural comparisons of native and mutant class I filamentous bacteriophages Ff (fd, f1, M13), If1 and IKe. *J. Mol. Biol.* **1994**, 235, 260-286.
192. Marvin, D.A. Filamentous phage structure, infection and assembly. *Curr. Opin. Struct. Biol.* **1998**, 8, 150-158.
193. Overman, S.A. & Thomas, G.J., Jr. Raman spectroscopy of the filamentous virus Ff (fd, fl, M13): structural interpretation for coat protein aromatics. *Biochemistry* **1995**, 34, 5440-5451.
194. Williams, K.A., Glibowicka, M., Li, Z., Li, H., Khan, A.R., Chen, Y.M., Wang, J., Marvin, D.A. & Deber, C.M. Packing of coat protein amphipathic and transmembrane helices in filamentous bacteriophage M13: role of small residues in protein oligomerization. *J. Mol. Biol.* **1995**, 252, 6-14.
195. Makowski, L. Terminating a macromolecular helix. Structural model for the minor proteins of bacteriophage M13. *J. Mol. Biol.* **1992**, 228, 885-892.
196. Deng, L.W., Malik, P. & Perham, R.N. Interaction of the globular domains of pIII protein of filamentous bacteriophage fd with the F-pilus of Escherichia coli. *Virology* **1999**, 253, 271-277.
197. Gray, C.W., Brown, R.S. & Marvin, D.A. Adsorption complex of filamentous fd virus. *J. Mol. Biol.* **1981**, 146, 621-627.
198. Armstrong, J., Perham, R.N. & Walker, J.E. Domain structure of bacteriophage fd adsorption protein. *FEBS Lett.* **1981**, 135, 167-172.
199. Crissman, J.W. & Smith, G.P. Gene-III protein of filamentous phages: evidence for a carboxyl-terminal domain with a role in morphogenesis. *Virology* **1984**, 132, 445-455.
200. Kremser, A. & Rasched, I. The adsorption protein of filamentous phage fd: assignment of its disulfide bridges and identification of the domain incorporated in the coat. *Biochemistry* **1994**, 33, 13954-13958.
201. Rakonjac, J., Feng, J. & Model, P. Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. *J. Mol. Biol.* **1999**, 289, 1253-1265.
202. Rakonjac, J., Bennett, N.J., Spagnuolo, J., Gagic, D. & Russel, M. Filamentous bacteriophage: biology, phage display and nanotechnology applications. *Curr. Iss. Mol. Biol.* **2011**, 13, 51-76.
203. Hill, D.F. & Petersen, G.B. Nucleotide sequence of bacteriophage f1 DNA. *J. Virol.* **1982**, 44, 32-46.
204. Lazzaroni, J.C., Germon, P., Ray, M.C. & Vianney, A. The Tol proteins of Escherichia coli and their involvement in the uptake of biomolecules and outer membrane stability. *FEMS Microbiol. Lett.* **1999**, 177, 191-197.
205. Webster, R.E. & Lopez, J. Structure and assembly of the class 1 filamentous bacteriophage. In *Virus structure and assembly*. Jones & Bartlett, Boston; **1985**. pp. 235-268.
206. Model, P. & Russel, M. Filamentous bacteriophage. In *The bacteriophages*. Plenum Publishing, New York; **1988**. pp. 375-456.
207. Russel, M. Moving through the membrane with filamentous phages. *Trends Microbiol.* **1995**, 3, 223-228.
208. Web site: http://rzv054.rz.tu-bs.de/Biotech/SD/M13livecycle.html, 24th November 2013.
209. Smith, G.P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **1985**, 228, 1315-1317.
210. Gao, C., Mao, S., Lo, C.H., Wirsching, P., Lerner, R.A. & Janda, K.D. Making artificial antibodies: a format for phage display of combinatorial heterodimeric arrays. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, 96, 6025-6030.
211. Jespers, L.S., Messens, J.H., De Keyser, A., Eeckhout, D., Van den Brande, I., Gansemans, Y.G., Lauwereys, M.J., Vlasuk, G.P. & Stanssens, P.E. Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. *Biotechnology (N Y)* **1995**, 13, 378-382.
212. Smith, G.P. & Petrenko, V.A. Phage Display. *Chem. Rev.* **1997**, 97, 391-410.

213. Iannolo, G., Minenkova, O., Petruzzelli, R. & Cesareni, G. Modifying filamentous phage capsid: limits in the size of the major capsid protein. *J. Mol. Biol.* **1995**, 248, 835-844.
214. Malik, P., Terry, T.D., Gowda, L.R., Langara, A., Petukhov, S.A., Symmons, M.F., Welsh, L.C., Marvin, D.A. & Perham, R.N. Role of capsid structure and membrane protein processing in determining the size and copy number of peptides displayed on the major coat protein of filamentous bacteriophage. *J. Mol. Biol.* **1996**, 260, 9-21.
215. Petrenko, V.A., Smith, G.P., Gong, X. & Quinn, T. A library of organic landscapes on filamentous phage. *Protein Eng.* **1996**, 9, 797-801.
216. Greenwood, J., Willis, A.E. & Perham, R.N. Multiple display of foreign peptides on a filamentous bacteriophage. Peptides from Plasmodium falciparum circumsporozoite protein as antigens. *J. Mol. Biol.* **1991**, 220, 821-827.
217. Kang, A.S., Barbas III, C.F., Janda, K.D., Benkovic, S.J. & Lerner, R.A. Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, 88, 4363-4366.
218. Sidhu, S.S., Weiss, G.A. & Wells, J.A. High copy display of large proteins on phage for functional selections. *J. Mol. Biol.* **2000**, 296, 487-495.
219. Weiss, G.A. & Sidhu, S.S. Design and evolution of artificial M13 coat proteins. *J. Mol. Biol.* **2000**, 300, 213-219.
220. Terry, T.D., Malik, P. & Perham, R.N. Accessibility of peptides displayed on filamentous bacteriophage virions: susceptibility to proteinases. *J. Biol. Chem.* **1997**, 378, 523-530.
221. Smith, G.P. Preface. *Gene* **1993**, 128, 1-2.
222. McConnell, S.J., Kendall, M.L., Reilly, T.M. & Hoess, R.H. Constrained peptide libraries as a tool for finding mimotopes. *Gene* **1994**, 151, 115-118.
223. Kishchenko, G., Batliwala, H. & Makowski, L. Structure of a foreign peptide displayed on the surface of bacteriophage M13. *J. Mol. Biol.* **1994**, 241, 208-213.
224. Dower, W.J., Miller, J.F. & Ragsdale, C.W. High efficiency transformation of E. coli by high voltage electroporation. *Nucleic Acids Res.* **1988**, 16, 6127-6145.
225. Parmley, S.F. & Smith, G.P. Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. *Gene* **1988**, 73, 305-318.
226. Lowman, H.B., Bass, S.H., Simpson, N. & Wells, J.A. Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry* **1991**, 30, 10832-10838.
227. Barbas III, C.F., Hu, D., Dunlop, N., Sawyer, L., Cababa, D., Hendry, R.M., Nara, P.L. & Burton, D.R. In vitro evolution of a neutralizing human antibody to human immunodeficiency virus type 1 to enhance affinity and broaden strain cross-reactivity. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, 91, 3809-3813.
228. Roberts, B.L., Markland, W., Ley, A.C., Kent, R.B., White, D.W., Guterman, S.K. & Ladner, R.C. Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 2429-2433.
229. Chen, Y., Wiesmann, C., Fuh, G., Li, B., Christinger, H.W., McKay, P., de Vos, A.M. & Lowman, H.B. Selection and analysis of an optimized anti-VEGF antibody: crystal structure of an affinity-matured Fab in complex with antigen. *J. Mol. Biol.* **1999**, 293, 865-881.
230. Di Niro, R., Sulic, A.M., Mignone, F., D'Angelo, S., Bordoni, R., Iacono, M., Marzari, R., Gaiotto, T., Lavric, M., Bradbury, A.R., Biancone, L., Zevin-Sonkin, D., De Bellis, G., Santoro, C. & Sblattero, D. Rapid interactome profiling by massive sequencing. *Nucleic Acids Res.* **2010**, 38, e110.
231. Rajotte, D., Arap, W., Hagedorn, M., Koivunen, E., Pasqualini, R. & Ruoslahti, E. Molecular heterogeneity of the vascular endothelium revealed by in vivo phage display. *J. Clin. Invest.* **1998**, 102, 430-437.
232. Koivunen, E., Restel, B.H., Rajotte, D., Lahdenranta, J., Hagedorn, M., Arap, W. & Pasqualini, R. Integrin-binding peptides derived from phage display libraries. *Methods Mol. Biol.* **1999**, 129, 3-17.
233. Adey, N.B., Mataragnon, A.H., Rider, J.E., Carter, J.M. & Kay, B.K. Characterization of phage that bind plastic from phage-displayed random peptide libraries. *Gene* **1995**, 156, 27-31.
234. Cortese, R., Felici, F., Galfre, G., Luzzago, A., Monaci, P. & Nicosia, A. Epitope discovery using peptide libraries displayed on phage. *Trends Biotechnol.* **1994**, 12, 262-267.
235. Sparks, A.B., Hoffman, N.G., McConnell, S.J., Fowlkes, D.M. & Kay, B.K. Cloning of ligand targets: systematic isolation of SH3 domain-containing proteins. *Nat. Biotechnol.* **1996**, 14, 741-744.
236. Cheadle, C., Ivashchenko, Y., South, V., Searfoss, G.H., French, S., Howk, R., Ricca, G.A. & Jaye, M. Identification of a Src SH3 domain binding motif by screening a random phage display library. *J. Biol. Chem.* **1994**, 269, 24034-24039.

237. Prezzi, C., Nuzzo, M., Meola, A., Delmastro, P., Galfre, G., Cortese, R., Nicosia, A. & Monaci, P. Selection of antigenic and immunogenic mimics of hepatitis C virus using sera from patients. *J. Immunol.* **1996**, 156, 4504-4513.
238. Felici, F., Luzzago, A., Folgori, A. & Cortese, R. Mimicking of discontinuous epitopes by phage-displayed peptides, II. Selection of clones recognized by a protective monoclonal antibody against the Bordetella pertussis toxin from phage peptide libraries. *Gene* **1993**, 128, 21-27.
239. Rebar, E.J. & Pabo, C.O. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **1994**, 263, 671-673.
240. Choo, Y., Sanchez-Garcia, I. & Klug, A. In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* **1994**, 372, 642-645.
241. Kriplani, U. & Kay, B.K. Selecting peptides for use in nanoscale materials using phage-displayed combinatorial peptide libraries. *Curr. Opin. Biotechnol.* **2005**, 16, 470-475.
242. Khoo, X., Hamilton, P., O'Toole, G.A., Snyder, B.D., Kenan, D.J. & Grinstaff, M.W. Directed assembly of PEGylated-peptide coatings for infection-resistant titanium metal. *J. Am. Chem. Soc.* **2009**, 131, 10992-10997.
243. Segvich, S.J., Smith, H.C. & Kohn, D.H. The adsorption of preferential binding peptides to apatite-based materials. *Biomaterials* **2009**, 30, 1287-1298.
244. Tomczak, M.M., Gupta, M.K., Drummy, L.F., Rozenzhak, S.M. & Naik, R.R. Morphological control and assembly of zinc oxide using a biotemplate. *Acta Biomater.* **2009**, 5, 876-882.
245. Nam, K.T., Wartena, R., Yoo, P.J., Liau, F.W., Lee, Y.J., Chiang, Y.M., Hammond, P.T. & Belcher, A.M. Stamped microbattery electrodes based on self-assembled M13 viruses. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, 105, 17227-17231.
246. Lee, Y.J., Lee, Y., Oh, D., Chen, T., Ceder, G. & Belcher, A.M. Biologically activated noble metal alloys at the nanoscale: for lithium ion battery anodes. *Nano Lett.* **2010**, 10, 2433-2440.
247. Lee, Y.J., Yi, H., Kim, W.J., Kang, K., Yun, D.S., Strano, M.S., Ceder, G. & Belcher, A.M. Fabricating genetically engineered high-power lithium-ion batteries using multiple virus genes. *Science* **2009**, 324, 1051-1055.
248. Dang, X., Yi, H., Ham, M.H., Qi, J., Yun, D.S., Ladewski, R., Strano, M.S., Hammond, P.T. & Belcher, A.M. Virus-templated self-assembled single-walled carbon nanotubes for highly efficient electron collection in photovoltaic devices. *Nat. Nanotechnol.* **2011**, 6, 377-384.
249. Dang, X., Qi, J., Klug, M.T., Chen, P.Y., Yun, D.S., Fang, N.X., Hammond, P.T. & Belcher, A.M. Tunable localized surface plasmon-enabled broadband light-harvesting enhancement for high-efficiency panchromatic dye-sensitized solar cells. *Nano Lett.* **2013**, 13, 637-642.
250. Nuraje, N., Dang, X., Qi, J., Allen, M.A., Lei, Y. & Belcher, A.M. Biotemplated synthesis of perovskite nanomaterials for solar energy conversion. *Adv. Mater.* **2012**, 24, 2885-2889.
251. Ghosh, D., Kohli, A.G., Moser, F., Endy, D. & Belcher, A.M. Refactored M13 bacteriophage as a platform for tumor cell imaging and drug delivery. *ACS Synth. Biol.* **2012**, 1, 576-582.
252. Ghosh, D., Lee, Y., Thomas, S., Kohli, A.G., Yun, D.S., Belcher, A.M. & Kelly, K.A. M13-templated magnetic nanoparticles for targeted in vivo imaging of prostate cancer. *Nat. Nanotechnol.* **2012**, 7, 677-682.
253. Yi, H., Ghosh, D., Ham, M.H., Qi, J., Barone, P.W., Strano, M.S. & Belcher, A.M. M13 phage-functionalized single-walled carbon nanotubes as nanoprobes for second near-infrared window fluorescence imaging of targeted tumors. *Nano Lett.* **2012**, 12, 1176-1183.
254. Nizak, C., Monier, S., del Nery, E., Moutel, S., Goud, B. & Perez, F. Recombinant antibodies to the small GTPase Rab6 as conformation sensors. *Science* **2003**, 300, 984-987.
255. Gao, J., Sidhu, S.S. & Wells, J.A. Two-state selection of conformation-specific antibodies. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, 106, 3071-3076.
256. Rouslahti, E. Drug targeting to specific vascular sites. *Drug discovery today* **2002**, 7, 1138-1143.
257. Kolonin, M., Pasqualini, R. & Arap, W. Molecular addresses in blood vessels as targets for therapy. *Curr. Opin. Chem. Biol.* **2001**, 5, 308-313.
258. Pasqualini, R. & Ruoslahti, E. Organ targeting in vivo using phage display peptide libraries. *Nature* **1996**, 380, 364-366.
259. Trepel, M., Arap, W. & Pasqualini, R. Modulation of the immune response by systemic targeting of antigens to lymph nodes. *Cancer Res.* **2001**, 61, 8110-8112.
260. Samoylova, T.I. & Smith, B.F. Elucidation of muscle-binding peptides by phage display screening. *Muscle & nerve* **1999**, 22, 460-466.
261. Arap, W., Haedicke, W., Bernasconi, M., Kain, R., Rajotte, D., Krajewski, S., Ellerby, H.M., Bredesen, D.E., Pasqualini, R. & Ruoslahti, E. Targeting the prostate for destruction through a vascular address. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 1527-1531.

262. Essler, M. & Ruoslahti, E. Molecular specialization of breast vasculature: a breast-homing phage-displayed peptide binds to aminopeptidase P in breast vasculature. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 2252-2257.
263. Kolonin, M.G., Pasqualini, R. & Arap, W. Teratogenicity induced by targeting a placental immunoglobulin transporter. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 13055-13060.
264. Kriplani, U. & Kay, B.K. Selecting peptides for use in nanoscale materials using phage-displayed combinatorial peptide libraries. *Current opinion in biotechnology* **2005**, 16, 470-475.
265. Joyce, J.A., Laakkonen, P., Bernasconi, M., Bergers, G., Ruoslahti, E. & Hanahan, D. Stage-specific vascular markers revealed by phage display in a mouse model of pancreatic islet tumorigenesis. *Cancer cell* **2003**, 4, 393-403.
266. Hoffman, J.A., Giraudo, E., Singh, M., Zhang, L., Inoue, M., Porkka, K., Hanahan, D. & Ruoslahti, E. Progressive vascular changes in a transgenic mouse model of squamous cell carcinoma. *Cancer cell* **2003**, 4, 383-391.
267. Arap, W., Kolonin, M.G., Trepel, M., Lahdenranta, J., Cardo-Vila, M., Giordano, R.J., Mintz, P.J., Ardelt, P.U., Yao, V.J., Vidal, C.I., Chen, L., Flamm, A., Valtanen, H., Weavind, L.M., Hicks, M.E., Pollock, R.E., Botz, G.H., Bucana, C.D., Koivunen, E., Cahill, D., Troncoso, P., Baggerly, K.A., Pentz, R.D., Do, K.A., Logothetis, C.J. & Pasqualini, R. Steps toward mapping the human vasculature by phage display. *Nat. Med.* **2002**, 8, 121-127.
268. Cardo-Vila, M., Zurita, A.J., Giordano, R.J., Sun, J., Rangel, R., Guzman-Rojas, L., Anobom, C.D., Valente, A.P., Almeida, F.C., Lahdenranta, J., Kolonin, M.G., Arap, W. & Pasqualini, R. A ligand peptide motif selected from a cancer patient is a receptor-interacting site within human interleukin-11. *PloS one* **2008**, 3, e3452.
269. Zurita, A.J., Troncoso, P., Cardo-Vila, M., Logothetis, C.J., Pasqualini, R. & Arap, W. Combinatorial screenings in patients: the interleukin-11 receptor alpha as a candidate target in the progression of human prostate cancer. *Cancer Res.* **2004**, 64, 435-439.
270. Kolonin, M.G., Sun, J., Do, K.A., Vidal, C.I., Ji, Y., Baggerly, K.A., Pasqualini, R. & Arap, W. Synchronous selection of homing peptides for multiple tissues by in vivo phage display. *FASEB J.* **2006**, 20, 979-981.
271. Pentz, R.D., Cohen, C.B., Wicclair, M., DeVita, M.A., Flamm, A.L., Youngner, S.J., Hamric, A.B., McCabe, M.S., Glover, J.J., Kittiko, W.J., Kinlaw, K., Keller, J., Asch, A., Kavanagh, J.J. & Arap, W. Ethics guidelines for research with the recently dead. *Nat. Med.* **2005**, 11, 1145-1149.
272. Pentz, R.D., Flamm, A.L., Pasqualini, R., Logothetis, C.J. & Arap, W. Revisiting ethical guidelines for research with terminal wean and brain-dead participants. *Hastings Cent. Rep.* **2003**, 33, 20-26.
273. Krag, D.N., Shukla, G.S., Shen, G.P., Pero, S., Ashikaga, T., Fuller, S., Weaver, D.L., Burdette-Radoux, S. & Thomas, C. Selection of tumor-binding ligands in cancer patients with phage display libraries. *Cancer Res.* **2006**, 66, 7724-7733.
274. Lindner, T., Kolmar, H., Haberkorn, U. & Mier, W. DNA libraries for the construction of phage libraries: statistical and structural requirements and synthetic methods. *Molecules* **2011**, 16, 1625-1641.
275. McBride, L.J. & Caruthers, M.H. An investigation of several deoxynucleoside phosphoramidites useful for synthesizing deoxyoligonucleotides. *Tetrahedron Lett.* **1983**, 24, 245-248.
276. Bossi, L. Context effects: translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message. *J. Mol. Biol.* **1983**, 164, 73-87.
277. Mena, M.A. & Daugherty, P.S. Automated design of degenerate codon libraries. *Protein Eng. Des. Sel.* **2005**, 18, 559-561.
278. Fellouse, F.A., Wiesmann, C. & Sidhu, S.S. Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101, 12467-12472.
279. Kayushin, A.L., Korosteleva, M.D., Miroshnikov, A.I., Kosch, W., Zubov, D. & Piel, N. A convenient approach to the synthesis of trinucleotide phosphoramidites--synthons for the generation of oligonucleotide/peptide libraries. *Nucleic Acids Res.* **1996**, 24, 3748-3755.
280. Yagodkin, A., Azhayev, A., Roivainen, J., Antopolsky, M., Kayushin, A., Korosteleva, M., Miroshnikov, A., Randolph, J. & Mackie, H. Improved synthesis of trinucleotide phosphoramidites and generation of randomized oligonucleotide libraries. *Nucleosides, nucleotides & nucleic acids* **2007**, 26, 473-497.
281. Gaytan, P., Contreras-Zambrano, C., Ortiz-Alvarado, M., Morales-Pablos, A. & Yanez, J. TrimerDimer: an oligonucleotide-based saturation mutagenesis approach that removes redundant and stop codons. *Nucleic Acids Res.* **2009**, 37, e125.

282. Neuner, P., Cortese, R. & Monaci, P. Codon-based mutagenesis using dimer-phosphoramidites. *Nucleic Acids Res.* **1998**, 26, 1223-1227.
283. Belcourt, M.F. & Farabaugh, P.J. Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell* **1990**, 62, 339-352.
284. DeGraaf, M.E., Miceli, R.M., Mott, J.E. & Fischer, H.D. Biochemical diversity in a phage display library of random decapeptides. *Gene* **1993**, 128, 13-17.
285. Rodi, D.J., Soares, A.S. & Makowski, L. Quantitative assessment of peptide sequence diversity in M13 combinatorial peptide phage display libraries. *J. Mol. Biol.* **2002**, 322, 1039-1052.
286. Makowski, L. & Soares, A. Estimating the diversity of peptide populations from limited sequence data. *Bioinformatics* **2003**, 19, 483-489.
287. Mandava, S., Makowski, L., Devarapalli, S., Uzubell, J. & Rodi, D.J. RELIC--a bioinformatics server for combinatorial peptide analysis and identification of protein-ligand interaction sites. *Proteomics* **2004**, 4, 1439-1460.
288. Pini, A. & Bracci, L. Phage display of antibody fragments. *Curr. Protein Pept. Sci.* **2000**, 1, 155-169.
289. Filpula, D. Antibody engineering and modification technologies. *Biomol. Eng.* **2007**, 24, 201-215.
290. Ladner, R.C. Constrained peptides as binding entities. *Trends Biotechnol.* **1995**, 13, 426-430.
291. McLafferty, M.A., Kent, R.B., Ladner, R.C. & Markland, W. M13 bacteriophage displaying disulfide-constrained microproteins. *Gene* **1993**, 128, 29-36.
292. Moffet, D.A. & Hecht, M.H. De novo proteins from combinatorial libraries. *Chem. Rev.* **2001**, 101, 3191-3203.
293. Heinis, C., Rutherford, T., Freund, S. & Winter, G. Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat. Chem. Biol.* **2009**, 5, 502-507.
294. Meyer, S.C., Shomin, C.D., Gaj, T. & Ghosh, I. Tethering small molecules to a phage display library: discovery of a selective bivalent inhibitor of protein kinase A. *J. Am. Chem. Soc.* **2007**, 129, 13812-13813.
295. Woiwode, T.F., Haggerty, J.E., Katz, R., Gallop, M.A., Barrett, R.W., Dower, W.J. & Cwirla, S.E. Synthetic compound libraries displayed on the surface of encoded bacteriophage. *Chem. Biol.* **2003**, 10, 847-858.
296. Dwyer, M.A., Lu, W., Dwyer, J.J. & Kossiakoff, A.A. Biosynthetic phage display: a novel protein engineering tool combining chemical and genetic diversity. *Chem. Biol.* **2000**, 7, 263-274.
297. Tian, F., Tsao, M.L. & Schultz, P.G. A phage display system with unnatural amino acids. *J. Am. Chem. Soc.* **2004**, 126, 15962-15963.
298. Liu, C.C., Mack, A.V., Tsao, M.L., Mills, J.H., Lee, H.S., Choe, H., Farzan, M., Schultz, P.G. & Smider, V.V. Protein evolution with an expanded genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, 105, 17688-17693.
299. Wang, L., Xie, J. & Schultz, P.G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, 35, 225-249.
300. Funke, S.A. & Willbold, D. Mirror image phage display--a method to generate D-peptide ligands for use in diagnostic or therapeutical applications. *Mol. Biosyst.* **2009**, 5, 783-786.
301. Sun, N., Funke, S.A. & Willbold, D. Mirror image phage display--generating stable therapeutically and diagnostically active peptides with biotechnological means. *J. Biotechnol.* **2012**, 161, 121-125.
302. Schumacher, T.N., Mayr, L.M., Minor, D.L., Jr., Milhollen, M.A., Burgess, M.W. & Kim, P.S. Identification of D-peptide ligands through mirror-image phage display. *Science* **1996**, 271, 1854-1857.
303. Taly, V., Kelly, B.T. & Griffiths, A.D. Droplets as microreactors for high-throughput biology. *Chembiochem* **2007**, 8, 263-272.
304. Bialkowska, A.B., Crisp, M., Bannister, T., He, Y., Chowdhury, S., Schurer, S., Chase, P., Spicer, T., Madoux, F., Tian, C., Hodder, P., Zaharevitz, D. & Yang, V.W. Identification of small-molecule inhibitors of the colorectal cancer oncogene Kruppel-like factor 5 expression by ultrahigh-throughput screening. *Mol. Cancer Ther.* **2011**, 10, 2043-2051.
305. Fallah-Araghi, A., Baret, J.C., Ryckelynck, M. & Griffiths, A.D. A completely in vitro ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution. *Lab on a chip* **2012**, 12, 882-891.
306. Krall, N., Scheuermann, J. & Neri, D. Small targeted cytotoxics: current state and promises from DNA-encoded chemical libraries. *Angew. Chem. Int. Ed.* **2013**, 52, 1384-1402.

RESULTS AND DISCUSSION

Design, synthesis and evaluation of small peptide libraries based on already described BBB-shuttles for performance and stability enhancement

Chapter 1

Assessing chirality role on passive transport across the BBB

1.1. Precedents

Physicochemical properties entirely determine the capacity of molecules to cross biological membranes by passive diffusion. Lipophilicity (governed by the presence of polar groups and/or H-bond donors/acceptors in the structure of a compound), molecular weight, peptide length and amino acid sequence are considered the main determinants of the capacity of peptides to cross the BBB by passive diffusion.[1]

Studies on tuning peptides that cross the BBB by passive diffusion had been carried out during last years in our group. Those studies assessed influence of diverse physicochemical properties on BBB passive transport, such as backbone *N*-methylation, side chain influence, halogenation, peptide length or flexibility among others. Best candidates of those studies were successfully used as BBB-shuttles.[2-4]

Here, we focused our effort to study another physicochemical feature that may be relevant: stereochemistry. While being always considered in active transport due to the well-accepted role of chirality in protein-protein and ligand-protein interaction, stereochemistry is a parameter overlooked on passive diffusion drug design. In this regard, we selected a well-known family of shuttles that cross the BBB by passive diffusion[5] to perform studies on stereochemistry.

Cell membrane components such as phospholipids and proteins are chiral. Since all these components are commonly found enantiomerically pure in nature, it has been postulated[6] that stereoisomers could be discriminated in cell membranes. It was previously shown that bio-membrane models, such as micelles and vesicles, discriminated between homochiral and heterochiral enantiomers of dipeptides.[7-10] Those studies hypothesized that in chiral aggregates recognition occurs in a chiral environment induced in an internal region of the aggregate by remote stereogenic centers. Chiral recognition cannot be simply ascribed to non-covalent specific

interactions between the solute and the monomers behaving as single entities, but due to the aggregate as a whole. Potential discrimination can occur in regions of the aggregate/membrane that are quite far away from those stereogenic centers. In addition, it is noteworthy that these chiral interactions seem to confer a particular conformation to the stereoisomers when approaching the membrane that finally determines the fate of the molecule when interacting with it. Furthermore, another study demonstrated that diastereomeric peptides show distinctive lipophilicity and permeability across the BBB.[11]

1.2. Design and synthesis

To study the effect of peptide stereochemistry on membrane permeation, we prepared a library consisting of 16 stereoisomeric peptides derived from all-L Ac-(*N*-MePhe)$_4$-NH$_2$, which was previously described as an efficient BBB-shuttle by passive diffusion.[2, 5]

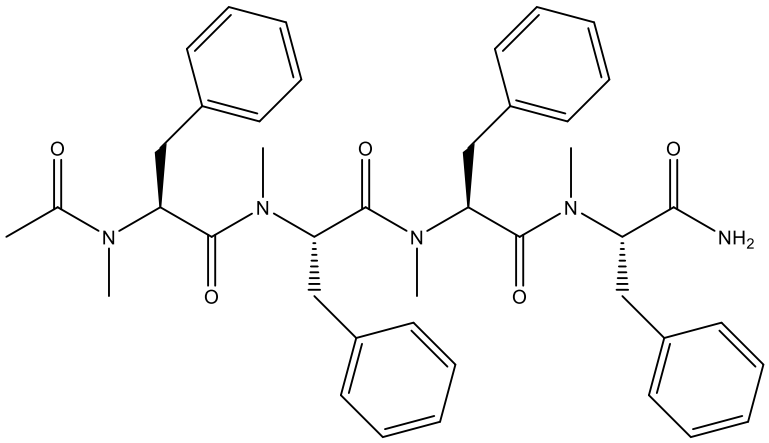


Figure 1.1. Structure of Ac-(*N*-MePhe)$_4$-NH$_2$ peptide in its all-L version.

Peptides were synthesized by solid-phase peptide synthesis with Fmoc/*t*Bu strategy, as it has several advantages over the Boc/Bzl strategy such a milder reaction conditions and ability to monitor deprotection by UV.[12]

Sieber resin was used in order to obtain *C*-terminal amide peptides. Sieber resin only requires 2% of trifluoroacetic acid (TFA) for cleavage. This allows skipping side reactions and by-product formation reported on cleavage step on acetylated *N*-metyl rich peptides (loss of *N*-acetyl-*N*-methyl aminoacid terminus and/or cleavage of amide bonds between consecutive *N*-methyl amino acids).[13]

The synthesis was performed at a 200 µmol scale. Oxima®, COMU and DIEA were used for couplings.[14, 15] Commercially available both Fmoc-L-*N*-MePhe-OH and Fmoc-D-*N*-MePhe-OH were used as building blocks. The coupling protocol was applied twice as the coupling reaction was always performed onto a secondary amine. All peptides were acetylated in the *N*-terminus on solid support. Preformation of the anhydride is required using acetic acid and DIPCDI. Coupling reactions were monitored by the chloranil test[16] or the *p*-nitrophenyl test (De Clercq test).[17]

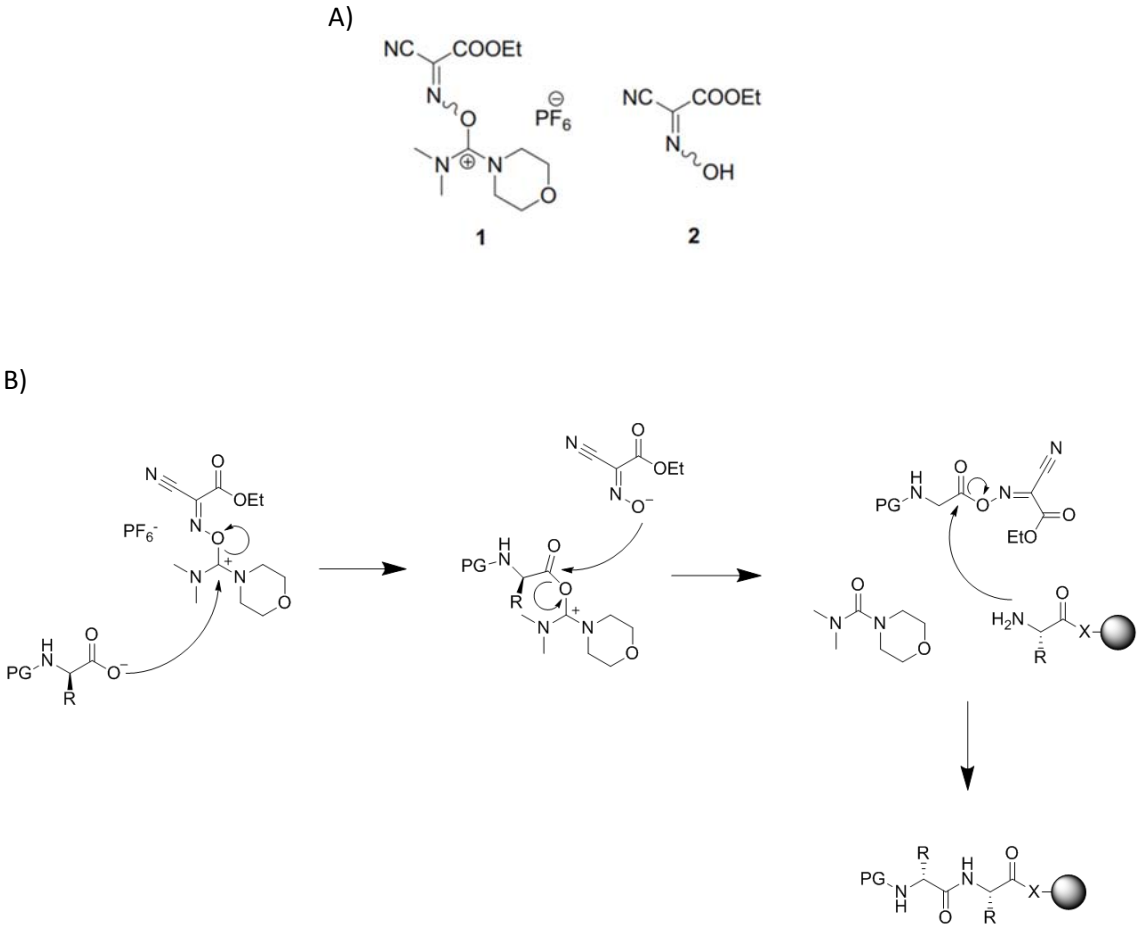
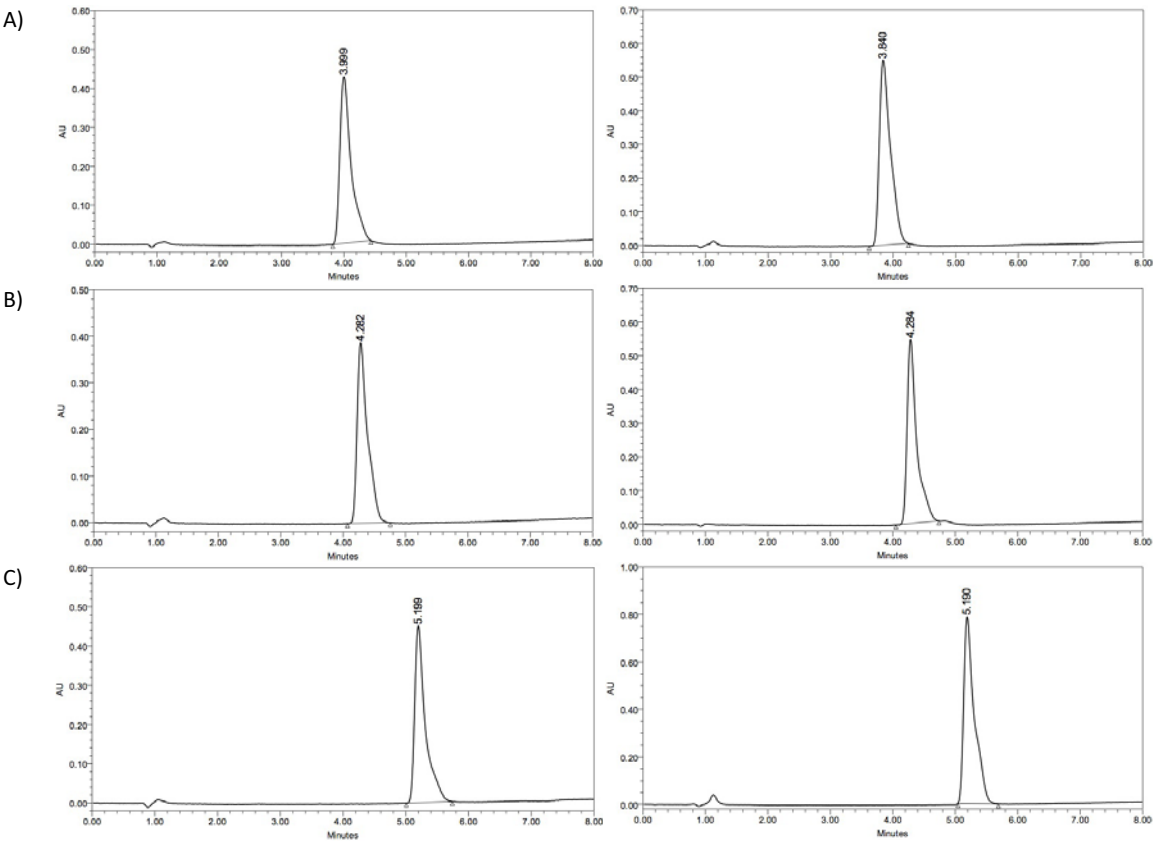


Figure 1.2. Chemical structure (A) of COMU (1) and Oxima Pure (2), and its reaction mechanism (B).

Peptides were obtained in purities higher than 95%, calculated by RP-HPLC.



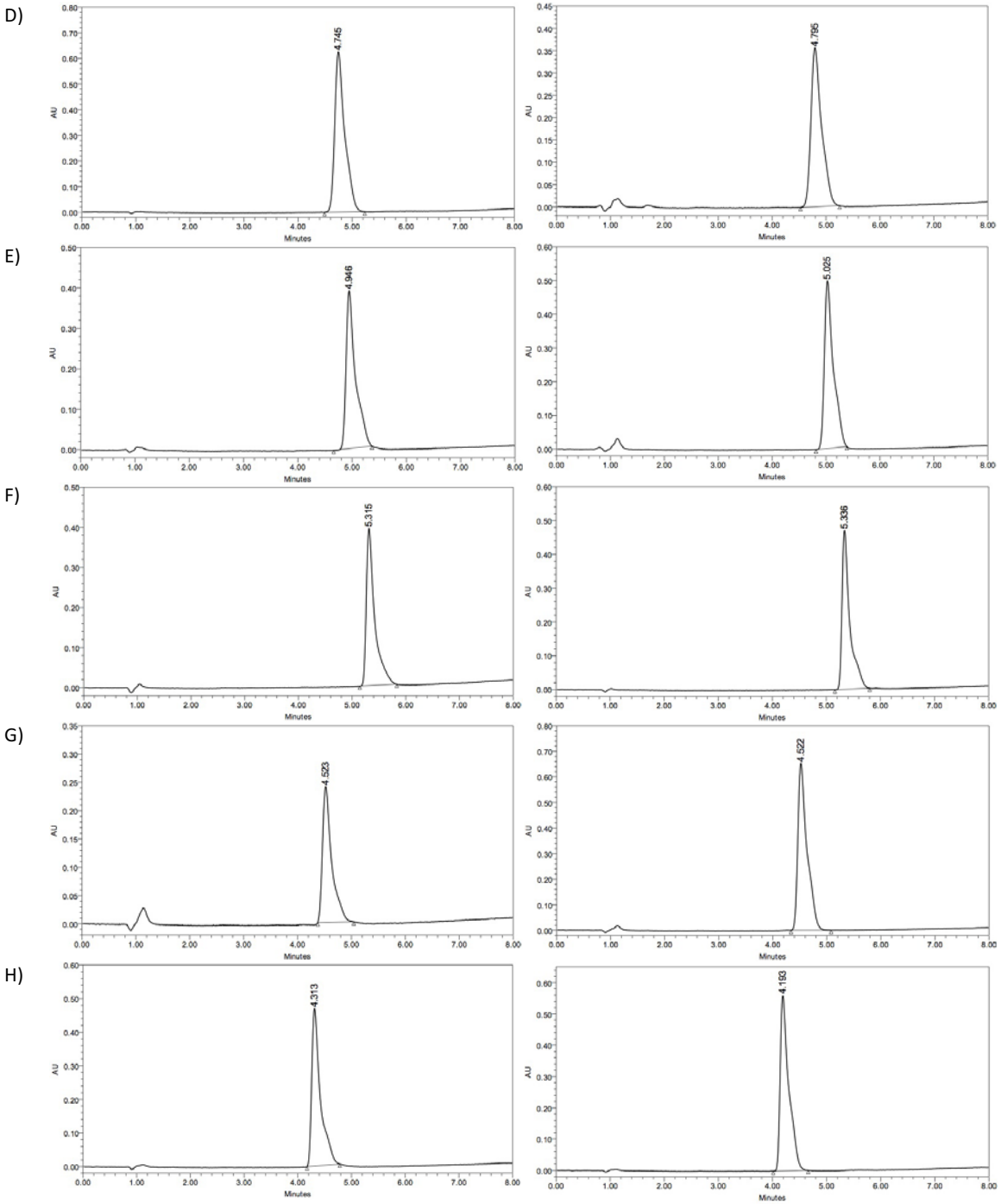


Figure 1.3. HPLC chromatograms of synthesized peptides (Ac-(*N*-MePhe)$_4$-NH$_2$) obtained in purities higher than 95%. D and L stand for D-*N*-MePhe and L-*N*-MePhe respectively. A) LLLL and DDDD; B) LDDD and DLLL; C) LDLL and DLDD; D) DDLL and LLDD; E) LLDL and DDLD; F) LDLD and DLDL; G) LDDL and DLLD; and H) DDDL and LLLD.

1.3. Evaluation: parallel artificial membrane permeability assay

Parallel artificial membrane permeability assay (PAMPA) is an *in vitro* transport assay based on artificial phospholipid membranes that only mimics passive diffusion properties of the BBB. Being an easy-handling and non-expensive assay, it is very convenient to assess distinct performance properties of our synthesized stereoisomeric library.

Diastereomers were placed in the donor compartment (lower compartment in PAMPA plate) at a concentration of 200 µM in pION commercial buffer with 20% of 1-propanol as a cosolvent to ensure complete solubility of candidates.[18, 19] The assay was performed at room temperature with GUT box shaker at an unstirred water layer (UWL) of 25 µm (commercial recommended

conditions) and saturated relative humidity for 4 hours. Shaking was performed to minimize UWL that may results in significant bias in permeability values not mimicking properly lipophilic *in vivo* conditions in the BBB.

1.4. Transport quantification of library peptides

Effective permeability, P_e, and percentage of transport (*%T*) were assessed with the following equations:

$$P_e = \frac{-218.3}{t} \cdot log\left[1 - \frac{2 \cdot Q_A(t)}{Q_D(t_0)}\right] 10^{-6} \ cm/s$$

$$\%T = \frac{Q_A(t)}{Q_D(t_0)} \cdot 100$$

where $Q_A(t)$ is the compound amount at the acceptor well at time *t* (area under the HPLC peak), and $Q_D(t_0)$ is the compound amount in the donor well at *t = 0* (area under the HPLC chromatographic peak).

Compound	P_e (× 10^{-6}) cm/s	Transport (%)	Membrane retention (%)	HPLC-RT min
Propranolol	5.7 ± 0.9	10.7 ± 1.4	55.1 ± 9.7	4.96
Carbamazepine	8.5 ± 0.3	15.1 ± 0.4	< 1	6.60
Ac-(DDDD)-NH_2	9.4 ± 0.9	16.4 ± 1.3	20.7 ± 5.0	3.8
Ac-(LLLL)-NH_2	6.6 ± 0.3	12.2 ± 0.5	42.0 ± 0.2	4
Ac-(LLLD)-NH_2	5.6 ± 0.2	10.5 ± 0.4	48.4 ± 0.5	4.2
Ac-(DDDL)-NH_2	6.1 ± 0.1	11.4 ± 0.1	45.4 ± 0.7	4.3
Ac-(LDDD)-NH_2	4.0 ± 0.9	7.8 ± 1.6	76.0 ± 2.5	4.3
Ac-(DLLL)-NH_2	4.0 ± 1.1	7.7 ± 1.9	58.1 ± 7.5	4.3
Ac-(DLLD)-NH_2	3.1 ± 0.1	6.2 ± 0.1	64.1 ± 0.4	4.5
Ac-(LDDL)-NH_2	4.4 ± 1.4	8.4 ± 2.5	43.3 ± 4.1	4.5
Ac-(DDLL)-NH_2	2.9 ± 0.4	5.9 ± 0.8	74.2 ± 6.2	4.7
Ac-(LLDD)-NH_2	2.3 ± 0.1	4.6 ± 0.2	81.6 ± 1.5	4.8
Ac-(LLDL)-NH_2	4.1 ± 0.2	8.0 ± 0.4	64.4 ± 0.8	4.9
Ac-(DDLD)-NH_2	2.7 ± 0.3	5.4 ± 0.7	75.1 ± 3.6	5
Ac-(DLDD)-NH_2	3.3 ± 0.1	6.6 ± 0.2	65.7 ± 0.6	5.2
Ac-(LDLL)-NH_2	5.0 ± 0.4	9.6 ± 0.7	56.7 ± 2.3	5.2
Ac-(LDLD)-NH_2	3.0 ± 0.2	6.0 ± 0.4	71.3 ± 0.1	5.3
Ac-(DLDL)-NH_2	2.0 ± 0.2	4.2 ± 0.4	78.4 ± 1.8	5.3

Table 1.1. Effective permeability (P_e), percentage of transport (*%T*), membrane retention (*%R*) in the PAMPA after 4 h and HPLC-RT of 16 stereoisomers and control compounds (Propranolol, Carbamazepine). Data are expressed as the mean ± SD. D and L stand for D-*N*-MePhe and L-*N*-MePhe respectively.

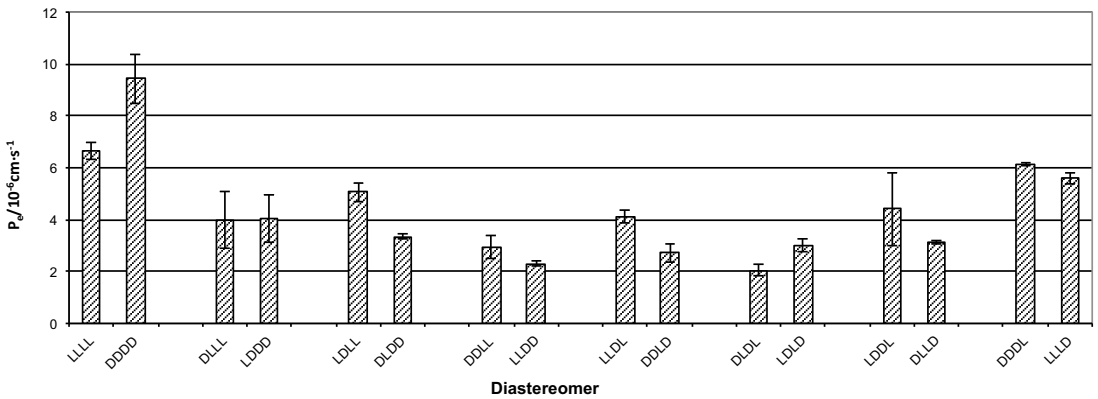


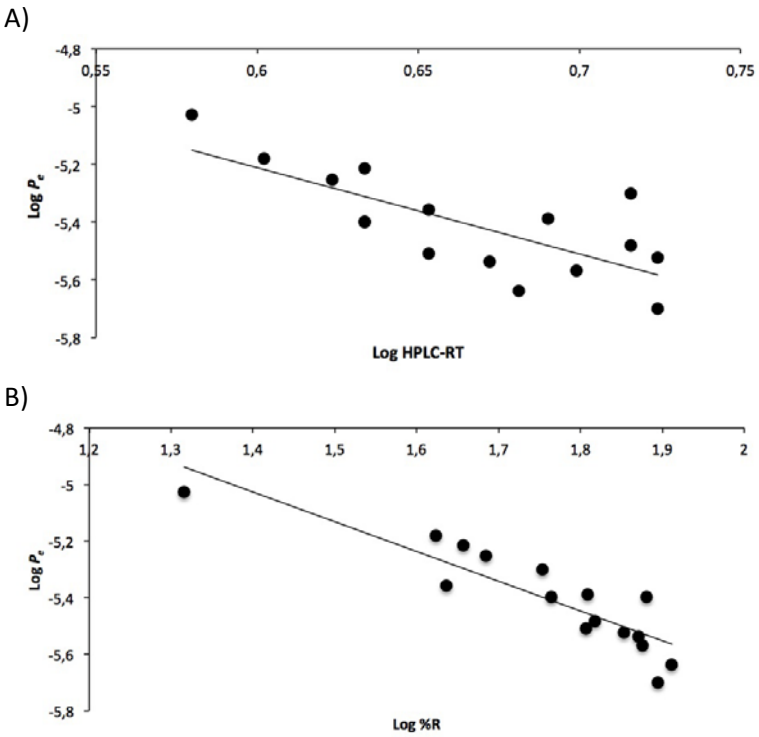
Figure 1.4. Illustration of results shown on Table 1.1. Enantiomeric couples are displayed together. D and L stand for D-*N*-MePhe and L-*N*-MePhe respectively.

It is widely agreed that lipophilicity plays a prominent role in molecular permeation across the BBB (by passive diffusion).[20] For that reason, at the beginning of the project, we were looking for a tool to measure lipophilicity, in order to rationalize our results and try to find a way to predict tendencies of behaviour of molecules across PAMPA assay.

We resolved ranking our molecules in terms of lipophilicity by HPLC retention time. HPLC retention time in reversed phase columns (HPLC-RT) is a good measure of the polarity of a molecule, especially when comparing stereoisomers of the same compound, being an indicator of relative lipophilicity.[21, 22]

We found that the homochiral peptides (LLLL and DDDD), although being less lipophilic than heterochiral peptides, exhibited higher permeability and lower membrane retention (%R) in PAMPA (Table 1.1). Between homochiral versions, the all-D version showed higher permeation than the previously reported all-L version.

Despite it is generally accepted that, in passive diffusion, the greater the lipophilicity, the greater the permeability across membranes, for these 16 peptides an inverse correlation was detected between lipophilicity and permeability (r = 0.75, Figure 1.5.A). Furthermore, permeability correlated inversely with membrane retention (%R) (r = 0.89, Figure 1.5.B). In contrast, there was a direct correlation between lipophilicity and membrane retention (%R), (r = 0.72, Figure 1.5.C). These results indicate that in this series of stereoisomeric peptides, homochiral peptides (LLLL and DDDD) had the optimal lipophilicity and/or conformational flexibility. The heterochiral peptides might show excessive lipophilicity, hampering their passage across the PAMPA assay due to an excessive retention within the membrane. The observed behaviour is probably consequence of differences in the more stable conformations adopted by each stereoisomer in the presence of the lipid bilayers. Our findings are consistent with previous studies reporting that heterochiral dipeptides (LD and DL) bond more strongly to chiral micellar phases (being trapped in the hydrophobic micellar portion of the described model) than homochiral enantiomers (LL and DD).[7-9]



C)

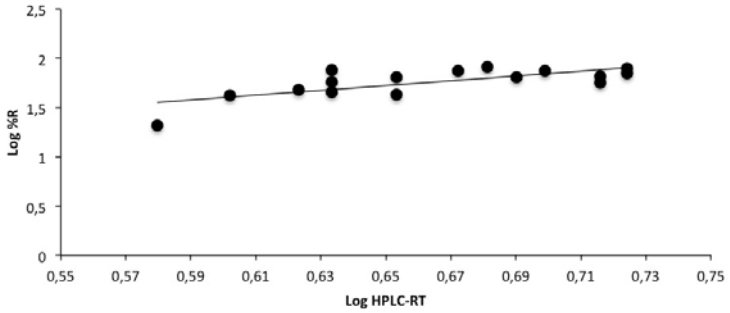


Figure 1.5. A) Inverse correlation between Log P_e of 16 stereoisomers and their Log HPLC-RT (r = 0.75). B) Inverse correlation between Log P_e of 16 stereoisomers and their Log %R (r = 0.89). C) Correlation between Log HPLC-RT of 16 stereoisomers and their Log *%R* (r = 0.72).

Looking now to the presence or not of enantiomeric discrimination along all the stereoisomers[8, 9, 23] our observations in the PAMPA membrane showed that enantiomeric discrimination was remarkable for most of the enantiomeric pairs, although it was very low for the pairs LDDD/DLLL and DDDL/LLLD. There is not a clear global trend; in general, the observed enantiomeric discrimination is higher for heterochiral enantiomers than for homochiral peptides. Enantiomeric discrimination is expressed in percentage and calculated in accordance with:

$$Enantiomeric\ discrimination\ (\%) = \frac{Best - Worst}{Worst} \cdot 100$$

where *best* and *worst* correspond to the species with higher and lower transport in the enantiomeric couple, respectively.

Enantiomeric couple	Enantiomeric discrimination (%)
DDDD/LLLL	42.2 ± 4.7
LDDD/DLLL	0.68 ± 0.25
LDLL/DLDD	51.5 ± 4.2
DDLL/LLDD	28.0 ± 2.5
LLDL/DDLD	51.4 ± 7.3
LDLD/DLDL	46.2 ± 6.0
LDDL/DLLD	41.4 ± 13.4
DDDL/LLLD	9.8 ± 0.4

Table 1.2. Enantiomeric discrimination results. D and L stand for D-*N*-MePhe and L-*N*-MePhe respectively.

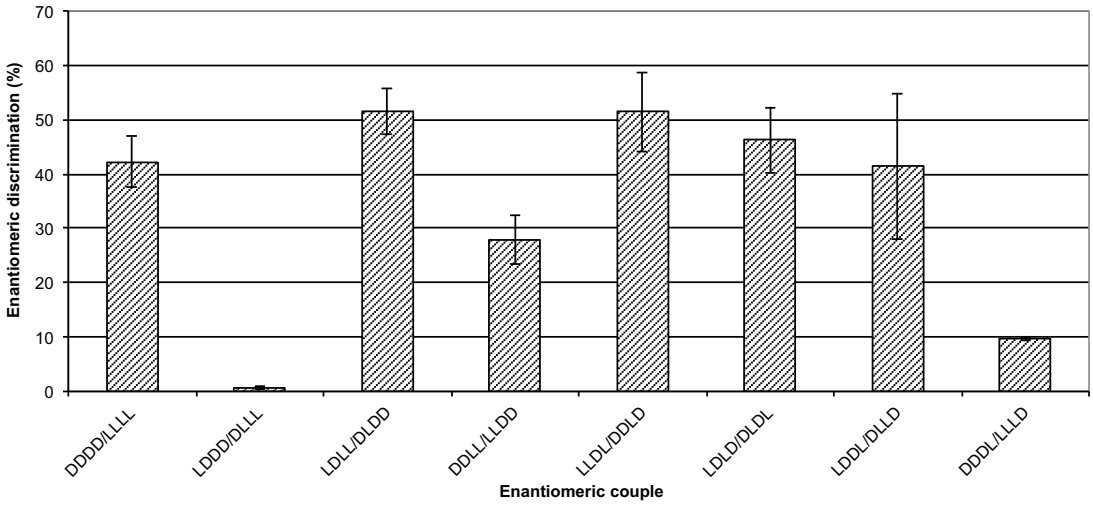


Figure 1.6. Enantiomeric discrimination results shown in Table 1.2. D and L stand for D-*N*-MePhe and L-*N*-MePhe, respectively.

We finally conclude that stereoisomers of peptide Ac-(*N*-MePhe)$_4$-NH$_2$ exhibit distinct crossing rates on PAMPA assay with a preference for homochiral peptide versions (especially all-D). An inverse correlation between membrane retention and transport was observed. We also detected a degree of enantiomeric discrimination, although we were unable to rationalize its behaviour.

1.5. UV detection vs MS detection for quantification

Integration of HPLC-UV peaks for transport quantification has been a widely used technique in our laboratory for peptide evaluation across biological models. However, it has some intrinsic limitations that move us one step forward to explore MS as a quantification technique. Generally speaking, MS allows higher sensitivity (targeted analysis), increase of selectivity by detection of specific m/z in complex samples and if necessary with MS/MS, and it is faster. On the other hand, it is quite more expensive.

MS it is being used for quantification of molecules in a wide variety of fields. Some examples recently reported compare both UV and MS techniques showing coherent data.[24-26]

1.5.1. MS technique selection for quantification

Both relative and absolute quantification are valid approaches for transport evaluation of *in vitro* experiments. Isotopically labelled pair-peptides (also so-called light and heavy versions) are needed for quantification since distinct peptides exhibit a non-comparable performance in front of MS detection. Those isotopically labelled peptides (heavy) are molecules equal to the one to be quantified (light) but with higher atomic mass.

Stable-isotope dilution (SID) theory, states that isotopomers are chemically identical to its native version, behaving identically during both chromatographic and mass spectrometry stages. Artificial isotopomers can be generated chemically, metabolically, enzymatically or by providing standard peptides containing heavy stable isotopes such as D, ^{13}C, ^{15}N or ^{18}O. Chemical labelling is performed at the peptide or protein level by chemical synthesis or modification, whereas metabolic labelling takes place during cell growth and enzymatic labelling during protein hydrolysis. On the other hand, standard peptides are synthesized chemically incorporating stable isotope-containing building blocks. Those approaches create the heavy version of the studied molecules (light version) or mixture. While comparing the intensities (or peak area) of both differently labelled peptides (or molecules) in the same spectra we can account for their relative quantities. Absolute quantification can be achieved by the addition of a known quantity of a stable isotope-labelled standard peptide and subsequent comparison of both mass spectrometric signals.

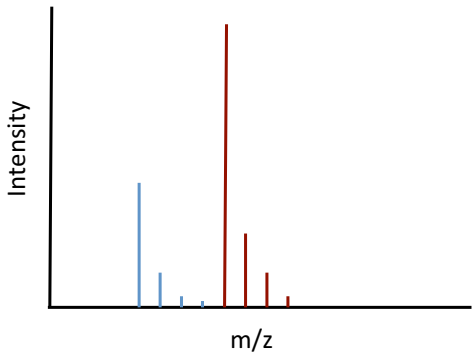


Figure 1.7. Schematic representation of MS spectra pattern on an spiking (with heavy peptide version, red, in this case being 4 a.m.u. heavier) experiment for quantification of a given peptide (light, blue). Intensities or peaks areas (of molecular peaks and isotopic envelopes) are compared for quantification assessment.

MS technique selection basically depends on sample complexity and equipment availability. MALDI-TOF was the technique of choice for single peptide analysis, and we postulate it could also be useful for simple peptide mixtures.

In the present case, our aim was to set up the methodology using an internal standard synthesized by SPPS. Although for Ac-(*N*-MePhe)$_4$-NH$_2$ or molecules tested alone, quantification could be achieved by UV standard methods we were looking for setting up the technique for future quantification analysis of simple peptide mixtures. Furthermore, UV methods work nicely when molecules absorb in wavelength comprised between 210 nm and 800 nm. However, there are molecules that might have poor absorbance or simply they are present in small amounts to be clearly detected and HPLC-UV peaks may be tiny, thus compromising quantification assessment.

Introduction of isotopically labelled tags during the synthetic workflow (SPPS) is the most suitable approach for synthetic isotopically labelled peptides. Tags should meet some requirements. Tags must not interfere in transport abilities of studied molecules. Ideally, tags should be placed near the *N*-terminus end of peptides facilitating the synthesis of molecules by splitting the synthesis resin as late as possible in the synthetic workflow. Isotopes must be stable. Finally, tags should be as cheap as possible.

1.5.2. Internal standard synthesis

In the bibliography there are a wide variety of described tags.[27] To meet requirements stated above each case must be studied according to nature of experiment (in the present work, a biological transport model) and MS techniques.

Burlina *et al.* introduced a biotin-(Gly)$_4$ tag in the *N*-terminus of studied peptides by chemical synthesis (SPPS).[28-30] This tag allows them to recover internalized peptides by washing and lysating the cells. Then, by adding streptavidin coated magnetic beads they collect internalized peptides with a magnet. This elegant strategy enables them to quantify the internalization into cells of studied cell-penetrating peptides (CPPs) and also study cargo intracellular degradation.

Derivatization-based strategies to improve the analysis of peptides by enhancement of MALDI MS signal intensities had been described. In these technologies, the matrix absorbs the laser UV light and transmits energy to the directly linked peptide molecules promoting their desorption and ionization. In this way, a peptide derivatization with α-cyano-4-hydorxycinnamic (ACH) acid, which absorbs UV light, improve identification through the relative enhancement over unlabelled/underivatized molecules in a neutral matrix, such as α-cyano-4-hydorxycinnamic methyl ester.[31] Paramelle *et al*. successfully used this strategy to quantify cellular uptake of a variety of CPPs.[32]

In our case, we want to quantify the transport across *in vitro* BBB models of peptides. PAMPA assay, mimicking exclusively passive diffusion from the BBB it is very sensible to any modification in the molecule affecting its physicochemical properties. For this reason, we didn't make any radical change to the studied molecule: Ac-(*N*-MePhe)$_4$-NH$_2$.

Even though there are isotopically labelled phenylalanines commercially available as building block for SPPS (phenylalanine-carboxy-^{13}C) they are very expensive. Furthermore, this approach would require synthesizing two different peptides from scratch: all phenylalanines must be isotopically labelled to get 4 atomic mass units difference when using phenylalanine-1-^{13}C.

Instead, by using the acetyl group as a tag we can easily reach three or four mass units by using CD$_3$-COOD or CD$_3$-^{13}COOD, respectively, to synthesize the heavy peptide. Interestingly, the synthesis of both peptides can be carried out in the same reactor until last synthetic step, where resin can be divided to perform capping with acetic acid and isotopically labelled acetic acid in two separate reactors.

In order to increase 5 atomic mass units the mass difference between pair-molecules, some authors suggest the use of another tag.[33] In this strategy benzyl group replace acetyl moiety in both peptides (light and heavy) by reaction of benzoic acid: C_6H_5COOH and C_6D_5COOH, respectively.

To set up the methodology, we tested all versions stated above to choose the best tag pair.

Synthesis of CH_3CO-(*N*-MePhe)$_4$-NH_2, CD_3CO-(*N*-MePhe)$_4$-NH_2, $CD_3{}^{13}CO$-(*N*-MePhe)$_4$-NH_2, C_6H_5CO-(*N*-MePhe)$_4$-NH_2, C_6D_5CO-(*N*-MePhe)$_4$-NH_2 was carried out by SPPS with Fmoc/*t*Bu strategy. Resin and coupling reagents conditions were the same described in this section (1.2).

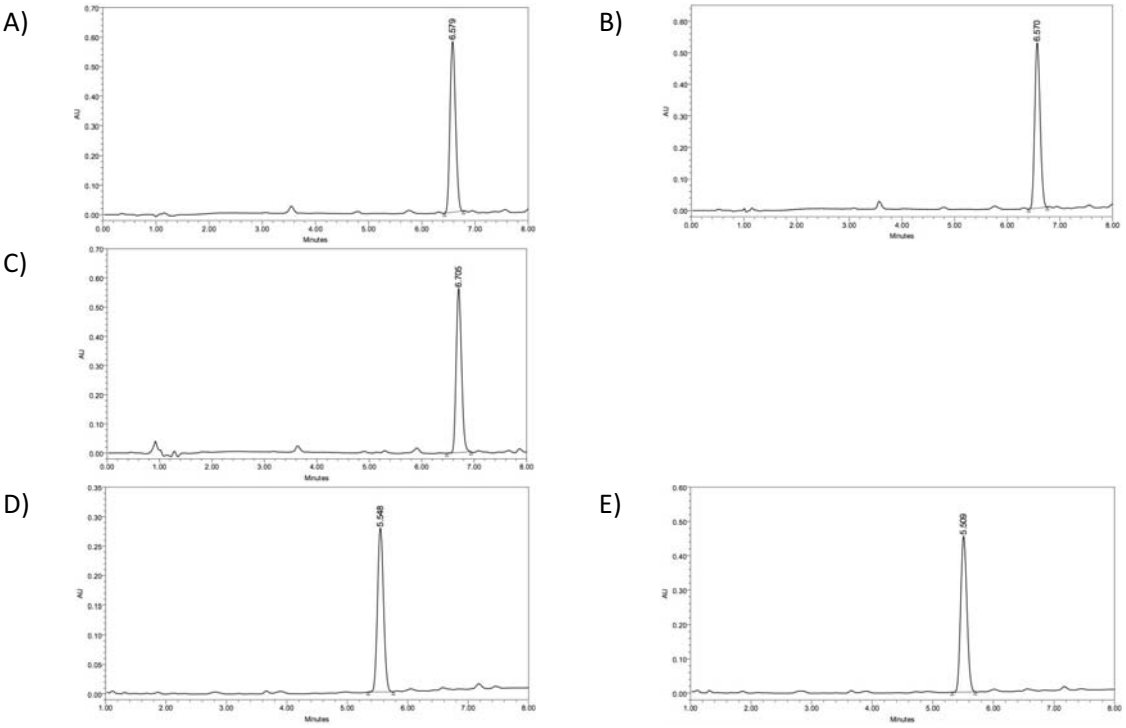


Figure 1.8. HPLC chromatograms of synthesized peptides. A) CH_3CO-(*N*-MePhe)$_4$-NH_2; B) CD_3CO-(*N*-MePhe)$_4$-NH_2; C) $CD_3{}^{13}CO$-(*N*-MePhe)$_4$-NH_2; D) C_6H_5CO-(*N*-MePhe)$_4$-NH_2; and E) C_6D_5CO-(*N*-MePhe)$_4$-NH_2.

Peptides were obtained in purities higher than 95%. Light versions were tested in PAMPA assay (donor, lower compartment) at a concentration of 200 µM in commercial buffer (pION) and 20% of 1-propanol as a cosolvent. The assay was performed at room temperature, with Gut-Box™ shaker, at an UWL of 25 µm and saturated relative humidity for 4 hours.

Transport rates were calculated and compared by HPLC-UV regular analysis.

Compound	Permeability /10^{-6} cm·s^{-1}	Transport (%)	Membrane retention (%)
CH_3CO-(*N*-MePhe)$_4$-NH_2	7.3 ± 0.2	13.2 ± 0.4	26.5 ± 1.8
CD_3CO-(*N*-MePhe)$_4$-NH_2	6.9 ± 0.4	12.6 ± 0.6	22.4 ± 2.3
$CD_3{}^{13}CO$-(*N*-MePhe)$_4$-NH_2	6.7 ± 0.4	12.3 ±0.7	19.3 ± 1.0
C_6H_5CO-(*N*-MePhe)$_4$-NH_2	1.4 ± 0.5	3.0 ± 1.0	83.5 ± 1.5
C_6D_5CO-(*N*-MePhe)$_4$-NH_2	1.3 ± 0.3	2.6 ± 0.6	83.0 ± 0.8
Propranolol	9.9 ± 1.0	17.1 ± 1.4	18.0 ± 2.6

Table 1.3. Study of permeability effect of *N*-terminus tags on peptide permeability and transport across PAMPA assay. Assessed parameters were permeability (P_e), percentage of transport (*%T*) and membrane retention (*%R*).

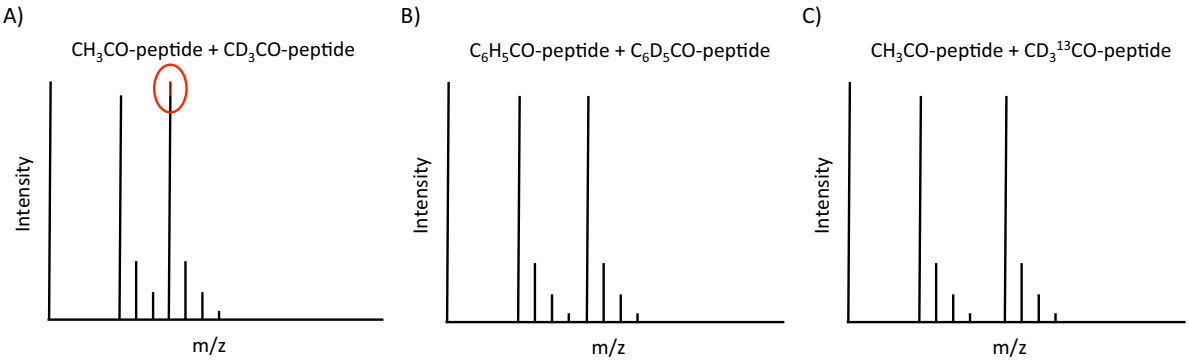


Figure 1.9. Expected pattern of molecular peaks of pair-peptides (light and heavy) on mass spectrometry. A) Molecular peak of heavy-peptide (CD_3CO-(*N*-MePhe)$_4$-NH_2; +3 a.m.u.) significantly overlaps with third isotopic peak of light-peptide discouraging its use for quantification. Higher mass differences between light and heavy peptides are required for proper quantification, such as $CD_3{}^{13}CO$-(*N*-MePhe)$_4$-NH_2; +4 a.m.u. (B) or C_6H_5CO-(*N*-MePhe)$_4$-NH_2 and C_6D_5CO-(*N*-MePhe)$_4$-NH_2; +5 a.m.u. (C).

1.5.3. MS quantification

As any other MS instruments, MALDI-TOF MS is not inherently quantitative. Actually, apparently it has some drawbacks such as its ionization method strongly depends on its shoot-to-shoot variability and co-crystallization of peptide molecules and matrix is not homogeneous on the sample spot. However the use of an isotopically labelled internal standard that behaves identically to its isotopic pair revert the situation. Indeed, all strengths of MALDI-TOF arise recommending the use of this technique for quantification of single peptides or simple mixtures in transport or internalization studies if isotopomers of tested molecules are easily available. Those MALDI-TOF strengths are speed of analysis, ease of use, relatively low cost, sensitivity, tolerance against detergents and contaminants in the sample and possibility of automation.

Use of deuterium containing isotopomers are discouraged in LC-MS or LC-MS/MS quantification in regard of reported slightly modification of chromatographic retention time in reverse phase HPLC.[34] However, since MALDI-TOF is not coupled to LC, deuterium can be safely used as no discrimination occurs between the analyte and the internal standard.

HPLC-UV detection

The area under the HPLC peak determined the amount of peptide in each compartment. Effective permeability (P_e) and percentage of transport *(%T*) were calculated with equations on section 1.4.

MALDI-TOF detection

Transport quantification of peptides across barriers such as PAMPA can be assessed both by absolute or relative quantification.

Absolute quantification relies on determining the exact amount of peptide present in the acceptor well at the end of the experiment ($t = t$) and in the donor well at the beginning of the experiment ($t = 0$). Those peptide amounts can be calculated by means of a calibration curve of ratio signals between a heavy version of the peptide and the tested peptide (light version). To build that curve, heavy peptide is spiked to the solution that has to be quantified and then analysed by MALDI-TOF for determining areas ratio. Absolute quantification of those peptide amounts requires perfectly know spiking concentration of heavy peptide solution (determined by amino acid analysis). Although this approach is effective, we argue that relative quantification has more sense because do not require amino acid analysis, which in this case do not mean any other advantage than allowing absolute quantification.

Since transport is indeed a relative measure, relative quantification specially suits on that purpose. This approach is quite similar in the sense that a calibration curve is still required. The same protocol was applied to build up the calibration curve to determine $Q_A(t)$ and $Q_D(t_0)$ by spiking the initial peptide sample ($C_{D,t0}$) and the acceptor sample at $t = t$ ($C_{A,t}$) of all tested peptides with heavy peptide spiking solution. Since those concentrations will differ by one or more fold orders, depending on transport ability of the tested molecule, ($C_{D,t0}$ >> $C_{A,t}$) heavy peptide spiking solutions must approximate $C_{D,t0}$ and $C_{A,t}$ to interpolate *heavy:light ratio signal = 1* into the calibration curve. Spiking solution prepared to determine $C_{D,t0}$ ($C_{0,heavy}$) was approximately 200 µM since this was the concentration of the assayed peptide. On the other hand, because we expect a transport around 10%, $C_{A,heavy}$ spiking solution to determine $C_{A,t}$ was prepared by 1:10 dilution of $C_{0,heavy}$. In case this solution do not interpolate heavy:light ratio = 1, another dilution would be required. Since spiking solutions are prepared by dilution, we do not introduce significant errors on solution preparations (as might happen preparing distinct solutions by weight).

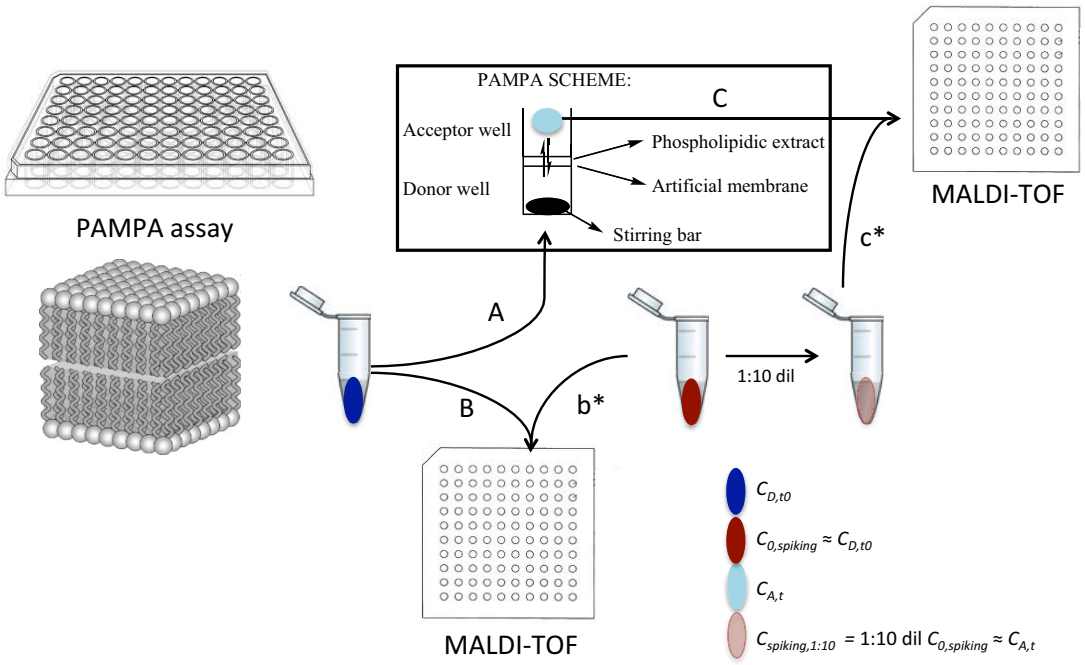


Figure 1.10. Spiking workflow for $Q_A(t)$ and $Q_D(t_0)$ quantification, B/b* and C/c* respectively. Spiking solutions of heavy peptide to determine $C_{D,t0}$ and $C_{A,t}$ must approximate the their concentration to fit in the calibration curve. Thus, $C_{A,heavy}$ was prepared by 1:10 dilution of $C_{0,heavy}$. Finally, 1 µL of spiked solution mixture was mixed with 1 µL of ACH matrix and analysed by MALDI-TOF.

To assess $Q_A(t)$, constant amounts of acceptor well final solution $C_{A,t}$ were spiked with different amounts of $C_{A,heavy}$ heavy-peptide solution with the following ratios (light:heavy, in µL): 2:1; 2:1.5; 2:2; 2:2.5; 2:3. Those solutions were mixed up and 1 µL of each was plated to the MALDI-TOF plate with 1 µL of α-cyano-4-hydroxycinnamic acid (ACH) matrix. Samples were analysed by MALDI-TOF. Peak area information was extracted and peak area ratios of heavy:light peptides was determined (calculations were performed with molecular peak, first and second isotopic peaks, but not the third one). Similarly, the very same protocol was used to determine $Q_D(t_0)$ by using $C_{0,heavy}$ as heavy peptide spiking solution.

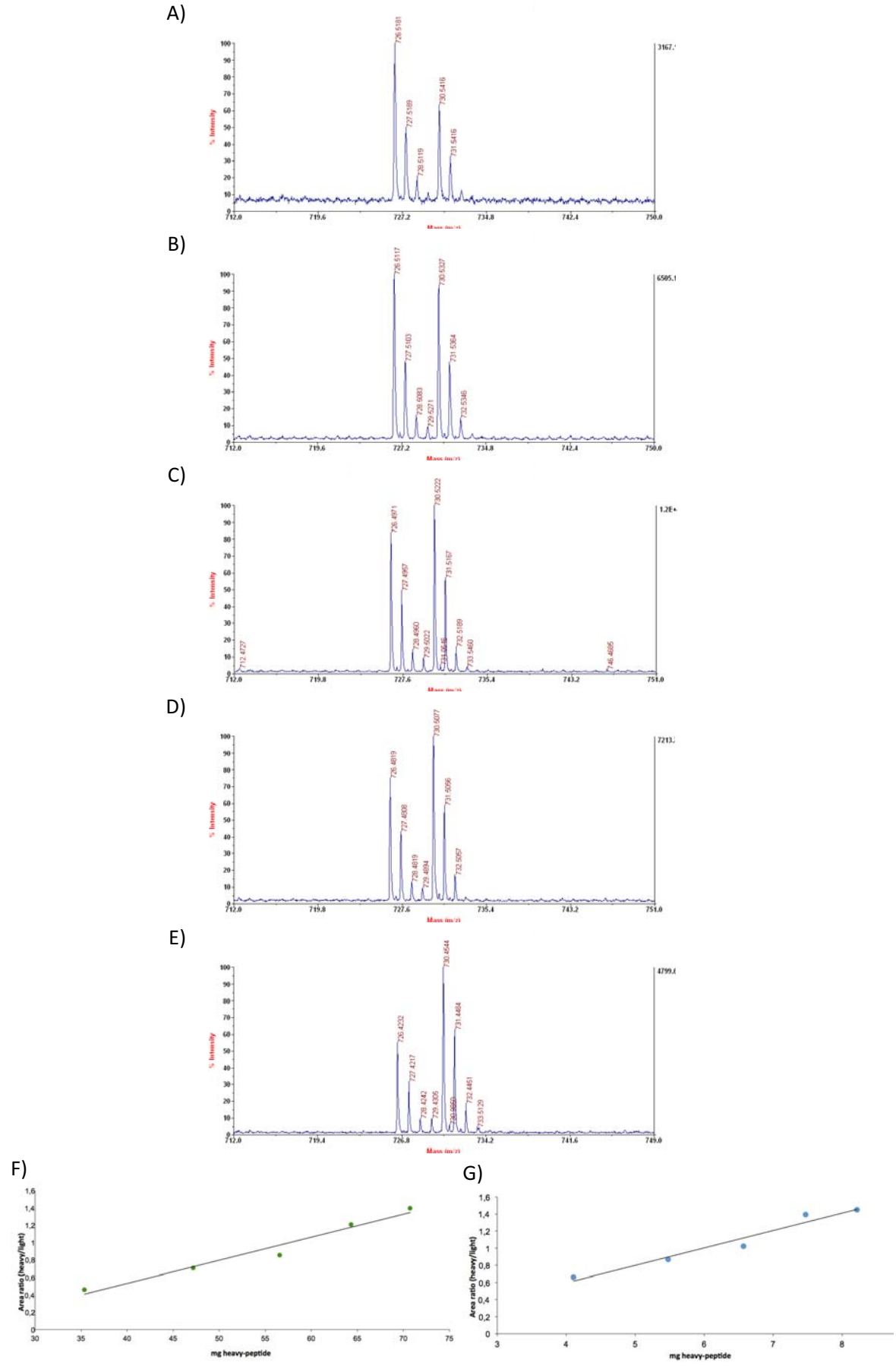


Figure 1.11. A-E) MALDI-TOF spectra series of increasing spiking amounts of heavy peptide in a PAMPA assay acceptor sample representing the following ratios (light:heavy, in µL): 2:1; 2:1.5; 2:2; 2:2.5; 2:3. F-G) Linearity curves (heavy/light areas ratio vs amount of heavy peptide) of $C_{D,t0}$ and $C_{A,t}$, respectively, with its regression lines (r = 0.98 and r = 0.97). Quantification is achieved from interpolation of this curve.

We build up a curve to assess the final amount of light version on acceptor compartment by interpolation. Ratio area of heavy:light peptides was drawn in *y*-axis whereas amount (mg, considering $C_{0,heavy}$ is 200µM) of spiked heavy peptide-version was represented in *x*-axis. Regression lines were assessed for all tested light-heavy versions and r^2 were superior than 0.95.

Percentage of transport was calculated with equations in 1.4, which was developed by using equations of regression curves from C_{t0} and C_A.

$$y_A = a_A x_A + b_A$$

$$y_0 = a_0 x_0 + b_0$$

$$\%T = \frac{Q_A(t)}{Q_0(t_0)} \cdot 100 = \frac{x_A}{x_0} \cdot 100$$

Since *b* has no means on the experiment because linearity is only checked nearby experiment concentrations and *y* = 1, formula is simplified to:

$$x_A = 1/a_A$$

$$x_0 = 1/a_0$$

$$\%T = \frac{Q_A(t)}{Q_0(t_0)} \cdot 100 = \frac{a_0}{a_A} \cdot 100$$

By this means, transport of Ac-(*N*-MePhe)$_4$-NH_2 is assessed using CD_3-^{13}CO-(*N*-MePhe)$_4$-NH_2 as internal standard.

% Transport	Quantification by HPLC-UV	Quantification by MALDI-TOF
Ac-(*N*-MePhe)$_4$-NH_2 - 1	11.3	13.1
Ac-(*N*-MePhe)$_4$-NH_2 - 2	11.6	13.0
Ac-(*N*-MePhe)$_4$-NH_2 - 3	11.0	10.7

Table 1.4. Comparison of triplicate transport quantification results of Ac-(*N*-MePhe)$_4$-NH_2 in PAMPA assay between HPLC-UV technique and MALDI-TOF-MS using CD_3-^{13}CO-(*N*-MePhe)$_4$-NH_2 as internal standard.

Quantification results from HPLC-UV and MALDI-TOF show a significant agreement that encourage us to apply this MS quantification approach to the next stages of projects described on this thesis.

Transport quantification of each acceptor well requires 5 spikes for each curve, then 10 spikes per sample quantified and its corresponding MALDI-TOF analysis. 10 spikes are time consuming when determining transport of libraries with several peptides. We intended to simplify this quantification protocol with minimum or no precision lost. In section 2.5 we assessed the linearity range in terms of heavy:light ratios and quantification limit of MALDI-TOF equipment used to perform those quantification experiments.

Bibliography

1. Witt, K.A., Gillespie, T.J., Huber, J.D., Egleton, R.D. & Davis, T.P. Peptide drug modifications to enhance bioavailability and blood-brain barrier permeability. *Peptides* **2001**, 22, 2329-2343.
2. Malakoutikhah, M., Teixido, M. & Giralt, E. Toward an optimal blood-brain barrier shuttle by synthesis and evaluation of peptide libraries. *J. Med. Chem.* **2008**, 51, 4881-4889.
3. Malakoutikhah, M., Prades, R., Teixido, M. & Giralt, E. N-methyl phenylalanine-rich peptides as highly versatile blood-brain barrier shuttles. *J. Med. Chem.* **2010**, 53, 2354-2363.
4. Teixido, M., Zurita, E., Malakoutikhah, M., Tarrago, T. & Giralt, E. Diketopiperazines as a tool for the study of transport across the blood-brain barrier (BBB) and their potential use as BBB-shuttles. *J. Am. Chem. Soc.* **2007**, 129, 11802-11813.
5. Chikhale, E.G., Ng, K.Y., Burton, P.S. & Borchardt, R.T. Hydrogen bonding potential as a determinant of the in vitro and in situ blood-brain barrier permeability of peptides. *Pharm. Res.* **1994**, 11, 412-419.
6. Arnett, E.M. & Gold, J.M. A Search for Stereospecific Interactions between Highly Purified Enantiomeric and Racemic Dipalmitoyl Phosphatidylcholines and Other Chiral Surfactants in Monolayers, Vesicles, and Gels. *J. Am. Chem. Soc.* **1982**, 104, 636-639.
7. Cruciani, O., Mannina, L., Sobolev, A.P., Cametti, C. & Segre, A. An improved NMR study of liposomes using 1-palmitoyl-2-oleoyl-sn-glycero-3-phospatidylcholine as model. *Molecules* **2006**, 11, 334-344.
8. Bombelli, C., Borocci, S., Lamanna, R., Mancini, G. & Segre, A.L. Chiral recognition of dipeptides in bio-membrane models: the role of amphiphile hydrophobic chains. *Tetrahedron-Asymmetry* **2008**, 19, 124-130.
9. Bombelli, C., Borocci, S., Lupi, F., Mancini, G., Mannina, L., Segre, A.L. & Viel, S. Chiral recognition of dipeptides in a biomembrane model. *J. Am. Chem. Soc.* **2004**, 126, 13354-13362.
10. Lopes, S.C., Fedorov, A. & Castanho, M.A. Chiral recognition of D-kyotorphin by lipidic membranes: relevance toward improved analgesic efficiency. *ChemMedChem* **2006**, 1, 723-728.
11. Witt, K.A., Slate, C.A., Egleton, R.D., Huber, J.D., Yamamura, H.I., Hruby, V.J. & Davis, T.P. Assessment of stereoselectivity of trimethylphenylalanine analogues of delta-opioid [D-Pen(2),D-Pen(5)]-enkephalin. *J. Neurochem.* **2000**, 75, 424-435.
12. Miller, S.C. & Scanlan, T.S. oNBS−SPPS: A New Method for Solid-Phase Peptide Synthesis. *J. Am. Chem. Soc.* **1998**, 120, 2690-2691.
13. Teixido, M., Albericio, F. & Giralt, E. Solid-phase synthesis and characterization of N-methyl-rich peptides. *Pept. Res.* **2005**, 65, 153-166.
14. Subiros-Funosas, R., Prohens, R., Barbas, R., El-Faham, A. & Albericio, F. Oxyma: an efficient additive for peptide synthesis to replace the benzotriazole-based HOBt and HOAt with a lower risk of explosion. *Chemistry* **2009**, 15, 9394-9403.
15. El-Faham, A., Subiros Funosas, R., Prohens, R. & Albericio, F. COMU: a safer and more effective replacement for benzotriazole-based uronium coupling reagents. *Chemistry* **2009**, 15, 9404-9416.
16. Christensen, T. Qualitative test for monitoring coupling completeness in solid-phase using chloranil. *Acta Chem. Scand. B. Org. Chem. Biochem.* **1979**, 33, 763-766.
17. Madder, A., Farcy, N., Hosten, N.G.C., De Muynck, H., De Clercq, P.J., Barry, J. & Davis, A.P. A novel sensitive colorimetric assay for visual detection of solid-phase bound amines. *Eur. J. Org. Chem.* **1999**, 1999, 2787-2791.
18. Sugano, K., Hamada, H., Machida, M., Ushio, H., Saitoh, K. & Terada, K. Optimized conditions of bio-mimetic artificial membrane permeation assay. *Int. J. Pharm.* **2001**, 228, 181-188.
19. Avdeef, A. Absorption and Drug Development: Solubility, Permeability, and Charge State. Wiley, New York; **2003**. pp. 226-227.
20. Banks, W.A. & Kastin, A.J. Peptides and the blood-brain barrier: lipophilicity as a predictor of permeability. *Brain Res. Bull.* **1985**, 15, 287-292.
21. Bowerman, C.J., Liyanage, W., Federation, A.J. & Nilsson, B.L. Tuning beta-sheet peptide self-assembly and hydrogelation behavior by modification of sequence hydrophobicity and aromaticity. *Biomacromolecules* **2011**, 12, 2735-2745.
22. Liederer, B.M., Fuchs, T., Vander Velde, D., Siahaan, T.J. & Borchardt, R.T. Effects of amino acid chirality and the chemical linker on the cell permeation characteristics of cyclic prodrugs of opioid peptides. *J. Med. Chem.* **2006**, 49, 1261-1270.

23. Sorrenti, A., Diociaiuti, M., Corvaglia, V., Chistolini, P. & Mancini, G. Chiral Recognition of Dipeptides in Langmuir Monolyers. *Tetrahedron-Asymmetry* **2009**, 20, 2737-2741.
24. Wilson, W.B., Wambua, D.M. & Chiu, N.H.L. Reduction of Internal Standard Signals in Quantitative MALDI-TOF Mass Spectrometry. *JASMI* **2012**, 2, 120-125.
25. Verdu, C.F., Gatto, J., Freuze, I., Richomme, P., Laurens, F. & Guilet, D. Comparison of Two Methods, UHPLC-UV and UHPLC-MS/MS, for the Quantification of Polyphenols in Cider Apple Juices. *Molecules* **2013**, 18, 10213-10227.
26. Pickering, M. & Brown, S. Quantification and validation of HPLC-UV and LC-MS assays for therapeutic drug monitoring of ertapenem in human plasma. *Biomed. Chromatogr.* **2013**, 27, 568-574.
27. Ong, S.E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **2005**, 1, 252-262.
28. Burlina, F., Sagan, S., Bolbach, G. & Chassaing, G. Quantification of the cellular uptake of cell-penetrating peptides by MALDI-TOF mass spectrometry. *Angew. Chem. Int. Ed.* **2005**, 44, 4244-4247.
29. Burlina, F., Sagan, S., Bolbach, G. & Chassaing, G. A direct approach to quantification of the cellular uptake of cell-penetrating peptides using MALDI-TOF mass spectrometry. *Nat. Protoc.* **2006**, 1, 200-205.
30. Aubry, S., Aussedat, B., Delaroche, D., Jiao, C.Y., Bolbach, G., Lavielle, S., Chassaing, G., Sagan, S. & Burlina, F. MALDI-TOF mass spectrometry: a powerful tool to study the internalization of cell-penetrating peptides. *Biochim. Biophys. Acta* **2010**, 1798, 2182-2189.
31. Lascoux, D., Paramelle, D., Subra, G., Heymann, M., Geourjon, C., Martinez, J. & Forest, E. Discrimination and selective snhancement of signals in the MALDI mass spectrum of a protein by combining a matrix-based label for lysine residues with a neutral matrix. *Angew. Chem. Int. Ed.* **2007**, 46, 5594-5597.
32. Paramelle, D., Subra, G., Vezenkov, L.L., Maynadier, M., André, C., Enjalbal, C., Calmès, M., Garcia, M., Martinez, J. & Amblard, M. A straightforward approach for cellular-uptake quantification. *Angew. Chem. Int. Ed.* **2010**, 49, 8240-8243.
33. Pastor, J.J., Lingard, I., Bhalay, G. & Bradley, M. Ion-extraction ladder sequencing from bead-based libraries. *J. Comb. Chem.* **2003**, 5, 85-90.
34. Zhang, R., Sioma, C.S., Wang, S. & Regnier, F.E. Fractionation of isotopically labeled peptides in quantitative proteomics. *Anal. Chem.* **2001**, 73, 5142-5149.

Chapter 2

Fine tuning of BBB-shuttle candidates to improve their efficiency and stability across *in vitro* cell-based BBB models

2.1. Precedents

As it was already discussed in the Introduction, phage display is a very powerful technique for the study of protein-protein and peptide-protein interactions. Phage display libraries nature allows screening deeply and precisely a chemical space comprising all proteinogenic amino acids as no other technique is able to do. However, hits found by phage display need a subsequent stage to fine-tune its structure to make the molecule resistant enough to proteases and/or solve solubility issues. Effective activity of peptides is a balanced equilibrium between activity, stability and solubility and when modifying molecules it is noteworthy to evaluate the overall effect.[1]

Despite many advantages offered by peptides as drug candidates, they show some drawbacks that limit their development as therapeutics. A major problem relies on their rapid *in vivo* degradation, often with half-lives on the order of minutes. Approaches to produce protease-resistant peptides are of special interest since pharmaceutical industry is willing to encompass the huge chemical space offered by peptides and derivatives.[1-4]

In our group a lot of effort have been devoted to modify peptides to increase their stability properties and thus make them suitable and appealing for drug development.

Lee *et al.* described a couple of peptides able to interact with human transferrin receptor (hTfR): H-HAIYPRH-NH_2 and H-THRPPMWSPVWP-NH_2.[5] Those peptides were found after an affinity selection protocol exploring two different commercial phage display libraries based on a combinatorial library of random 7- or 12- amino acid peptides fused to pIII minor coat protein of the filamentous coliphage M13 to find 7-mer and 12-mer candidates that could bind the hTfR expressed on the surface of CEF + hTfR cells (human transferrin receptor expressing chicken

embryo fibroblasts). Those peptides are described to interact in a region of TfR that does not overlap with the native binding site of transferrin receptor.

Some years ago, in our laboratory, Roger Prades tested (in his PhD thesis) these peptides in an *in vitro* BBB cell-based model. Transport results were promising and in the frame of his thesis, he tuned those peptides to improve the transport of the peptides by increasing their stability. He synthesized and tested the enantio and retro-enantio versions of the peptide and an *N*-alkylated version by *N*-methylation of the most labile amide groups. In general, all modifications improved peptide permeability but results were especially promising on retro-enantio versions where permeability was enhanced by 65-70%. Finally THR retro-enantio version was chosen as the best candidate and it became a BBB-shuttle involved in different studies for several applications.[6]

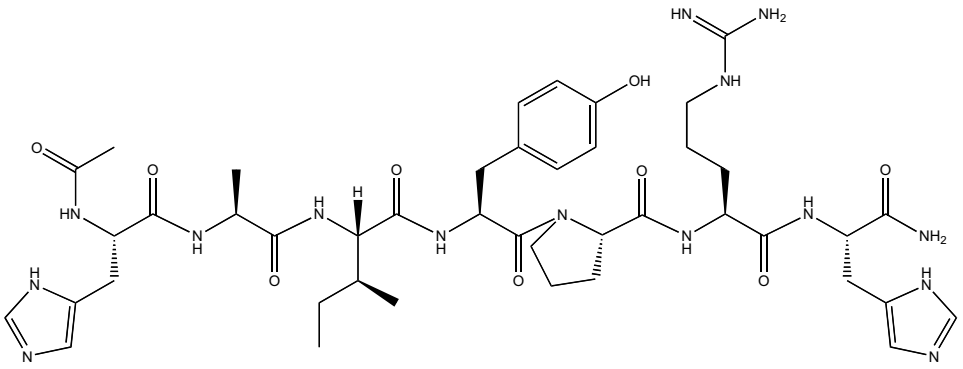


Figure 2.1. Ac-HAIYPRH-NH_2 structure.

Even though THR showed slightly better features than HAI peptide on BBB transport ability, we decided to move a step further tuning the shorter HAIYPRH to both try to improve its stability and transport ability by synthesizing a 20 HAIYPRH-analogues library. The synthesis of this library was also regarded as an opportunity to proof the MALDI-TOF transport quantification method described and set up in section 1.5 in a complete library study.

2.2. Analogue library design

Tuning phage display peptides to improve their performance is a risky task. Selected peptides from panning phage display are leads of an extremely high competitive protocol, where up to 10^8 peptides are competing in a particular assay. So that, to find a better candidate is not easy. Of course, to fine tune the lead sequence by modifying any amino acid by other proteinogenic amino acids has no sense. This selection was already explored in phage display panning rounds. Instead, the use of non-natural amino acids surveys an unexplored chemical space. Obviously this effect might be either beneficial or detrimental. In our opinion, it is not worth trying to make radical changes in such an already very specialized sequence after phage display screening. As a matter of fact, Lee *et al*. report a synthesis error when synthesizing HAIYPRH. HAIYPNH was synthesized and tested with no significant effect in competing assays with HAIYPRH-phage. This result suggests that arginine is a crucial residue for peptide activity and that simple changes can dramatically spoil the effectiveness of the molecule.[5]

In previous work in the lab, Roger Prades assessed the most protease-labile peptide positions in HAIYPRH: those being HA-I-YPR-H. On library design, we mainly modified those protease-labile positions to better address peptide stability.

To plan this library we selected a wide list of commercially available non-natural amino acids closely related to amino acids present in Ac-HAIYPRH-NH_2 peptide (both L and D versions, if available). These broad selection of commercial non-natural amino acid screen a wide chemical space by exploring side chain variations such as hydrophobicity/lipophylicity, steric volume or

length, halogenation, conformational restriction, π-π interaction by addition or removal of aromatic moieties and so on.

For synthetic reasons, the library was restricted to 21 peptides. Each analogue introduced a single amino acid substitution. A variety of non-natural amino acids enhancing the explored chemical space were considered (listed below). Library was divided in two areas, depending on the chirality of the amino acids forming the peptide backbone: L-library (10 members) and D-library (11 members). Thus, each member of L-library was based on modifying one single amino acid of Ac-HAIYPRH-NH_2 peptide. On the other hand, members from D-side library were based on the modification of a single amino acid of the HAIYPRH's retro-enantio version: Ac-hrpyiah-NH_2.

Library also contained both enantio and retro versions of Ac-HAIYPRH-NH_2. Indeed, two scrambled enantiomeric versions were designed and synthesized as negative controls: Ac-RIHHAYP-NH_2 and Ac-pyahhir-NH_2. Moreover, analogue 2 corresponding to L-arginine substitution by a L-lysine was included as a negative control analogue since Arg residue was described to play a crucial role on peptide activity as stated above.

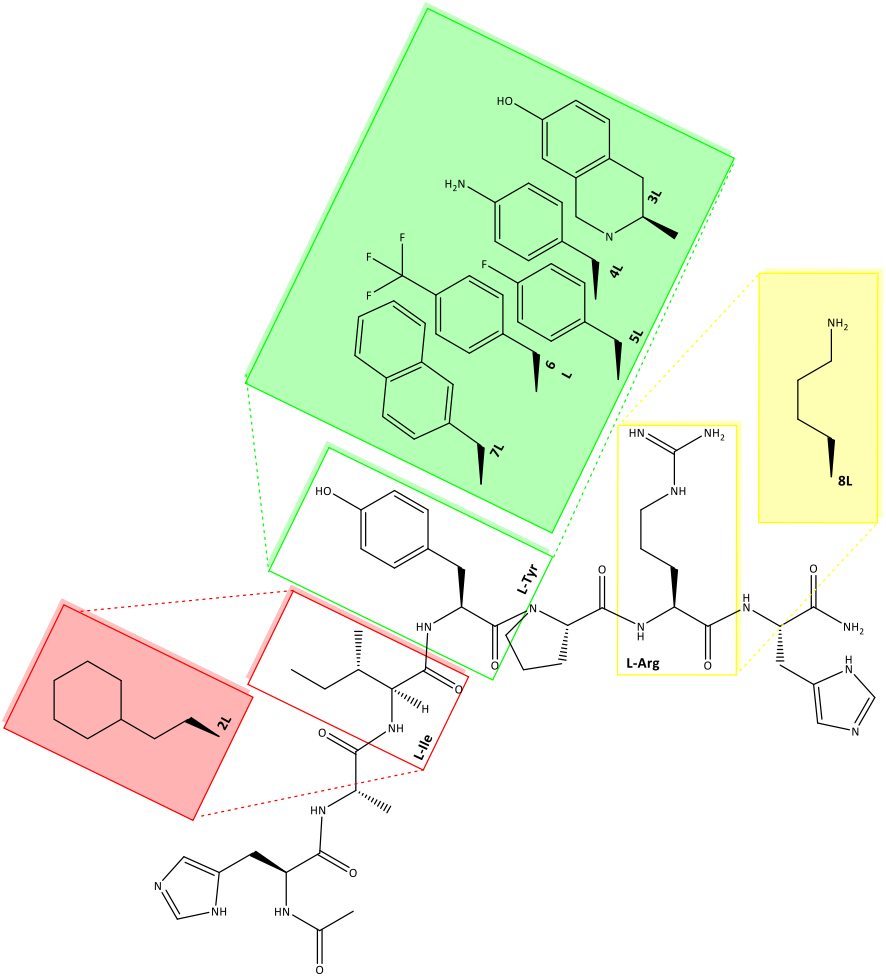
All peptides in the library were amidated in their *C*-terminus and acetylated in their *N*-terminus. The acetyl tag differs from the reported by Dr. Roger Prades work and act both as a small cargo and also allows to easily generate an internal standard of each peptide for further quantification by mass spectrometry (as previously discussed in section 1.5.2).

Peptide library members are listed below and structural changes can be pictured in Figure 2.2.

Peptide ID	Substituted amino acid	Substitution
1L	Ac-HAIYPRH-NH_2	-
2L	L-Ile	homocyclohexyl- L-alanine
3L	L-Tyr	7-hydroxy-(S)-1,2,3,4-tetraydroisoquinoline-3-carboxylic acid
4L	L-Tyr	4-amino-L-phenylalanine
5L	L-Tyr	4-fluoro-L-phenylalanine
6L	L-Tyr	4-trifluoromethyl-L-phenylalanine
7L	L-Tyr	L-2-naphthylalanine
8L	L-Arg	L-lysine
9L	Scrambled	Ac-RIHHAYP-NH_2
11L	Retro version	Ac-HRPYIAH-NH_2
1D	Retro-enantio version	Ac-hrpyiah-NH_2
2+D	D-Pro	(2R, 3R)-3-phenylpirrolidine-2-carboxilic acid
2-D	D-Pro	(2S, 3S)-3-phenylpirrolidine-2-carboxilic acid
3D	D-Pro	D-pipecolic acid
4D	D-His (*N*-terminus)	4-thiazoylalanine
5D	D-Tyr	7-hydroxy-(R)-1,2,3,4-tetrahydroisoquinoline-3-carboxilic acid
6D	D-Tyr	4-trifluoromethyl-D-phenylalanine
7D	D-Tyr	D-2-naphthyalanine
8D	D-His (*C*-terminus)	4-thiazoylalanine
9D	Scrambled	Ac-pyahhir-NH_2
11D	Enantio version	Ac-haiyprh-NH_2

Table 2.1. Amino acids substituted from the original structure to create the analogue HAIYPRH library.

A)



B)

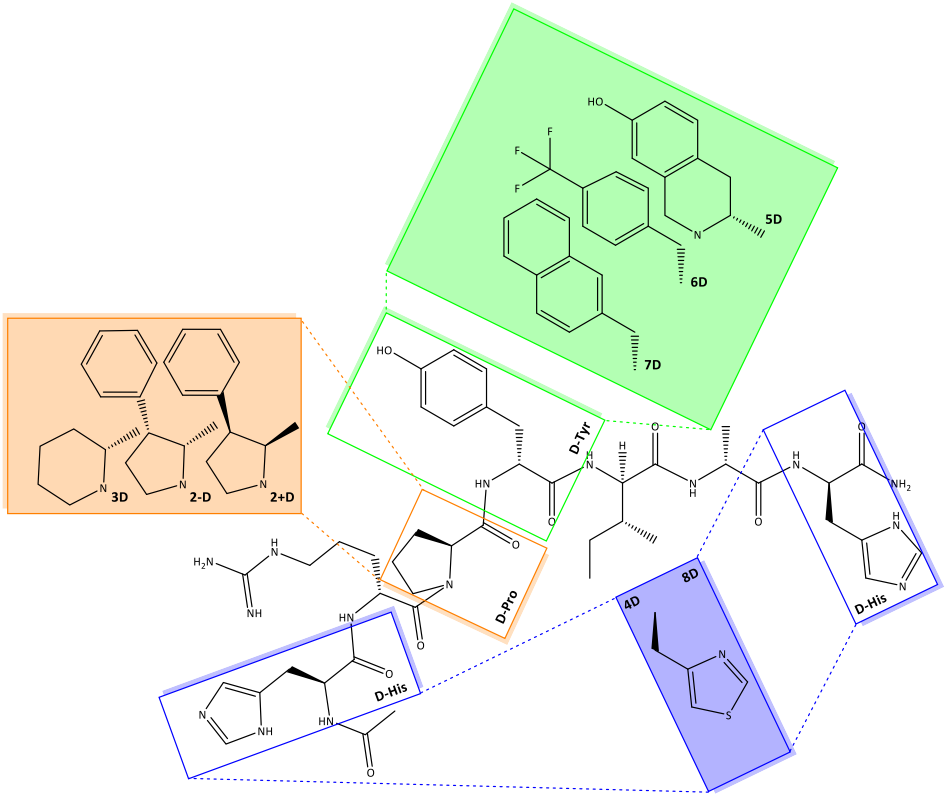


Figure 2.2. On this figure, side chain structures of substituted amino acids are shown. L-library (A) and D-library (B). Figure from Pol Arranz.

2.3. Peptide synthesis

Peptides were synthesized by solid-phase peptide synthesis with Fmoc/*t*Bu strategy. In order to obtain *C*-terminal amide peptides Rink AM resin was chosen. The synthesis was performed at a 200 µmol scale. Oxima®, COMU and DIEA were used for coupling reactions. The coupling protocol was applied twice in case coupling was performed onto a secondary amine. All peptides were acetylated in *N*-terminus. Coupling reactions were monitored by the ninhydrin Kaiser test,[7] the chloranil test[8] or the *p*-nitrophenyl test (De Clercq test).[9] In order to create two versions of each peptide the resin was split and acetylation was performed with acetic acid (for the light version) and with isotopically labelled acetic acid ($CD_3{}^{13}COOD$, for the heavy version). DIPCDI was used as coupling reagent. A preactivation step was required to allow anhydride formation.

As reported in 1.5.2, heavy versions are 4 atomic mass units heavier than the light peptides avoiding molecular peaks and their isotopic envelopes overlap on MALDI-TOF spectra.

Peptide coupling and deprotection steps were followed by either the Kaiser colorimetric test[7] or the De Clercq or chloranil colorimetric assay depending on the type of amine (primary or secondary, respectively).

Cleavage of the peptide from the resin and complete removal of side chain protecting groups was accomplished by acidolytic treatment with a TFA solution containing appropriate scavengers: TFA/TIS/H_2O (95:2.5:2.5). Crude products were analysed by reverse phase high-performance liquid chromatography (RP-HPLC), coupled either to a UV or mass spectrometry (MS) detector and by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF). Peptides were purified by RP-HPLC at semi-preparative scale and the pure product was fully characterized by RP-HPLC, HPLC-MS and MALDI-TOF. Altogether, all 42 synthesized peptides were obtained with purity higher than 95%.

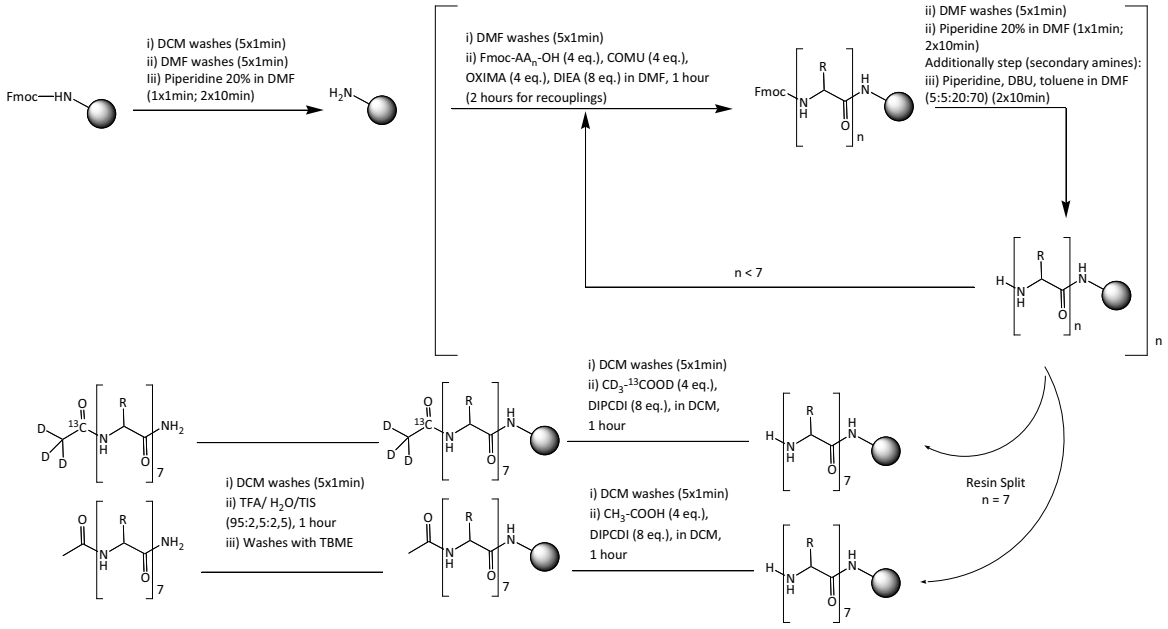
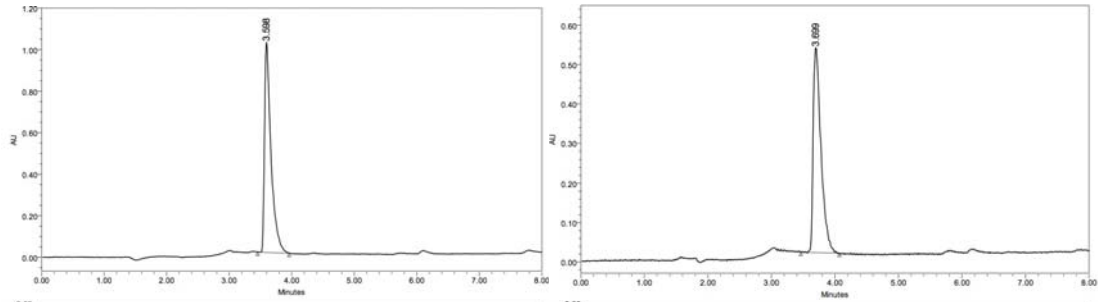
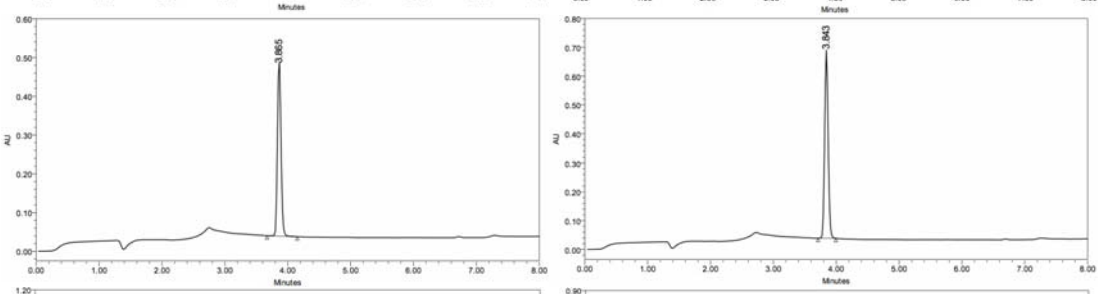


Figure 2.3. Synthetic scheme for HAIYPRH library synthesis.

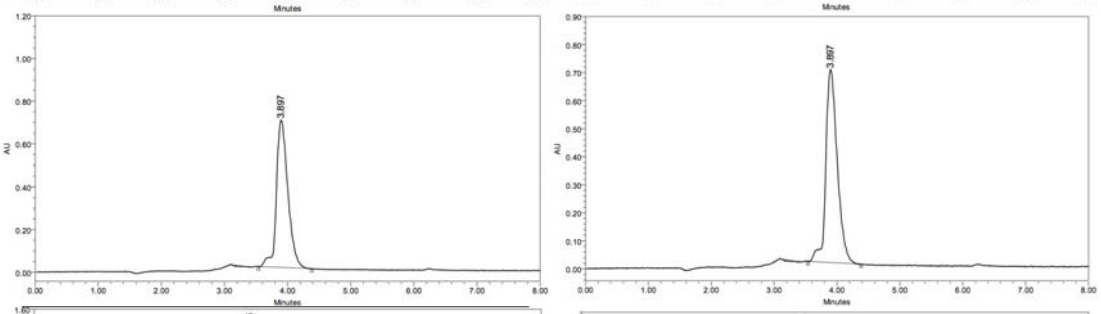
1L



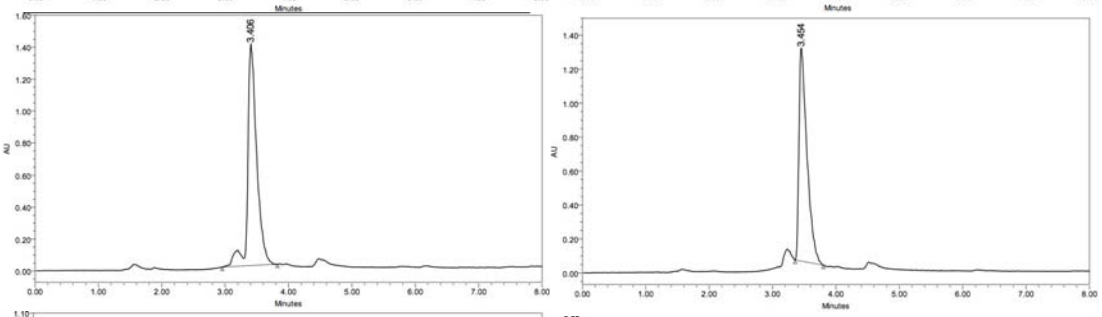
2L



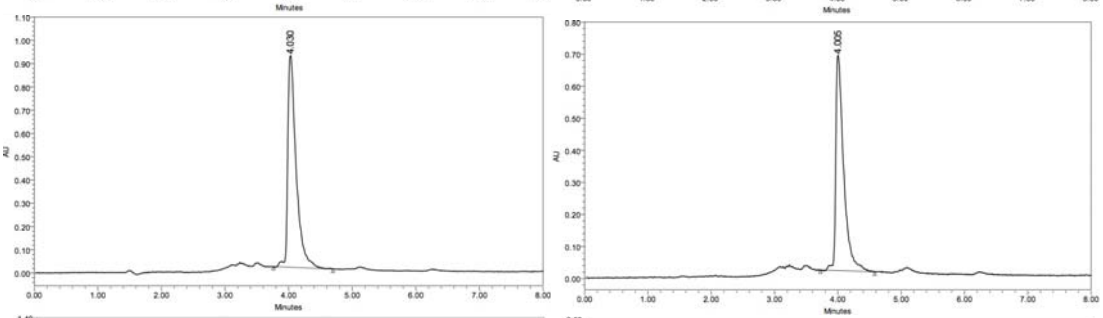
3L



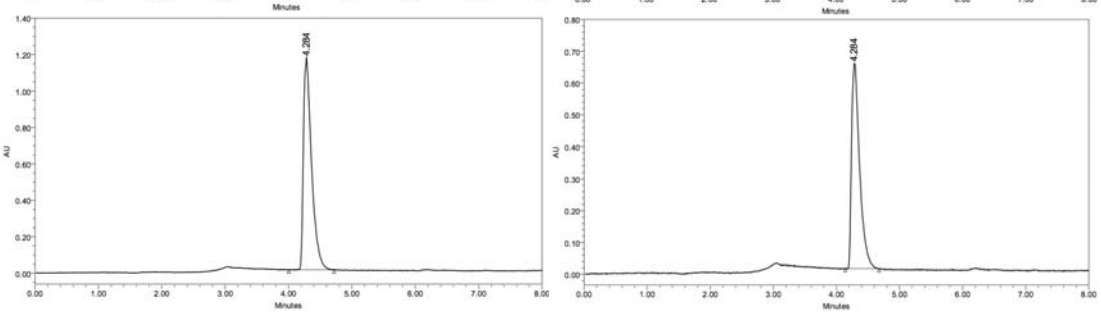
4L



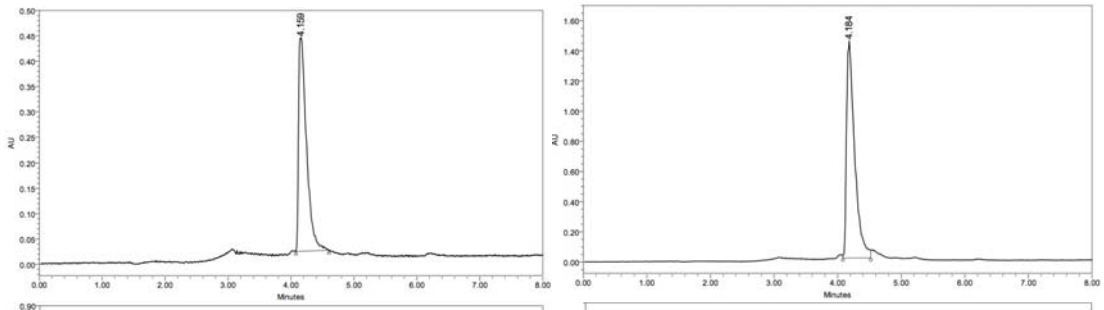
5L



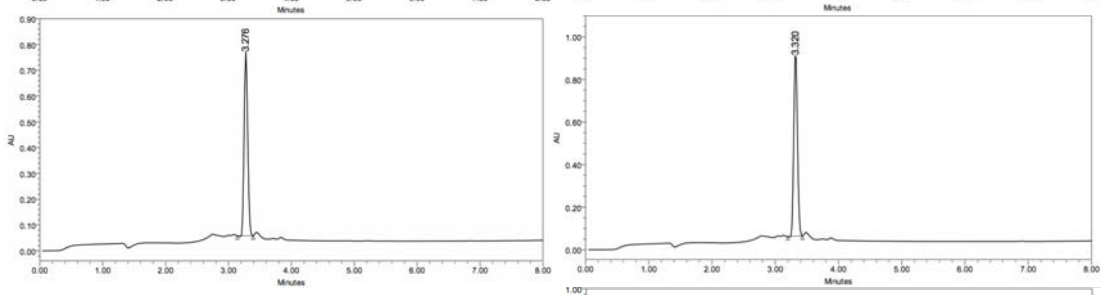
6L



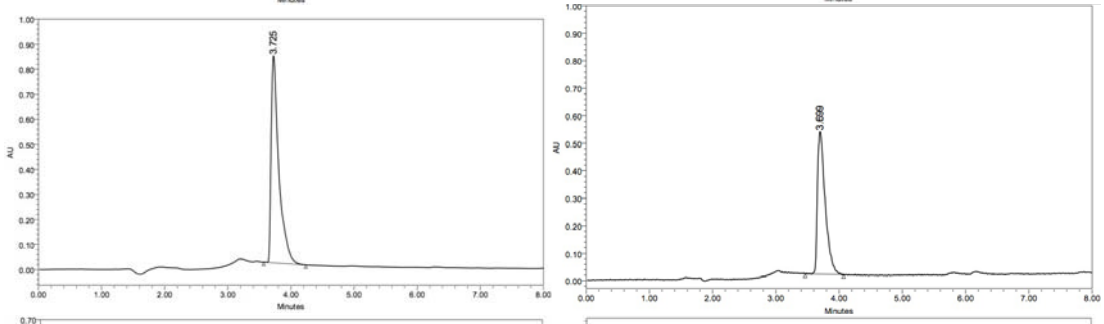
7L



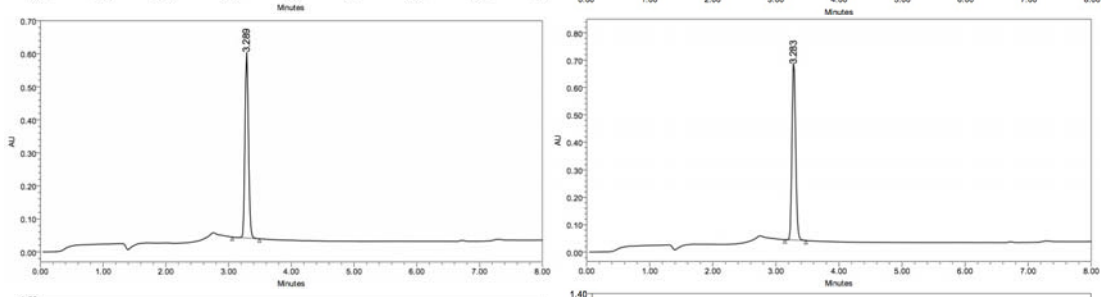
8L



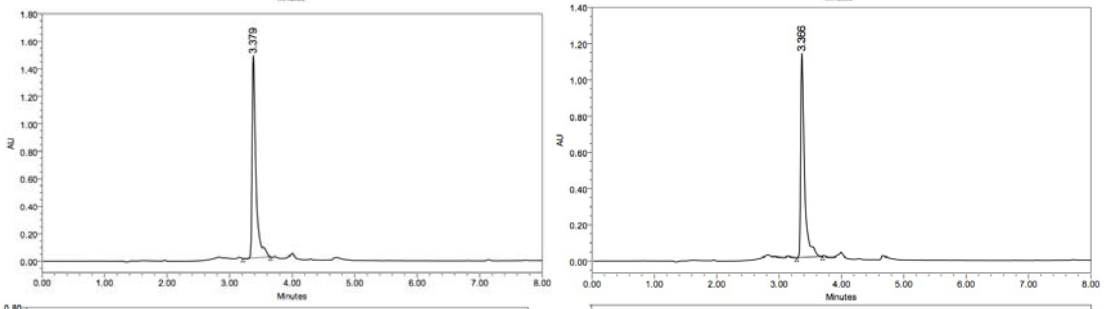
9L



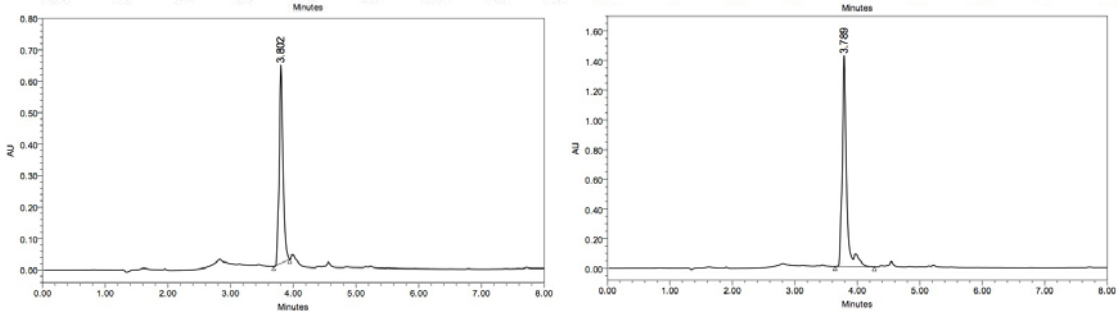
11L



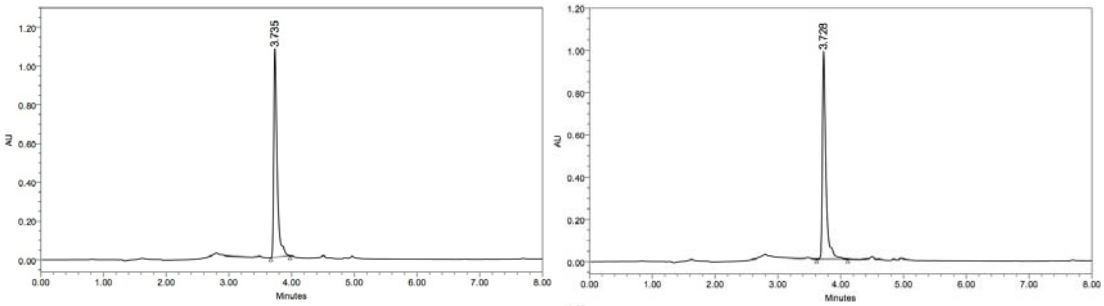
1D



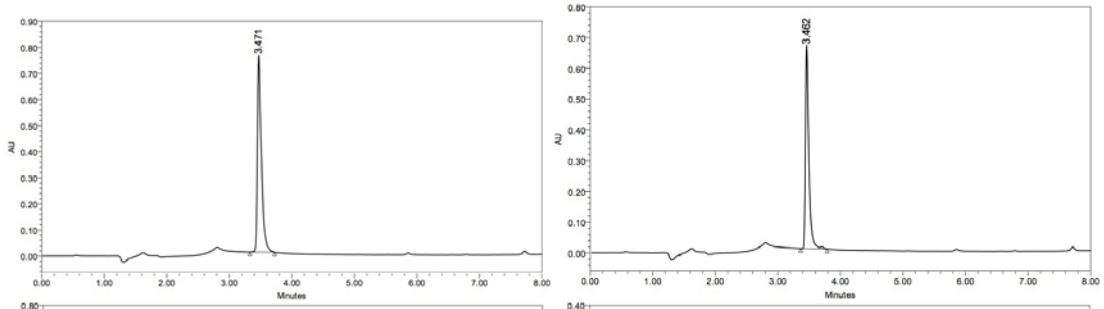
2+D



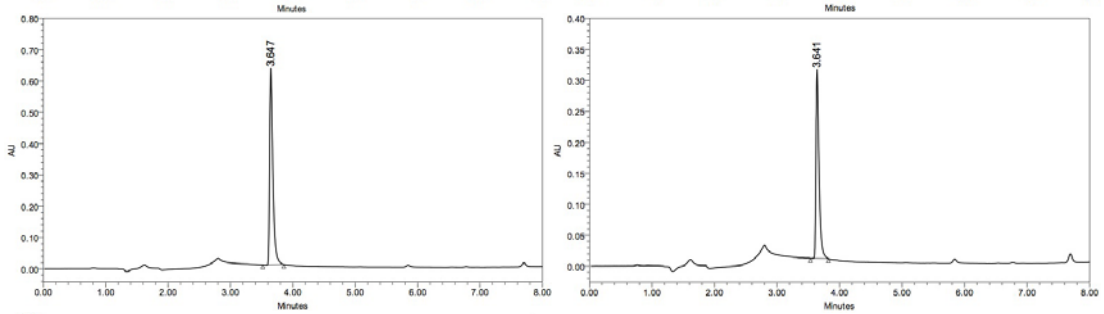
2-D



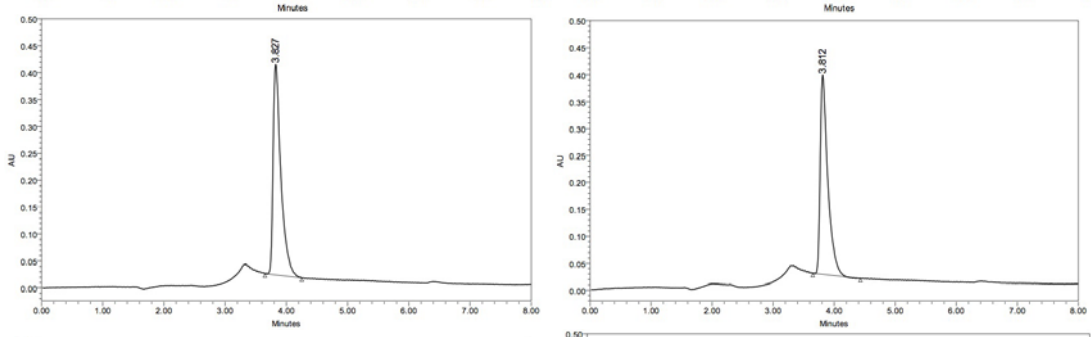
3D



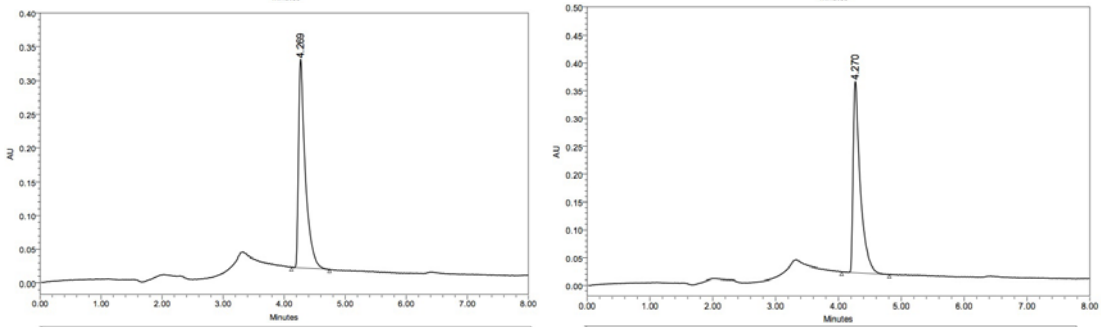
4D



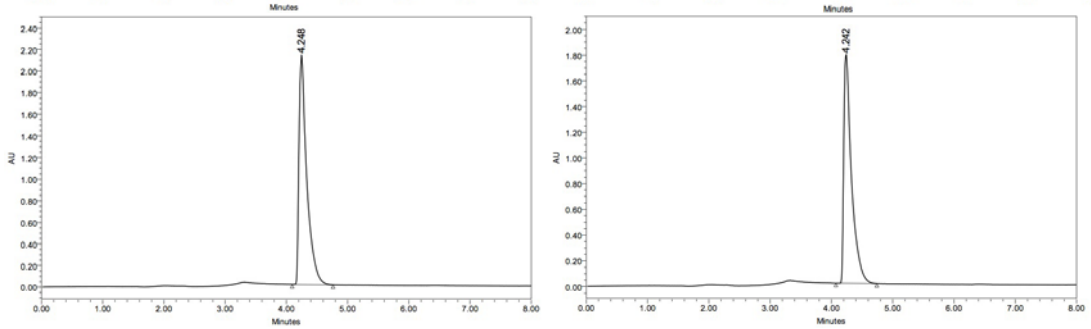
5D



6D



7D



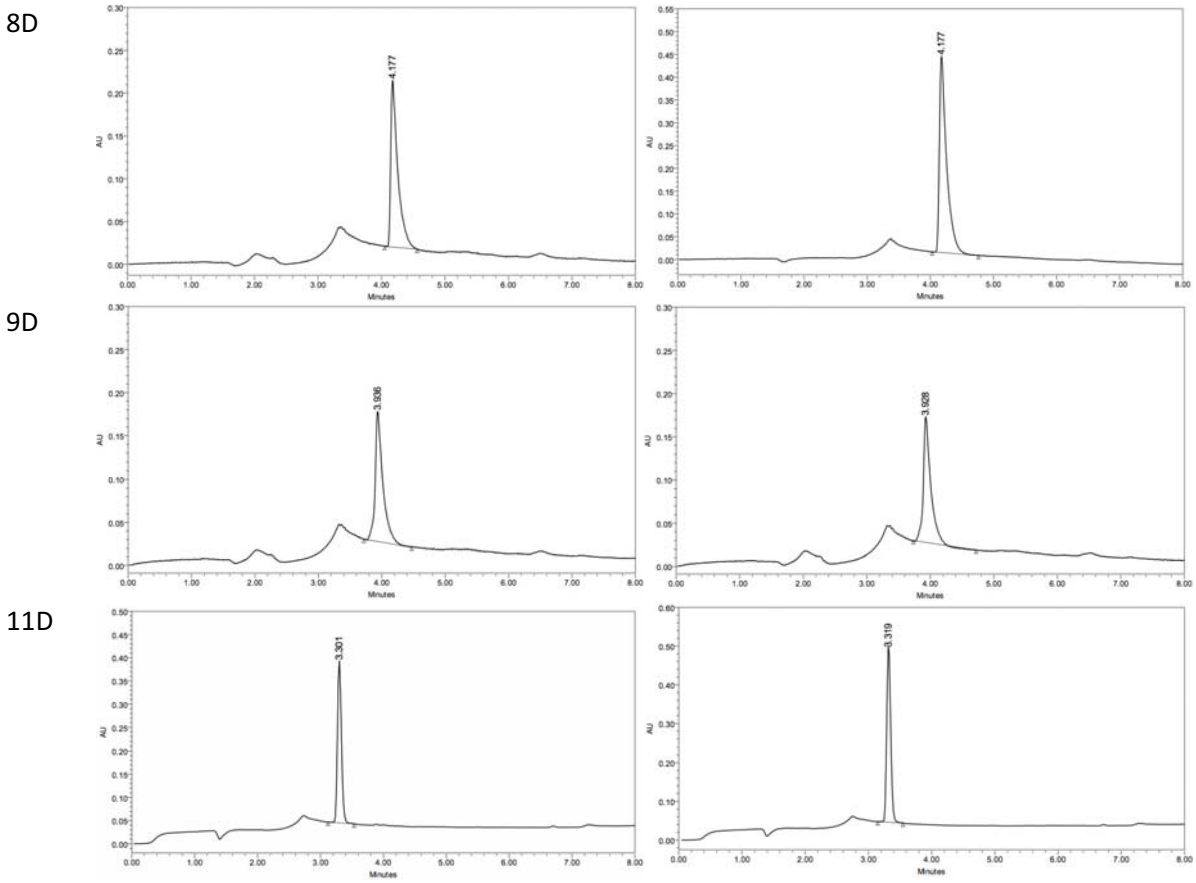


Figure 2.4. HPLC chromatograms of synthesized analogues of HAIYPRH library obtained in purities higher than 95% in their acetylated versions (left) and isotopically acetylated ($CD_3{}^{13}CO$-) versions (right).

2.4. Transport experiments: *in vitro* cell-based BBB model

As discussed in the Introduction, *in vitro* cell-based BBB models are still far from fully mimicking the *in vivo* BBB. Several models based on cell co-cultures had been proposed. These models show a significant correlation between *in vitro* and *in vivo* BBB permeability.

Several combinations of endothelial cells and astrocytes from distinct animal origins had been proposed. Our model approach is based on a co-culture of bovine brain endothelial cells and rat astrocytes, being one of the most accepted and validated BBB model.[10, 11] Although isolation and maintenance of primary cultures is not a trivial task we considered it is worthy since cultured cells have an original BBB phenotype and, under adequate conditions, they are able to express the main features of the BBB such as specific markers, tight junctions formation and cell polarization.

In vitro cell models are a simplified version of BBB, so that, some specific features cannot be completely achieved thus provoking some drawbacks on BBB-resemblance. *In vivo* BBB endothelial cells are influenced by many other stimuli than those present *in vitro* due to astrocyte presence. Those stimuli can be both mechanical and biochemical. Then, changes in performance are expected between *in vitro* and *in vivo* BBB.

Even batch-to-batch endothelial cells and rat astrocytes used to build up the cellular model or cell pass, strongly influence on final *in vitro* BBB features. Tight junction formation, variety and amount of expressed receptors can influence the transport assay performance. Important influences can range from handling to external factors that can disturb cell co-culture.

Altogether, it is of paramount importance to have some tools to evaluate, predict and diagnose the maturity of *in vitro* BBB to validate both the quality of batch-to-batch of the *in vitro* model and to compare our model with those previously described in the bibliography.

We used two tools to address this purpose. Transendothelial electrical resistance (TEER) measurement is non-destructive and allows monitoring tightness of the cell monolayer during maturity process. Since TEER measurement involve introducing an electrode into the transwell, it has some drawbacks such as risk to disturb or break the cell layer and introduction of contaminants that would spoil the correct performance of the model. Usually accepted TEER measures range from 100 to 400 $\Omega \cdot cm^2$. *In vivo* TEER is estimated to be around 1800-2000 $\Omega \cdot cm^2$.[12]

On the other hand, Lucifer Yellow (LY) is usually employed as an integrity marker of *in vitro* cellular models. It is known that crosses the BBB by paracellular pathway and its passage is considered to be below $1.7 \cdot 10^{-5}$ cm/s in well-performing BBB cellular barriers.[13] If LY transport is higher than this value, results cannot be considered valid. LY is co-incubated with tested peptide or mixture of peptides to validate each transwell assay. This tool only looks for barrier integrity during the assay and cannot give information on barrier maturity. LY transport also offers information on peptide BBB-cytotoxicity.

A) B)

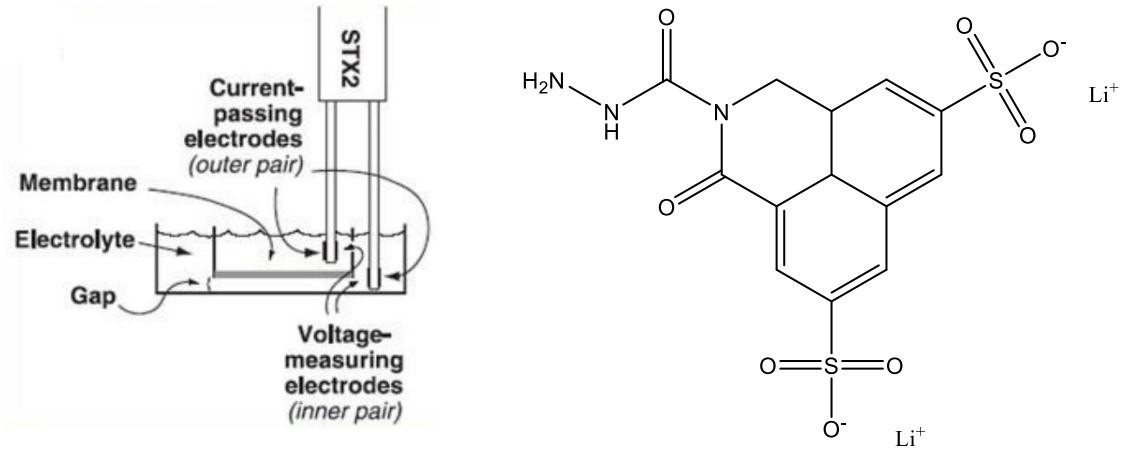


Figure 2.5. A) TEER measurement scheme. Chopstick electrode is introduced in the transwell system to measure the resistance through the cell monolayer. A trasnwell blank sample not containing the cell monolayer is required for accurate measurement; B) and Lucifer Yellow molecule.

Previous work in the lab, performed by Dr. Roger Prades on his thesis project, focused on setting up the cell-based BBB *in vitro* model. He established a model based on a co-culture of bovine brain endothelial cells and rat astrocytes. Despite failing initial attempts to exactly reproduce the model described by Gaillard and de Boer,[11] he managed to establish a well-performing model by considering other critical issues stated by experts on the field.[10, 14] A combination of parameters were screened to fully check model performance. Immunochemistry analysis to assess the presence of membrane proteins responsible for tight junctions such as ZO-1 and occludin and also transporters such as transferrin receptor or P-gp efflux pump were performed at that time. Those were considered fingerprint membrane proteins defining the BBB, thus showing evidence of correct BBB-features displayed by the *in vitro* cell-based BBB assay.[15]

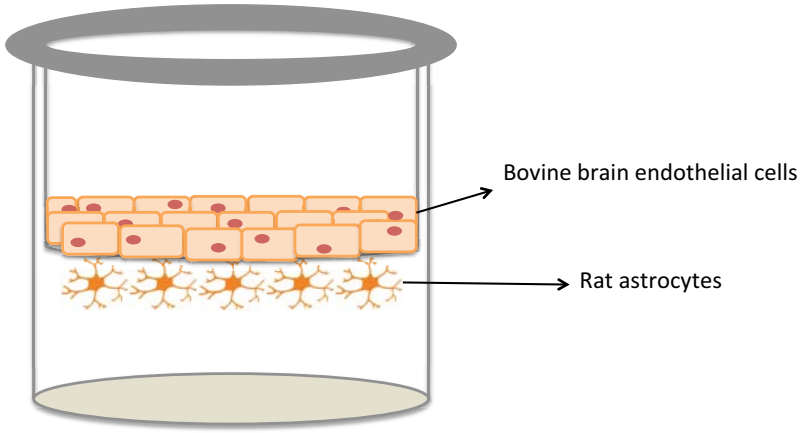


Figure 2.6. Scheme of *in vitro* cell-based BBB assay.

Inheritance of this valuable tool for BBB-related projects in the lab has required significant amount of work to learn the basic technical mechanisms and biological basis to constantly and robustly set up the model for required experiments. During the present thesis we faced significant trouble on regularly establish the model and continued support from Dr. Roger Prades and Benjamí Oller has been essential to eventually perform satisfactory *in vitro* cell-based BBB model experiments.

Peptides were evaluated at a concentration of 200 µM for 2 h at 37º°C in Ringer/HEPES buffer. They were co-incubated with LY at a concentration of 20 µM to ensure the integrity of the cell membrane during the assay. Previous to the assay, TEER measures of cell membrane were 116 ± 22 Ω·cm^2. Apparent permeability (P_{app}) and percentage of transport (*%T*) were determined by means of both RP-HPLC and MALDI-TOF as discussed in next section.

2.5. Quantification

Quantification of peptide library was assessed both by HPLC-UV and MALDI-TOF. Samples were injected to RP-HPLC with the same adjusted gradient for each peptide. Percentage of transport was calculated by applying the following formula,

$$\%T = \frac{Q_A(t)}{Q_D(t_0)} \cdot 100$$

where $Q_A(t)$ is the compound amount at the acceptor (baso-lateral) well at time *t* (area under the HPLC peak), and $Q_D(t_0)$ is the compound amount in the donor (apical) well at *t = 0* (area under the HPLC chromatographic peak).

Similarly, apparent permeability (P_{app}) was calculated as follows,

$$P_{app} = \frac{\partial Q}{\partial t} \cdot \frac{1}{A \cdot C_0}$$

where *d*Q/*d*t is the transport rate of the compound (mol/s), *A* is the area of the cell monolayers (cm^2) and C_0 is the initial donor concentration (mol/L, M). P_{app} units are cm/s.

In contrast to percentage of transport value *(%T)*, apparent permeability (P_{app}) considers the area of the well membrane and the initial concentration of the assayed sample. Since in the present work all peptides were tested at the same concentration and with the same model assay both measures are equally valid. However, in case we would like to compare data from distinct models or assess transport at distinct initial concentrations only P_{app} should be considered.[16]

Compound	%T	P_{app} / 10^{-6} cm/s
1L	2.9 ± 1.4	2.5 ± 1.2
2L	3.9 ± 0.9	3.3 ± 0.8
3L	2.0 ± 0.7	1.7 ± 0.6
4L	2.0 ± 0.9	1.7 ± 0.8
5L	3.4 ± 1.6	2.9 ± 1.4
6L	2.3 ± 1.4	1.9 ± 1.2
7L	3.8 ± 0.8	3.2 ± 0.7
8L	5.5 ± 1.2	4.6 ± 1.0
9L	2.4 ± 1.2	2.0 ± 1.0
11L	5.3 ± 2.6	4.4 ± 2.2
1D	5.3 ± 1.8	4.5 ± 1.5
2+D	5.7 ± 1.8	4.8 ± 1.5
2-D	6.8 ± 1.2	5.7 ± 1.0
3D	7.2 ± 0.9	6.1 ± 0.8
4D	4.4 ± 0.5	3.7 ± 0.4
5D	5.0 ± 2.3	4.2 ± 1.9
6D	5.3 ± 2.5	4.4 ± 2.1
7D	4.5 ± 2.0	3.8 ± 1.7
8D	4.8 ± 3.4	4.1 ± 2.9
9D	4.8 ± 1.9	4.0 ± 1.6
11D	4.6 ± 1.9	3.9 ± 1.3

Table 2.2. Transport results measured by HPLC-UV method on the basis of area under chromatographic peaks.

As discussed in section 1.5, MALDI-TOF relative quantification is the most convenient methodology for determining transport across *in vitro* barriers with MS techniques. Samples are spiked with different amounts of heavy version peptides. If quantification of both $Q_{D,t0}$ and $Q_{A,t}$ is determined with the same spiking solution (only dilutions applied), there is no need to perfectly know the exact amount of internal standard. The area of all isotope peaks of the light and heavy compounds corresponding to $[M+H^+]$, $[M+Na^+]$ and $[M+K^+]$ were used for quantification.

However, a calibration curve as described in section 1.5 should be assessed and experimental points fit within this curve. After sample spiking, quantification relies on ratio value between signals from isotopically labelled and non-modified molecules (quantification ratio).

To simplify MS transport evaluation, we intended to skip calibration curve for routine transport estimation. Building a curve requires 10 spikes and this can be a tedious task when analysing several peptides within a library. Hence, we suggested using a simplified method for high-throughput transport estimation to rank all analogue candidates in the library. Then, transport quantification of the best candidates would be confirmed by using the best-accepted calibration curve spiking method.

Instead of spiking each sample with 5 amounts of heavy peptide to build a whole calibration curve, this simplified quantification method consist on a single spiking of heavy peptide to each sample. Since detector linearity on heavy-light area ratios is not ensured without calibration curve, we assessed the linear range and the limit of quantification of two instruments available (MALDI-4800, *Applied Biosystem 4800*; MALDI-4700, *Applied Biosystem 4700*) for two peptides (11L and 4D) out of 21 in the library. Although each peptide could perform distinctly in MS due to their ionization capabilities and/or detection, we assumed all peptides within the library would perform similarly than those chosen as model peptides.

This allowed us to work with a single spike per sample instead of building up a complete calibration curve for each one, with minimum precision lost. Experimental errors in different steps were determined.

2.5.1. Limit of quantification

Limit of quantification of each instrument for 11L and 4D peptides was determined by mixing pair-peptides in 1:2 proportion (light:heavy) with the assay starting concentration (light, 200 µM; heavy, 400 µM). Ratio between pair-peptides signal was determined and continuous 1:3 dilutions were performed. We determined the error on area ratios (A_{heavy}/A_{light}) throughout dilutions. We observed that this parameter strongly depends on the equipment used. MALDI-4800 is able to detect lower concentrations than MALDI-4700 and exhibits much better sensitivity and signal/noise ratio. Moreover, all peaks of isotopic envelopes of analysed species are clearly detected with MALDI-4800, while MALDI-4700 shows poor sensitivity on early dilutions.

Dilution	MALDI-4800		MALDI-4700	
	11L	4D	11L	4D
Light, 200 µM : heavy, 400 µM	-	-	-	-
Dilution 1 (1:3)	3.8	2.5	3.0	0.7
Dilution 2 (1:3, dilution 1)	4.4	5.0	6.1	0.7
Dilution 3 (1:3, dilution 2)	0.3	2.0	4.4	0.8
Dilution 4 (1:3, dilution 3)	1.6	3.9	6.0	0.3
Dilution 5 (1:3, dilution 4)	4.3	7.8	7.9	3.4
Dilution 6 (1:3, dilution 5)	3.0	0.8	44.2	6.8
Dilution 7 (1:3, dilution 6)	3.2	6.3	27.0	20.4
Dilution 8 (1:3, dilution 7)	3.1	2.0	67.1	39.8
Dilution 9 (1:3, dilution 8)	11.7	0.4	ND	ND
Dilution 10 (1:3, dilution 9)	8.5	26.1	ND	ND
Dilution 11 (1:3, dilution 10)	15.6	55.4	ND	ND

Table 2.3. Limit of quantification comparison between MALDI-4800 and MALDI-4700 instruments by testing light:heavy mixtures at starting concentration of: light, 200 µM and heavy, 400 µM; and subsequent dilutions (1:3) for analogues 11L and 4D. Errors were always calculated respect the initial solution and expressed as % of error. ND: non-detected.

2.5.2. Linear range assessment

As already stated above, peptide quantification by a single spike could be unreliable due to non-linear performance of MS detector if peaks of pair-peptides (light-heavy) have large size variation. To explore signal ratio limits for feasible detection we performed a series of test mixtures of light-heavy peptides with a variety of proportions.

Ratio detection limits will depend on the instrument used and amounts of detected peptides. So that, we postulated different conditions closely mimicking those that can be found on peptide transport experiments. Transport values are expected to fit within 0.1% and 10% transport range. Transports lower than 0.1% can be dismissed since they will not have any biological relevance, whereas transports higher than 10% are unexpected. Since initial concentration of tested peptides was around 200 µM, we assessed ratio linear range for peptides 11L and 4D at three different concentrations: 20 µM, 2 µM and 0.2 µM corresponding to 10%, 1% and 0.1% transport. Each concentration was independently spiked (equal volumes) with heavy peptide at a concentration of 20 µM, 6.67 µM, 2 µM, 0.67 µM and 0.2 µM. Since limit of quantification was not a concern at these concentrations we hypothesize there was no error on the measure when spiking the light peptide sample with the same amount of heavy peptide. Thus, we assessed the error of quantification of each light peptide sample using all other heavy peptide spiking concentrations prepared (compared to 20:20, 2:2 and 0.2:0.2; light:heavy concentration, µM, ratios).

MALDI-4800 (*Applied Biosystem 4800*):

Peptide 11L		Light peptide concentration/µM		
		20	2	0.2
Heavy peptide concentration/ µM	20	0	13.6	62.7
	6.67	11.2	11.9	53.6
	2	13.3	0	20.9
	0.67	29.8	1.1	6.6
	0.2	113.6	1.1	0

Peptide 4D		Light peptide concentration/µM		
		20	2	0.2
Heavy peptide concentration/ µM	20	0	62.3	76.9
	6.67	24.8	10.5	37.2
	2	38.8	0	30.8
	0.67	36.2	27.3	6.5
	0.2	353.2	59.2	0

MALDI-4700 (*Applied Biosystem 4700*):

Peptide 11L*		Light peptide concentration/µM		
		20	2	0.2
Heavy peptide concentration/ µM	20	0	21.4	0.1
	6.67	12.2	25.1	19.1
	2	9.4	0	67.9
	0.67	4.5	0.9	12.4
	0.2	24.7	9.7	0

Peptide 4D*		Light peptide concentration/µM		
		20	2	0.2
Heavy peptide concentration/ µM	20	0	13.2	14.5
	6.67	14.1	3.4	38.5
	2	0.3	0	17.5
	0.67	12.9	29.9	24.1
	0.2	82.5	2.0	0

Table 2.4. Linear range assessment results for peptides 11L and 4D in MALDI-4800 and MALDI-4700 instruments. It was hypothesized that there was no error on the measure when spiking the light peptide sample with the same amount of heavy peptide. Errors on quantification due to distinct peak size were assessed by mixing equal amounts of distinct light and heavy peptide concentrations. Errors are expressed in percentage (%) compared to mixtures of pair-peptides with equal concentrations. *Spectra from MALDI-4700 instrument suffered from low sensitivity showing poor signal/noise ratios. In most cases isotopic envelope was incomplete.

We conclude that concentrations of spiking peptide (heavy) one order of magnitude lower or higher than light peptide sample lead to acceptable quantification results. Nonetheless, we strongly suggest working with peptide amounts leading to signals ratios 1:3 or closer for better identification precision.

Although results from linear range assessment experiments show that MALDI-4700 seems to perform slightly better than MALDI-4800, MALDI-4700 spectra show low sensitivity and poor signal/noise ratio. In most cases, isotopic envelope was incomplete. Therefore, limit of quantification assessment experiments (see 2.5.1) confirmed that MALDI-4800 has far more sensitivity, which appeared to be crucial on faithfully determining signal areas on quantification experiments. Altogether, we decided to use MALDI-4800 for library transport quantification.

2.5.3. Library transport quantification

Quantification was determined by assessing the amount of peptide present in the initial peptide sample and the amount of peptide present on the acceptor compartment at *t = t* by means of:

$$\%T = \frac{Q_A(t)}{Q_D(t_0)} \cdot 100$$

$Q_A(t)$ and $Q_D(t_0)$ were assessed by spiking the initial peptide sample ($C_{D,t0}$) and the acceptor sample at *t = t* ($C_{A,t}$) of all tested peptides with heavy peptide spiking solution following the scheme below.

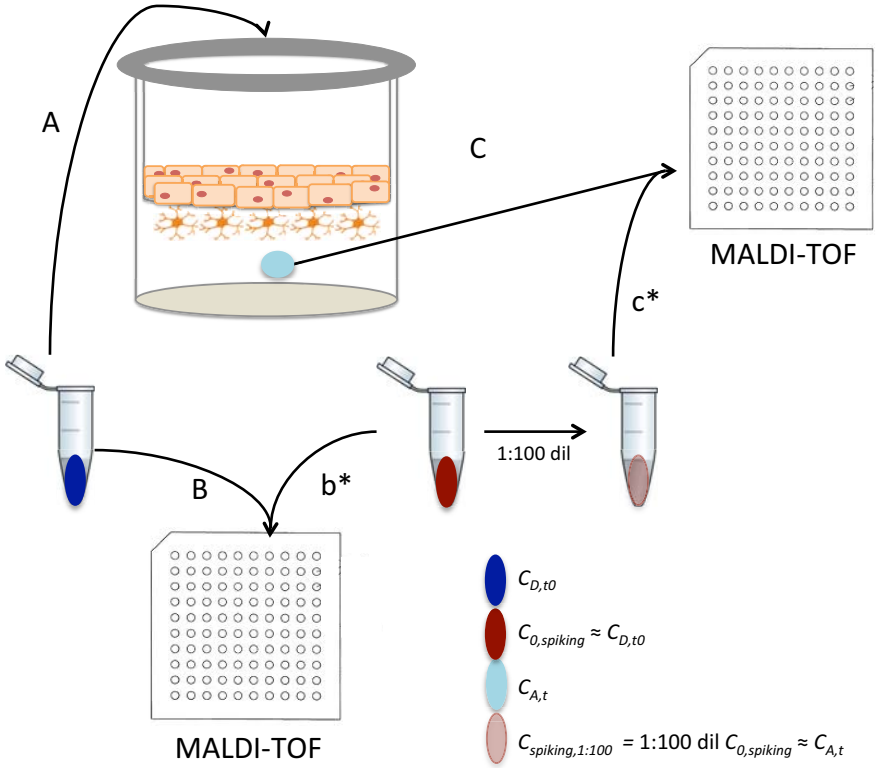


Figure 2.7. Spiking scheme for simplified MALDI quantification method. Heavy peptide initial solution ($C_{0,spiking}$ = 200 µM) its diluted to 2 µM ($C_{spiking,1:100}$) by subsequent 1:10 dilutions. $C_{0,spiking}$ is used for spiking of $C_{D,t0}$ of to achieve similar signals intensities of pair-peptides. Similarly, $C_{spiking,1:100}$ is used for spiking acceptor (baso-lateral) assay solution ($C_{A,t}$), which is expected to cross around 1%. Spiked solutions are plated in MALDI-TOF plates mixed with ACH matrix. In A, Sample $C_{D,t}$ is introduced to the donor (apical) compartment. In B/b* and C/c* assessment of $Q_{D,t}$ and $Q_{A,t}$ was performed. Finally, 1 µL of spiked solution mixture was mixed with 1 µL of ACH matrix and analysed by MALDI-TOF.

Peptides were tested at a concentration of 200 µM. Heavy peptide spiking solutions were set up at the same concentration ($C_{0,spiking}$ = 200 µM). Two 1:10 dilutions on the $C_{0,spiking}$ were applied to get $C_{spiking,1:10}$ and $C_{spiking,1:100}$. C_{t0} of tested peptides were spiked with $C_{0,spiking}$ and C_A obtained after the *in vitro* assay were spiked with $C_{spiking,1:100}$. In order to have heavy:light ratios close to 1 we will spike with $C_{spiking,1:10}$ or $C_{spiking,1:100}$ depending on the transport expected. Since we already calculated the transport with HPLC-UV method, we know the expected transport of each peptide, which is close to 1%. Then, in this case, we used $C_{spiking,1:100}$. C_A results were normalized depending on the dilution factor used for quantification, multiplying by 100 the signal values corresponding to heavy peptide (because we used $C_{spiking,1:100}$).

$$\%T = \frac{Q_A(t)}{Q_D(t_0)} = \frac{\left[A_{light}/dil \cdot A_{heavy}\right]_{A,t}}{\left[A_{light}/A_{heavy}\right]_{D,t0}}$$

where A_{light} refers to the areas of all isotopic peaks of the light compound and A_{heavy} refers to the areas of all isotopic peaks of the heavy compounds; [M+H$^+$], [M+Na$^+$] and [M+K$^+$]. Also, *dil* refers

to dilution of used spiking solution for determining peptide amount on acceptor compartment. Then, *dil* could be 10 or 100 when 1:10 or 1:100 dilutions are used, respectively.

Quantification results from both HPLC-UV and MALDI-TOF are shown in Table 2.5 and Figure 2.8.

A)

Compounds	*%T* (by HPLC-UV)	*%T* (by MALDI-TOF)
1L	2.9 ± 1.4	3.8 ± 1.4
2L	3.9 ± 0.9	3.4 ± 0.7
3L	2.0 ± 0.8	2.5 ± 1.2
4L	2.0 ± 0.9	2.1 ± 0.5
5L	3.4 ± 1.6	3.1 ± 1.6
6L	2.3 ± 1.4	-
7L	3.8 ± 0.8	3.7 ± 1.1
8L	5.5 ± 1.2	5.8 ± 0.6
9L	2.7 ± 1.2	4.9 ± 2.1
11L	5.3 ± 2.6	5.4 ± 3.1
1D	5.3 ± 1.8	3.1 ± 0.9
2+D	5.7 ± 1.8	7.5 ± 2.9
2-D	6.8 ± 1.2	6.8 ± 2.2
3D	7.2 ± 0.9	7.5 ± 1.0
4D	4.4 ± 0.5	2.9 ± 1.2
5D	5.0 ± 2.3	6.5 ± 2.4
6D	5.3 ± 2.5	7.2 ± 3.1
7D	4.5 ± 2.0	3.1 ± 1.2
8D	4.8 ± 3.4	5.2 ± 3.8
9D	4.8 ± 1.9	4.9 ± 1.8
11D	4.6 ± 1.9	5.9 ± 2.7

B)

Compounds	P_{app} (by HPLC-UV) / 10^{-6} cm/s	P_{app} (by MALDI-TOF)/ 10^{-6} cm/s
1L	2.5 ± 1.2	3.2 ± 1.2
2L	3.3 ± 0.8	2.8 ± 0.6
3L	1.7 ± 0.6	2.1 ± 1.0
4L	1.7 ± 0.8	1.8 ± 0.4
5L	2.9 ± 1.4	2.6 ± 1.3
6L	1.9 ± 1.2	-
7L	3.2 ± 0.7	3.2 ± 0.9
8L	4.6 ± 1.0	5.0 ± 0.5
9L	2.0 ± 1.0	4.1 ± 1.8
11L	4.4 ± 2.2	4.5 ± 2.6
1D	4.5 ± 1.5	2.6 ± 0.8
2+D	4.8 ± 1.5	6.3 ± 2.4
2-D	5.7 ± 1.0	5.7 ± 1.9
3D	6.1 ± 0.8	6.3 ± 0.8
4D	3.7 ± 0.4	2.4 ± 1.0
5D	4.2 ± 1.9	5.5 ± 2.0
6D	4.4 ± 2.1	6.1 ± 2.6
7D	3.8 ± 1.7	2.6 ± 1.0
8D	4.1 ± 2.9	4.4 ± 3.2
9D	4.0 ± 1.6	4.1 ± 1.5
11D	3.9 ± 1.3	5.0 ± 2.2

Table 2.5. Comparison of transport results between HPLC-UV method and MALDI-TOF method. A) Percentage of transport *(%T*); B) Apparent permeability (P_{app}).

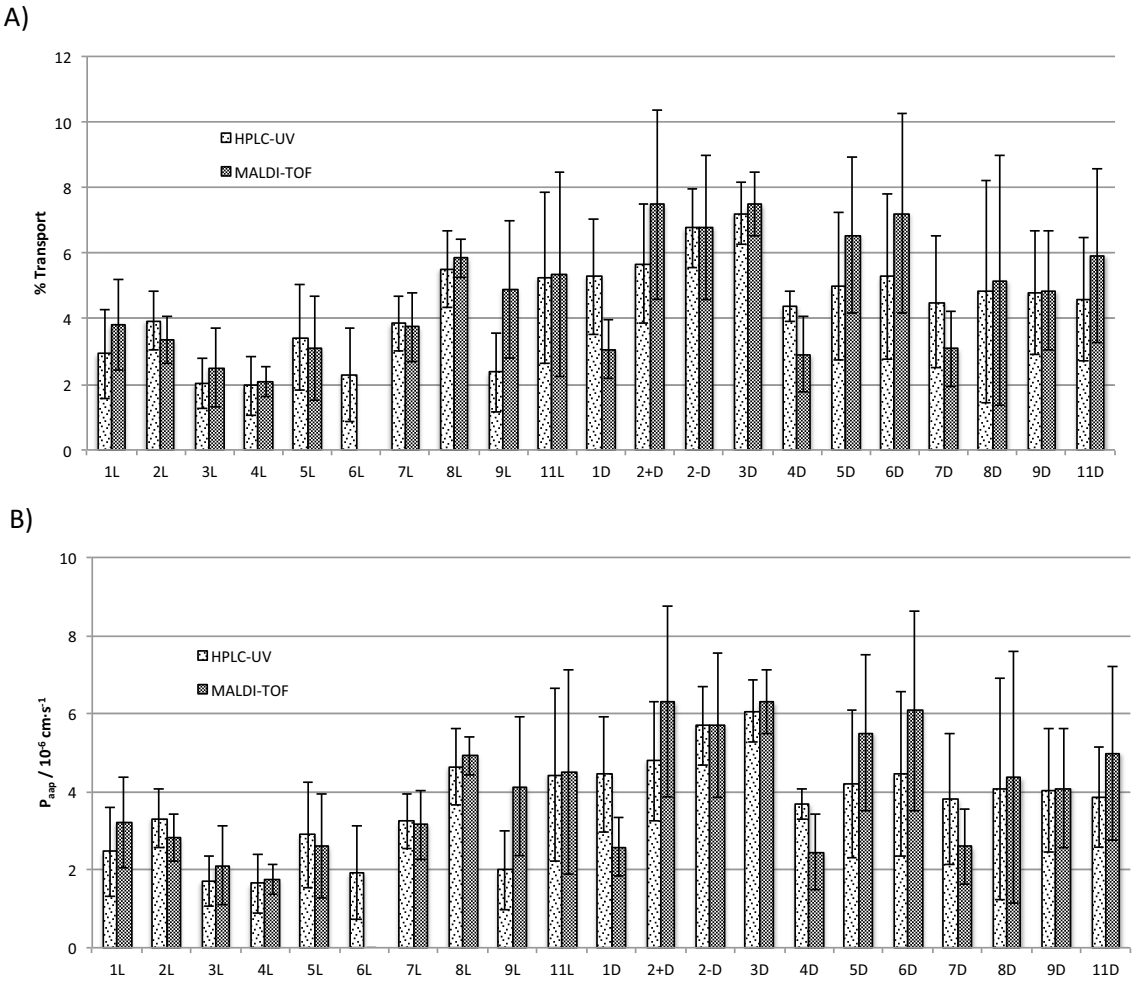


Figure 2.8. Illustration of comparison results by HPLC-UV method and MALDI-TOF method shown in Table 2.5. A) Percentage of transport *(%T*); B) Apparent permeability (P_{app}).

These results were analysed in terms of discovery of improved BBB-shuttles candidates and also on the assessment of MALDI-TOF MS quantification technique.

2.5.4. Discovery of improved BBB-shuttles candidates

Transport results (*%T* or P_{app} values) from library analogues tested by triplicate across the *in vitro* cell-based BBB model and the variability among each triplicate impede to choose any analogue significantly performing better as BBB-shuttle than the original one, neither with HPLC-UV nor MALDI-TOF quantification techniques.

Scrambled peptides tested as negative controls (9L and 9D) showed medium transport abilities on crossing *in vitro* BBB model among all analogues in the library. Evidences lead us hypothesize that *in vitro* cell-based BBB model was not performing as desired (probably not expressing required amounts of TfR) even though standard assay quality parameters (TEER and LY-P_{app}) were fitting requirements described in the bibliography. This hypothesis is supported by previous results obtained in the lab, where HAIYPRH-peptide exhibited transport values significantly higher.[15]

Anyway, D-library showed slightly better transport performance than L-library. We hypothesize this general behaviour from most peptides on D-library is due to its protease resistance structure while L-peptides are faster degraded by proteases secreted by cells during the assay.

2.5.5. Evaluation of quantification techniques

To determine errors intrinsic to each quantification technique we first considered the whole library as an entity instead of 21 individual analogues. To this purpose, we averaged the difference between relative errors of both techniques to exclude the biological error. We conclude both techniques performed equivalently, only differing on 0.6%.

Compound	P_{app} (HPLC-UV) / 10^{-6} cm/s	P_{app} (MALDI-TOF) / 10^{-6} cm/s	Relative error HPLC-UV	Relative error MALDI-TOF	Δ Relative error
1L	2.5 ± 1.2	3.2 ± 1.2	0.48	0.38	0.10
2L	3.3 ± 0.8	2.8 ± 0.6	0.24	0.21	0.03
3L	1.7 ± 0.6	2.1 ± 1.0	0.35	0.48	-0.13
4L	1.7 ± 0.8	1.8 ± 0.4	0.47	0.22	0.25
5L	2.9 ± 1.4	2.6 ± 1.3	0.48	0.50	-0.02
6L	1.9 ± 1.2	-	0.63	-	-
7L	3.2 ± 0.7	3.2 ± 0.9	0.22	0.28	-0.06
8L	4.6 ± 1.0	4.9 ± 0.5	0.22	0.10	0.12
9L	2.0 ± 1.0	4.1 ± 1.8	0.50	0.44	0.06
11L	4.4 ± 2.2	4.5 ± 2.6	0.50	0.58	-0.08
1D	4.5 ± 1.5	2.6 ± 0.8	0.33	0.31	0.02
2+D	4.8 ± 1.5	6.3 ± 2.4	0.31	0.38	-0.07
2-D	5.7 ± 1.0	5.7 ± 1.9	0.18	0.33	-0.15
3D	6.1 ± 0.8	6.3 ± 0.8	0.13	0.13	0.00
4D	3.7 ± 0.4	2.4 ± 1.0	0.11	0.42	-0.31
5D	4.2 ± 1.9	5.5 ± 2.0	0.45	0.36	0.09
6D	4.4 ± 2.1	6.1 ± 2.5	0.48	0.41	0.07
7D	3.8 ± 1.7	2.6 ± 1.0	0.45	0.38	0.07
8D	4.1 ± 2.9	4.4 ± 3.2	0.70	0.73	-0.03
9D	4.0 ± 1.6	4.1 ± 1.5	0.40	0.37	0.03
11D	3.9 ± 1.3	5.0 ± 2.2	0.33	0.44	-0.11
				Average:	**0.006**

Table 2.6. Comparison of relative errors achieved by HPLC-UV method and MALDI-TOF method considering apparent permeability (P_{app}).

However, when looking closer to every single analogue within the library we realized both techniques dramatically disagree on transport quantification. Listing results in terms of HPLC-UV or MALDI-TOF quantification leads to a completely distinct ranking as shown in Table 2.7 and Figure 2.9, (P_{app} values).

From these results we were not able to conclude which quantification technique was the most trustable for determining peptide transport across the *in vitro* cell-based BBB model. At least on peptide libraries where analogues exhibit such similar transport rates.

Despite being the most-used quantification technique, HPLC-UV is far from performing ideally. First, it depends on the UV absorption of each analysed molecule, which is indeed strongly influenced by the concentration of that molecule on the analysed sample. Even injecting the highest volume allowed by the equipment, we usually face tiny peaks that difficult area integration for quantification assessment. Second, other molecules (such as endogenous molecules from the *in vitro* assay) with similar retention time can contribute to the peak area, further biasing real results. These factors may lead to inconsistent results. Contrarily, MALDI-TOF technique shows a much better sensitivity for all library compounds. Moreover, higher resolution, precision and sensitivity of MALDI-TOF compared to HPLC-UV ensure higher reliable data recorded by MALDI-TOF. Since transport calculations are based on peak area, sensitivity and precision are crucial parameters to avoid including any effect from impurities present in the

sample. Analogues comprised on the HAIYPRH-library appear in MALDI-TOF spectra as three sets of peaks corresponding to [$M_{analogue}$ + H^+], [$M_{analogue}$ + Na^+] and [$M_{analogue}$ + K^+], each of them carrying their isotopic envelope peaks. In addition, MALDI-TOF equipment showing enough sensitivity, such as MALDI-4800, leads to much more robust data acquisition.

A)

Compound	P_{app} (by HPLC-UV) / 10^{-6} cm/s
3D	6.06 ± 0.79
2-D	5.69 ± 1.01
2+D	4.78 ± 1.54
8L	4.64 ± 0.98
1D	4.45 ± 1.49
6D	4.44 ± 2.11
11L	4.43 ± 2.21
5D	4.20 ± 1.91
8D	4.06 ± 2.85
9D	4.04 ± 1.59
11D	3.86 ± 1.28
7D	3.79 ± 1.68
4D	3.68 ± 0.40
2L	3.31 ± 0.75
7L	3.24 ± 0.69
5L	2.89 ± 1.35
1L	2.47 ± 1.15
9L	1.99 ± 1.03
6L	1.92 ± 1.19
3L	1.70 ± 0.64
4L	1.65 ± 0.76

B)

Compound	P_{app} (by MALDI-TOF) / 10^{-6} cm/s
2+D	6.30 ± 2.44
3D	6.29 ± 0.82
6D	6.07 ± 2.56
2-D	5.71 ± 1.86
5D	5.51 ± 2.01
11D	4.99 ± 2.23
8L	4.92 ± 0.49
11L	4.51 ± 2.62
8D	4.35 ± 3.22
9L	4.13 ± 1.77
9D	4.09 ± 1.53
1L	3.20 ± 1.16
7L	3.15 ± 0.89
2L	2.82 ± 0.62
5L	2.60 ± 1.33
7D	2.60 ± 0.97
1D	2.58 ± 0.75
4D	2.44 ± 0.97
3L	2.11 ± 1.01
4L	1.75 ± 0.38
6L	-

Table 2.7. Ranking of peptides based on permeability results. A) HPLC-UV; B) MALDI-TOF.

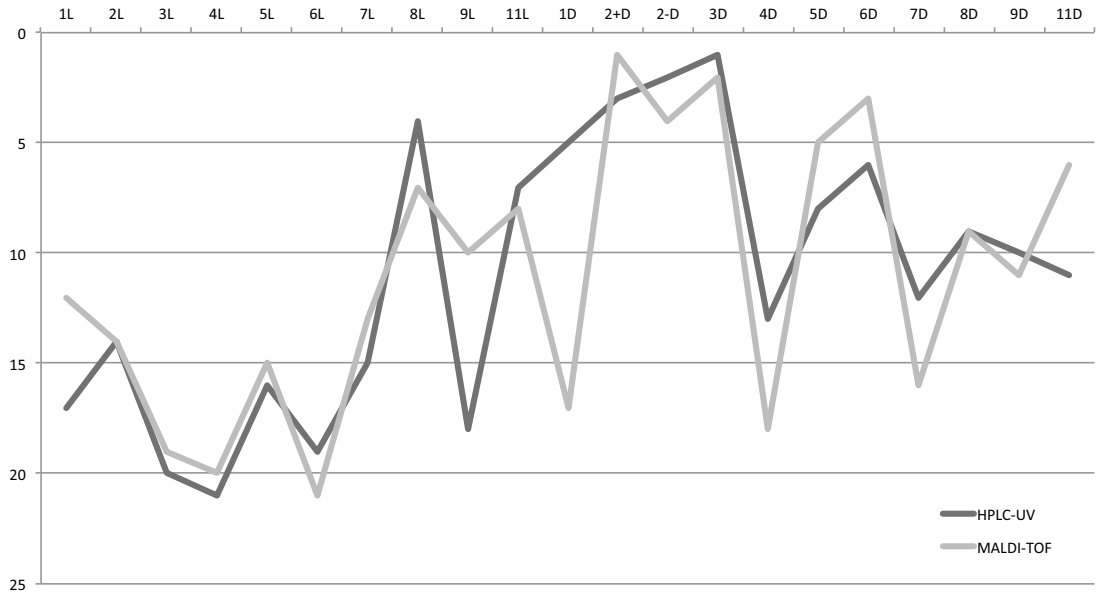


Figure 2.9. Comparison of relative ranking positions regarding transport abilities depending on used quantification method: HPLC-UV (dark grey) or MALDI-TOF (light grey) of library analogues across the *in vitro* cell-based BBB model.

Unfortunately, we finally conclude that these assay results were not conclusive to select peptide analogues excelling on transport ability across the *in vitro* cell-based BBB model. However, experience obtained on the project has been useful since it has been a test bench for setting up MALDI-TOF quantification technique in a real case scenario for quantification of a notably large peptide library.

2.6. Future prospects: Analysis of mixtures

Altogether, we did not consider performing analogue mixture assays at this point. We need a robust and reproducible *in vitro* cell-based BBB model to test analogue mixtures. When testing mixtures, other factors such as interaction among tested peptides and competition among them will influence transport rates. On peptide mixtures, interaction among them cannot be controlled in any way. On the other hand, competition of mixed analogues will depend on the used assay. Passive diffusion assays such as PAMPA assay do not screen on competition among candidates because passive transport mechanism is not saturable (at least at low concentrations such as the ones assayed in this chapter). However, *in vitro* cell-based BBB assay mimics BBB features and it is described to express TfR receptor, among others. Since peptide library is based on similar analogues of HAIYPRH peptide, all of them are expected to interact with TfR receptor in a similar way. Active transport mechanisms are saturable, thus providing a competitive environment, which enhance stringency on peptide selection. On this very competitive environment relies the added value of mixture approaches, although it cannot be regarded as a transport validation approach.

It is noteworthy to highlight that MALDI-TOF quantification technique would be the best candidate to quantify transport of peptides in such mixtures. MALDI-TOF has a valuable characteristic that recommends its use over HPLC-UV. Peptides within a mixture of analogues show very similar polarity. Thus, analogues would appear in the HPLC spectra in very similar retention times. Although peaks can be slightly separated modifying the gradient applied, they appear broader. In contrast, most peptides within the mixture have distinct mass. Mass spectrometry is able to quantify peptides showing distinct m/z because peaks do not overlap. Since signal of each peptide is composed of a molecular peak and 3 isotopic peaks and for quantification a internal patron is required (being 4 a.m.u. heavier), molecules showing mass variations larger than 8 atomic mass units would be easily quantified by MALDI-TOF.

However when analysed mixtures are larger and more complex, other mass spectrometry techniques with higher resolution such as Orbitrap, FT-ICR MS or triple quadrupole (QqQ) instruments would be required.

Bibliography

1. McGregor, D.P. Discovering and improving novel peptide therapeutics. *Curr. Opin. Pharmacol.* **2008**, 8, 616-619.
2. Vlieghe, P., Lisowski, V., Martinez, J. & Khrestchatisky, M. Synthetic therapeutic peptides: science and market. *Drug Discov. Today* **2010**, 15, 40-56.
3. Weinstock, M.T., Francis, J.N., Redman, J.S. & Kay, M.S. Protease-resistant peptide design-empowering nature's fragile warriors against HIV. *Biopolymers* **2012**, 98, 431-442.
4. Antosova, Z., Mackova, M., Kral, V. & Macek, T. Therapeutic application of peptides and proteins: parenteral forever? *Trends Biotechnol.* **2009**, 27, 628-635.
5. Lee, J.H., Engler, J.A., Collawn, J.F. & Moore, B.A. Receptor mediated uptake of peptides that bind the human transferrin receptor. *Eur. J. Biochem.* **2001**, 268, 2004-2012.
6. Giralt, E., Teixidó, M. & Prades, R. Protease-resistant compounds useful as shuttles through the blood-brain barrier and shuttle-cargo constructs. **2013**, WO2013127829.
7. Kaiser, E., Colescott, 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.
8. Christensen, T. Qualitative test for monitoring coupling completeness in solid-phase using chloranil. *Acta Chem. Scand. B. Org. Chem. Biochem.* **1979**, 33, 763-766.
9. Madder, A., Farcy, N., Hosten, N.G.C., De Muynck, H., De Clercq, P.J., Barry, J. & Davis, A.P. A novel sensitive colorimetric assay for visual detection of solid-phase bound amines. *Eur. J. Org. Chem.* **1999**, 1999, 2787-2791.
10. Cecchelli, R., Dehouck, B., Descamps, L., Fenart, L., Buee-Scherrer, V.V., Duhem, C., Lundquist, S., Rentfel, M., Torpier, G. & Dehouck, M.P. In vitro model for evaluating drug transport across the blood-brain barrier. *Adv. Drug Deliv. Rev.* **1999**, 36, 165-178.
11. Gaillard, P.J. & de Boer, A.G. 2B-Trans technology: targeted drug delivery across the blood-brain barrier. *Methods Mol. Biol.* **2008**, 437, 161-175.
12. Crone, C. & Olesen, S.P. Electrical resistance of brain microvascular endothelium. *Brain Res.* **1982**, 241, 49-55.
13. Culot, M., Lundquist, S., Vanuxeem, D., Nion, S., Landry, C., Delplace, Y., Dehouck, M.P., Berezowski, V., Fenart, L. & Cecchelli, R. An in vitro blood-brain barrier model for high throughput (HTS) toxicological screening. *Toxicol. In Vitro* **2008**, 22, 799-811.
14. Hoheisel, D., Nitz, T., Franke, H., Wegener, J., Hakvoort, A., Tilling, T. & Galla, H.J. Hydrocortisone reinforces the blood-brain barrier properties in a serum free cell culture system. *Biochem. Biophys. Res. Commun.* **1998**, 244, 312-316.
15. Prades, R. PhD thesis. Towards a universal blood-brain barrier shuttle: protease-resistant peptide shuttles with capacity to deliver cargos into the central nervous system. **2012**, Universitat de Barcelona.
16. Garberg, P., Ball, M., Borg, N., Cecchelli, R., Fenart, L., Hurst, R.D., Lindmark, T., Mabondzo, A., Nilsson, J.E., Raub, T.J., Stanimirovic, D., Terasaki, T., Oberg, J.O. & Osterberg, T. In vitro models for the blood-brain barrier. *Toxicol. In Vitro* **2005**, 19, 299-334.

High-throughput screening methods
for new BBB-shuttles discovery

Chapter 3

Phage display: Exploring simplicity

Work presented on this chapter was fully performed at Drs. Wadih Arap and Renata Pasqualini's laboratory at the University of Texas MD Anderson Cancer Center, Houston, TX (USA), during my stay from July 22nd 2012 to January 21st 2013.

3.1. Library design

On the advent of phage display technology, a wide range of random phage libraries has appeared. Generally speaking, the success of any combinatorial strategy aimed at selecting ligands for a given target strongly depends on the quality of the library, and phage display is not an exception. Hence, not only a complete and homogeneous distribution of variability is required within the library, but also favourable scaffolds to display this variability (by introducing randomized residues) had to be designed to orient the library compounds in a biologically relevant way, thus conditioning all subsequent processes. Several peptide scaffolds have been designed to introduce variability to phage display libraries.[1] As discussed in the Introduction, even chemical actions over peptides displayed on phage have been exploited to broaden the explored chemical space.

Contrarily, in the frame of this thesis we considered a phage library to explore simplicity. We synthesized one of the most basic phage display libraries we could conceive: the CX_3C random peptide phage library. Several drawbacks deserves further consideration when it comes to the simplistic design of this particular library, such as lack of diversity to fulfil specificity requirements of drug candidate molecules, or excessive conformational stiffness of such short and thus constrained cyclic peptides. However, strong literature supports otherwise, for instance, analysis of CX_7C libraries screened *in vivo* in a human patient by Arap *et al.*[2] seeking for peptides targeting specific tissues revealed that tripeptide motifs were recurrently present on clones found on organs/tissues. These observations made us wonder whether tripeptide motifs are sufficient to drive specific and strong interactions that are yet biologically relevant.

3.2. Library synthesis

Three randomized positions constrained by two cysteines forming a disulphide bridge led to a library with 8000 distinct possible variants. Since the CX_3C library has much less variability than other libraries previously synthesized in the lab, or those described by Smith (such as CX_7C and CX_8C libraries), the amount of phage DNA vector (fUSE55) required to start the phage library synthesis was much lower. According to the protocol described by G.P. Smith, the synthesis of a CX_7C library requires approximately 194 µg. Therefore, the CX_3C library could be synthesized with only 1.2 ng of fUSE55. This DNA amount would ensure around $1.5 \cdot 10^4$ copies of phage of each displayed peptide. However, to ensure quality, we started the synthesis with 500 ng of fUSE55. Our phage display library synthesis protocol was adapted from the one reported by G.P. Smith.[3]

3.2.1. Amplification of fUSE55 vector

Production of the fUSE55 plasmid consists of electroporation of pure fUSE55 in commercially available Electromax DH5α bacteria cells. Then, by culturing those DH5α bacteria cells we achieve amplification of fUSE55. Cells were harvested by centrifugation and resuspended in a cell-lysing buffer to release the fUSE55 plasmid. The partially purified fUSE55 needs to be purified by double purification with *Qiagen Giga prep* purification kits and CsCl/EtBr gradient purification.

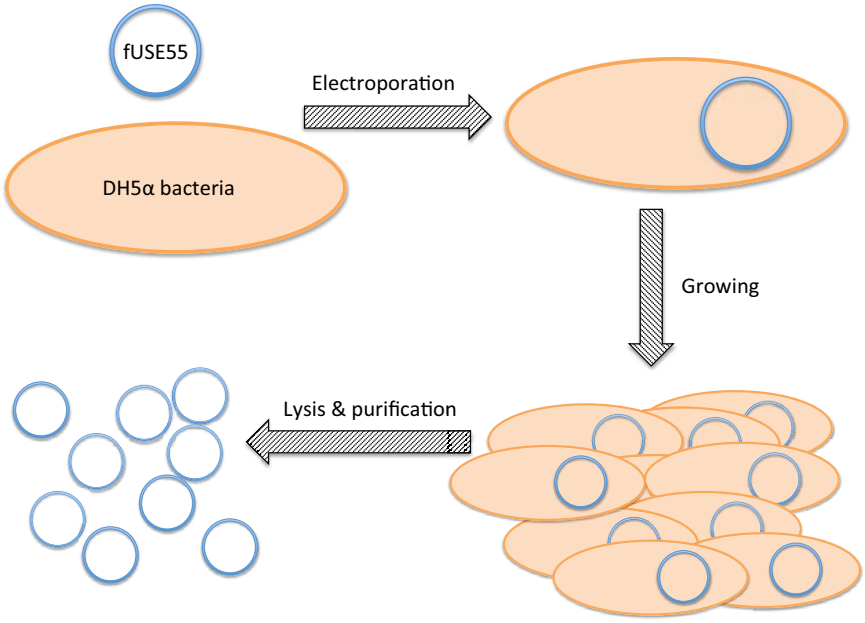


Figure 3.1. Schematic illustration of amplification of fUSE55 DNA vector into DH5α bacteria. Electroporation, growth, lysis and purification steps are required.

3.2.2. Oligonucleotide library conditioning

Oligonucleotide library consisting of a $(NNF)_3$-type library was purchased from a commercial supplier. As discussed in Introduction (see 4.8.2) *N* corresponds to any oligonucleotide (thymine, adenine, guanine or cytosine), while *F* corresponds to only guanine and cytosine. Oligonucleotide library conditioning consists of two steps. First, annealing of oligonucleotides with its corresponding primers as shown in Figure 3.2. Second, during the elongation step, the annealed primer becomes the starting point for the enzyme to create the antiparallel strand of the oligonucleotide library. Finally, a cleaning step was required.

3.2.3. fUSE55 and oligonucleotide insert-library digestion

Enzyme BGLI is used for independent digestion of both oligonucleotide insert-library and fUSE55 vector. Another purification step is further required after each digestion. BGLI enzyme leads to DNA sticky ends (as shown in Figure 3.2.B) for eventual ligation reaction between digested vector (fUSE55) and digested library-insert. Whereas initially considered standard digestion conditions were successfully applied to fUSE55 digestion, poor digestion performance for oligonucleotide insert-library was observed and some slight modifications on Smith protocol were applied to better set up conditions for insert-DNA material. Eventually, for oligonucleotide insert-library digestion, three subsequent overnight reactions with addition of fresh BGLI enzyme were performed with appropriate overall yield.

A) **CX_3C library**

5' – CACTCGGCCGACGGGGCTTGTNNKNNKNNKTGTGGGCCGCTGGGGCCGAA – 3'

Library antisense oligonucleotide

5' – TTCGGCCCCAGCGGC

B) **Annealing**

5' – CACTCGGCCGACGGGGCTTGTNNKNNKNNKTGTGGGCCGCTGGGGCCGAA – 3'
3' - CGGCGACCCCGGCTT – 5'

Elongation

5' – CACTCGGCCGACGGGGCTTGTNNKNNKNNKTGTGGGCCGCTGGGGCCGAA – 3'
3' – GTGAGCCGGCTGCCCCGAACANNKNNKNNKACACCCGGCGACCCCGGCTT – 5'

Digestion (BGLI)

5' – CACTCG**GCC***G*ACG*G***GGC**TTGTNNKNNKNNKTGTGG**GCC***G*CTG*G***GGC**CGAA – 3'
3' – GTGAGC**CGG***C*TGC*C***CCG**AACANNKNNKNNKACACC**CGG***C*GAC*C***CCG**GCTT – 5'

5'…GCCNNNNGGC…3'
3'…CGGNNNNCCG…5'

5' – GGGCTTGTNNKNNKNNKTGTGGGCCGCTG – 3'
3' – TGCCCCGAACANNKNNKNNKACACCCGGC – 5'

Figure 3.2. A) CX_3C purchased oligonucleotide library and its antisense oligonucleotide primer sequence are shown. B) Annealing, elongation and digestion for considered oligonucleotide insert-library are disclosed.

3.2.4. Miniaturized setting up of ligation conditions and whole library ligation

Ligation conditions are one of the most crucial steps on the phage display library synthesis. Low ligation rates would lead to poor oligonucleotide insertions and thus rather non-viable phage or biased variability libraries. As described by Smith, a miniaturized ligation is required to better set up reaction conditions. Five miniaturized reactions testing 1:1, 1:3, 1:5, 1:10, 1:30 vector:insert ratios were tested in order to determine the ideal ratio for best ligation efficiency.

Another important step on library synthesis is the electroporation of the vector-containing library inserts to electrocompetent cells (MC-1061) and culture of electroporated cells to eventually produce phage. Electroporation rates are usually quite low. In some cases, electroporation can collapse and electrocute cells, then loosing all DNA contained on that electroporation cuvette. This is a critical experimental step and it strongly depends on experimental conditions. There are

some parameters that can be optimized for electroporation such as voltage and time of electroporation, although usually standard conditions are used (2kV, 200 Ω, 25 µF). In addition, the size of electroporation cuvettes can lead to variable performance, but the amount of DNA and cells required for library synthesis usually dictate the cuvette size.

To test overall efficiency, electroporated cells were plated on tetracycline plates and colonies counted to determine the efficiency of the process.

fUSE55:insert	Plate 1	Plate 2	Average	Efficiency*
1:1	163	134	148.5	605880
1:3	203	221	212	864960
1:5	766	789	777.5	3172200
1:10	276	261	268.5	1095480
1:30	161	168	164.5	671160

Table 3.1. Ligation efficiency calculated after electroporating MC-1061 electrocompetent cells with fUSE55 vector incorporating oligonucleotide inserts. 1:5 ratio showed the best efficiency of all tested fUSE55:insert ratios. This condition was selected for library synthesis. *Efficiency units are: transformats/µL DNA.

3.2.5. Phage library growth

Once ligation conditions were optimized a new ligation reaction, with 1:5 (fUSE55:oligonucleotide insert) ratio, was performed for phage library synthesis. Vector was electroporated into electrocompetent cells. On one hand, a small aliquot of electroporated cells was plated to calculate process efficiency and on the other hand, library was grown overnight. Electroporation efficiency was determined as $4.6 \cdot 10^6$ transformats/µL DNA.

Bacteria were centrifuged and supernatant (containing phage) was kept as the 1st generation of CX_3C library. Pelleted cells were frozen in case future growing of a 2nd generation of CX_3C phage library would be required (Figure 3.3.B). Phage from the culture supernatant were precipitated and cleaned. Finally, the library was tittered and stored at 4ºC.

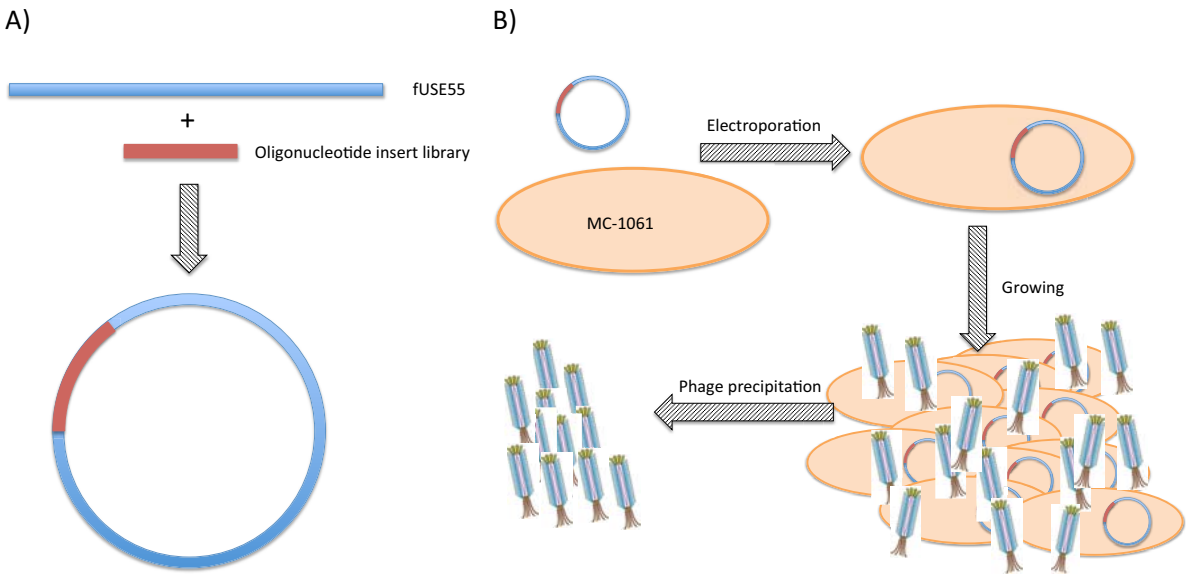


Figure 3.3. A) Ligation of oligonucleotide insert and fUSE55 DNA vector. B) Electroporation of vector construct in MC-1061 electrocompetent bacteria cells, growing, and phage isolation by precipitation.

3.3. Phage library validation and control

As previously stated, the quality of a library strongly determines any screening process performed with that library. To assess the quality of the synthesized phage library we carried out the library

titration and sequencing of 192 random phage. Library titration was determined as $2.4 \cdot 10^9$ TU/µL while sequencing results (after PCR DNA amplification) showed that most phage (over 95%) indeed contained oligonucleotide inserts on its genotype.

Amino acids present in those 192 sequenced inserts were listed and compared. All phage sequenced incorporated an insert, meaning ligation step worked nicely. However, from those 192 picked colonies, 15 sequences were repeated: CESGC (x2), CGFVC (x2), CGGGC (x2), CGGVC (x6), CGRGC (x2), CGVGC (x2), CGVVC (x2), CGWEC (x2), CGWWC (x2), CKKKC (x5), CNKKC (x2), CNNKC (x2), CVFFC (x2) and CVGGC (x3). Those repetitions highly express Gly, Lys and Val.

CAFVC	CERGC	CGFGC	CGLGC	**CGWEC**	CKGKC	CKSRC	CNFDC	CQFAC	CRKNC	CSVQC	CVGQC
CAGDC	CERIC	**CGFVC**	CGLVC	CGWKC	**CKKKC**	CKYSC	CNGGC	CQGGC	CRMGC	CSVSC	CVPMC
CAGVC	**CESGC**	**CGFVC**	CGMEC	CGWVC	**CKKKC**	CLEQC	CNHKC	CQNKC	CRRGC	CTHSC	CVQLC
CAGWC	**CESGC**	CGGAC	CGMVC	**CGWWC**	CKKNC	CLGEC	**CNKKC**	CQTMC	CRRYC	CTKNC	CVRGC
CAQKC	CEWMC	CGGEC	CGQKC	**CGWWC**	CKMEC	CLGYC	**CNKKC**	CQVGC	CRSKC	CTMKC	CVRGC
CAVGC	CEWVC	**CGGGC**	**CGRGC**	CGYKC	CKMFC	CLIWC	CNKLC	CQVVC	CRVGC	CTNIC	CVRLC
CDETC	CFFAC	**CGGGC**	**CGRGC**	CHQDC	**CKNKC**	CLKLC	CNMNC	CRAFC	CRVKC	CTTNC	CVRRC
CDGGC	CFFVC	CGGLC	CGRKC	CHSEC	**CKNKC**	CLPPC	CNNEC	CRDRC	CSDFC	CTVNC	CVRSC
CDKKC	CFGVC	**CGGVC**	CGRLC	CILAC	**CKNKC**	CLWFC	**CNNKC**	CRENC	CSFSC	CTWPC	CVRVC
CDKVC	CFIPC	**CGGVC**	CGRVC	CIMRC	**CKNKC**	CMDKC	**CNNKC**	CRERC	CSFTC	CVEEC	CVWLC
CDWGC	CFPVC	**CGGVC**	**CGVGC**	CITRC	**CKNKC**	CMDPC	CNQKC	CRFAC	CSFYC	**CVFFC**	CVWMC
CEGNC	CFSFC	**CGGVC**	**CGVGC**	CITSC	CKRDC	CMGGC	CNRMC	CRGGC	CSGAC	**CVFFC**	CVWVC
CEGRC	CFVRC	**CGGVC**	CGVRC	CKDDC	CKRKC	CMVKC	CNRSC	CRGMC	CSLGC	**CVGGC**	CWLGC
CEIKC	CGDAC	**CGGVC**	**CGVVC**	CKDQC	CKRRC	CMWVC	CNYKC	CRHEC	CSMGC	**CVGGC**	CWMVC
CEKKC	CGDGC	CGKWC	**CGVVC**	CKERC	CKSGC	CMYGC	CPNVC	CRHHC	CSNKC	**CVGGC**	CYGVC
CEKMC	CGERC	CGLEC	**CGWEC**	CKFMC	CKSKC	CNEVC	CQDNC	CRIQC	CSTMC	CVGLC	CYVSC

Table 3.2. All sequenced colonies (96) from a plated aliquot of library electroporated bacteria previous library growing.

Amino acid abundance proportions present in sequenced peptides displayed on the library are listed below. We compared the amino acid abundance of the synthesized phage library with the expected theoretically and with a commercially available CX_7C phage library from BioLabs© (which is comparable in terms of amino acid variability, since also has an *NNF* nature). Since oligonucleotide library purchased for phage library synthesis was *NNF*-type, theoretically expected proportions of amino acids are listed and included in Table 3.2 and Figure 3.4.

Amino acid	CX_3C library	Theoretical (*NNF*)	Commercial library
A	2.43	7.14	5.02
C	0.00	3.57	0.16
D	3.30	3.57	6.12
E	5.21	3.57	4.08
F	5.03	3.57	2.51
G	18.58	7.14	4.40
H	1.22	3.57	5.18
I	1.74	3.57	2.35
K	11.81	3.57	4.71
L	3.65	10.71	8.79
M	4.34	3.57	4.40
N	6.25	3.57	5.49
P	1.39	7.14	6.28
Q	2.95	3.57	4.87
R	8.51	3.57	4.71
S	4.69	7.14	11.46
T	2.43	7.14	8.95
V	11.11	7.14	4.71
W	3.82	3.57	2.04
Y	1.56	3.57	3.77

Table 3.3. Values comparison between CX_7C commercial library by BioLabs©, theoretical amounts expected for a *NNF* library and experimental presence of amino acids in synthesized CX_3C library. Values expressed in percentage.

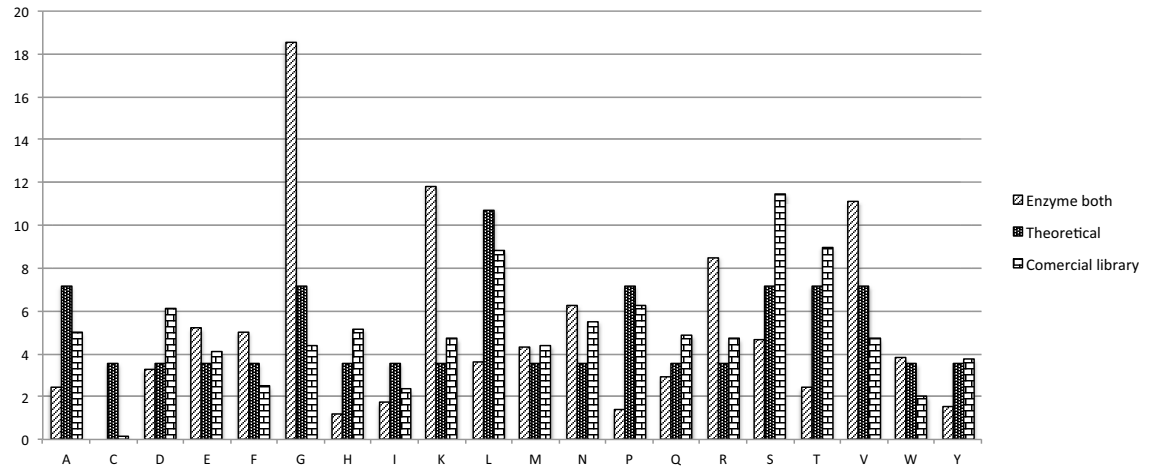


Figure 3.4. Figure illustrating data shown on Table 3.3.

Both libraries (synthesized CX_3C and commercial CX_7C) show significant differences on amino acid abundance compared to the expected theoretical values (according to *NNF* insert library pattern). As discussed above and in the introduction chapter, several factors may influence the amino acid abundance and variability. Some of these factors cannot be fully controlled or optimized, specially when oligonucleotides are purchased commercially. Since the library apparently presents adequate features such as presence of all amino acids (except cysteine) and presence of inserts on most selected and analysed phage genome, we accepted the library for screening experiments.

3.4. *In vivo* screening and evaluation of CX_3C phage library

CX_3C phage library was evaluated *in vivo* with the goal of seeking for molecules targeting the brain, as peptides displayed on phage could represent new BBB-shuttles. Phage library was intravenously injected on the tail vein of Blab/c mice. Phage circulating blood throughout brain capillaries can undergo four distinct fates on the brain. First, they may cross the BBB and reach the brain parenchyma. Second, they may internalize capillary endothelial cells, but not cross the BBB. Third, they may interact with receptors, protein or membrane surface of luminal side of brain capillary endothelial cells. Those interactions can be either specific or non-specific. Finally, phage circulating with blood might not experience any interaction on brain capillaries thus continuing with blood flow.

Before harvesting the whole brain after each panning round, mice were perfused to remove all blood from the capillaries. This step allows the removal of phage not interacting with brain capillaries or those interacting non-specifically with the luminal surface of endothelial cells. Also, specifically bound molecules to any receptor or in any part of the membrane could be removed during perfusion, but depends on the affinity of each interaction. Thus, by harvesting the whole brain after each panning round we cannot ensure that those sequences are BBB-shuttles crossing to brain parenchyma. Hence, other experiments need to be performed to validate found sequences as BBB-shuttles.

Although capillary depletion protocols after brain homogenization had been described, we discarded this option considering they were not 100% efficient and would also require further experiments to validate selected sequences to ensure its BBB-shuttle abilities.[4, 5]

Three biopanning rounds and a post-panning selection round were performed. Biopanning rounds were set up for brain enrichment. Each panning round was performed with 3 mice. $5 \cdot 10^9$ TU of CX_3C phage library were intravenously injected in BALB/c mice tail vein. After 6 hours of circulation time, mice were deeply anesthetized and perfused with PBS. Brain (targeted organ) and kidney and liver (control organs) were harvested and weighted for all triplicates. Weighted portions of tissues were grinded.

Log phase *E. coli* K91/kan bacteria were added to homogenized tissue for infection. Dilutions of this mixture for all tissues were plated to determine the amount of phage on each organ by colony counting. 96 colonies from brain tissue were picked up to perform PCR reaction on its DNA for sequencing after each panning round. Only bacteria infected by phage in brain-homogenized tissue were grown overnight to amplify phage for the next panning round and continue the affinity selection process of phage targeting the brain. Phage was precipitated twice and tittered previous injection to the subsequent panning round. This protocol was performed for each panning round.

Sequences present in all panning rounds and those most repeated ones in the last round were selected for a post-panning round. We designed a post-panning round to avoid biased and false positive results. Since infectivity efficiency depends on displayed sequence on pIII, phage amplification after each panning round could bias phage population. All selected phage were grown, precipitated and tittered separately.

AFG	FGF	GGW	GWG	KVK	MTH	**RGG**	SFG	VGY	WVG
AFV	FGV	GHG	**GWR**	KVK	MVG	**RGG**	SFY	VKD	YDW
AGG	FHV	GIV	**GWR**	KWK	MWL	**RGG**	SGG	VKS	YGV
AGM	FIL	GIV	**GWR**	LAR	NEL	**RGG**	SGV	VLA	YKS
AHR	FIR	GKL	GWR	LER	NKK	**RGG**	SIM	VLF	YQR
AHR	**FLF**	GLV	GWV	LFG	NKK	RGV	SLV	VLQ	YQV
AKG	**FLF**	GNQ	IFR	LFL	NLM	RGV	SRF	VLV	YRM
AKK	**FLF**	GNQ	IFS	LFL	NNK	RKS	SRF	VMV	
AKK	FLV	GPF	IGR	LFR	NNK	RNG	SSW	VNF	
ATG	FPV	GQH	IGS	LFR	NNN	RNN	SYE	VRF	
AVV	FPV	GRE	IHH	LFT	NNR	RNQ	TDK	VSV	
AWF	FSF	GRN	IHH	LGD	NNY	RPL	TGH	VTD	
AWG	FSF	**GRV**	IRE	LGF	NQF	RRF	TKG	VTK	
CVS	FSR	**GRV**	ITR	LGR	NRW	RRP	TKG	VVF	
DDG	FVF	**GRV**	KEM	LKG	NSK	RSG	TKQ	VVG	
DGA	FVF	**GRV**	**KGK**	LNR	NSK	RSM	TLL	VVS	
DGA	FVS	**GRV**	**KGK**	LQM	*NSQ*	RVV	TNT	VVV	
DGA	FVV	**GRV**	**KGK**	LSL	*NSQ*	*RWE*	TNT	VWG	
DGT	FYV	**GRV**	**KGK**	LTD	*NSQ*	*RWE*	TTF	VWV	
DSQ	GFG	GSK	**KGK**	LVG	*NSQ*	*RWE*	VDT	VWW	
DSR	GFR	GSK	KGW	LVG	*NSQ*	*RWE*	VFD	VYG	
EGG	GFW	GSN	KIN	LVG	*NSQ*	*RWE*	*VFF*	VYR	
EKK	GGE	GTL	KKK	LVG	*NSQ*	*RWE*	*VFF*	WGF	
ELR	GGE	GVE	KKM	MAV	*NSQ*	*RWE*	*VFF*	**WGG**	
EMR	**GGG**	GVG	KKM	MGK	PTQ	*RWE*	*VFF*	**WGG**	
EPG	**GGG**	GVL	KNK	MKN	QTM	*RWE*	*VFF*	**WGG**	
ERR	**GGG**	GVR	KNK	MKR	QVS	*RWE*	VFV	**WGG**	
FAS	**GGG**	GVV	KQK	MKT	RDG	RWG	VGE	**WGG**	
FAS	GGR	GVW	KQK	MKT	RFF	RWG	VGG	**WGG**	
FAS	GGV	GWA	KQK	MRR	RGE	SAR	VGG	WRG	
FFS	GGV	GWD	KVG	MRR		SDL	VGM	WSQ	

Table 3.4. Selection of post-panning phage candidates. Table shows sequenced colonies of all three panning rounds: round 1, blue; round 2, green; round 3, red. In **bold**, those sequences found in all panning rounds. In *italics*, sequences highly repeated but not present in all panning rounds, which are considered worth including.

Post-panning round consisted on injecting two BALB/c mice with $5 \cdot 10^9$ TU of each selected phage. Thus, post-panning round concentrations of all tested phage were initially the same. The same protocol as for panning rounds was performed and finally 96 colonies were sequenced. The following sequences were selected for further tests: CNSQC, CRWEC and CFLFC. CNSQC (51%) and CRWEC (33%) were selected, as they were the most abundant phage sequences found in the brain of injected mice. Although CFLFC (4%) was far less abundant in the post-panning selection round, it was still included on the selection to explore such distinct sequence in terms of physicochemical properties.

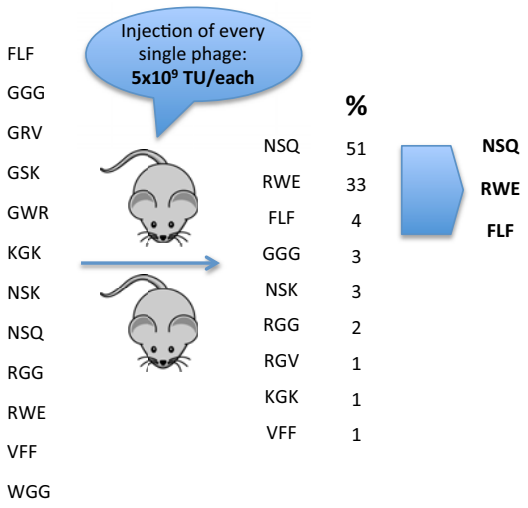


Figure 3.5. Post-panning round results. Two BALB/c mice were injected with $5 \cdot 10^9$ TU of each selected phage. CNSQC (51%) and CRWEC (33%) were selected, as they were the most abundant phage sequences found in the brain of injected mice. Although CFLFC (4%) was far less abundant in the post-panning selection round it was included on the selection.

3.5. Biodistribution and validation of selected sequences

Biopanning and post-panning round selected for three peptide sequences displayed on phage showing high affinity to the brain. However, validation experiments should be performed to better track the fate of those molecules. On one hand, biodistribution experiments were performed by quantitative RT-PCR (q-PCR). Those experiments enlighten how each selected sequence behaves in the entire mouse body (and specially the brain) compared to a phage control not containing the displayed peptide insert, thus not exhibiting the affinity that results from the peptide insert. On the other hand, immunofluorescence experiments were performed by immunostaining to localize phage-containing selected inserts (CFLFC, CRWEC and CNSQC) on the brain.

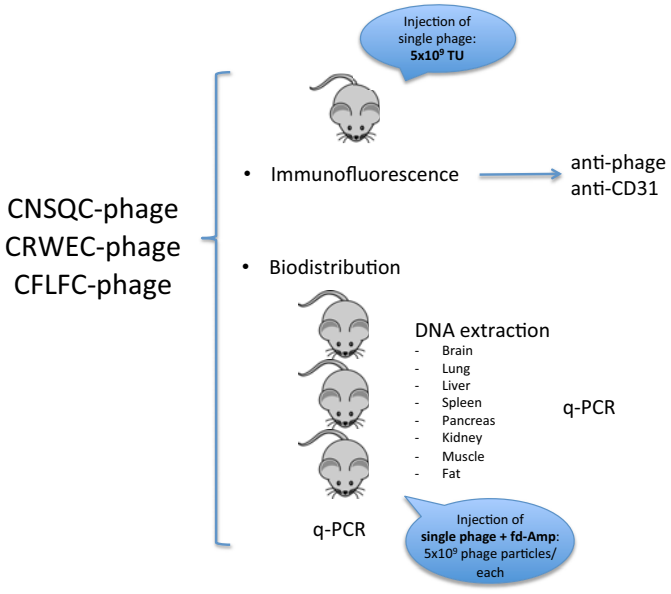


Figure 3.6. Validation experimental set up. Immunofluorescence experiments to find phage location on the brain and biodistribution experiments with q-PCR to quantify affinity to the brain and other organs.

3.5.1. Biodistribution by q-PCR

Biodistribution was assessed by q-PCR phage DNA from brain, lung, liver, spleen, pancreas, kidney, muscle and fat. Phage were q-PCR tittered determining phage particles (not transducing units). Then, each phage sequence was injected ($5{\cdot}10^9$ phage particles) into triplicate sets of mice with $5{\cdot}10^9$ phage particles of fd-Amp phage as a negative control. Fd-Tet and fd-Amp phage show the same phenotype but they have slight differences on their genome. Then, they require distinct primers for DNA amplification on q-PCR and fd-Amp phage do not interfere when calculating amounts of selected sequences and vice versa. Our library phage are fd-Tet based phage. This methodology allows injecting both phage into the same animal, CXXXC-Tet-phage and fd-Amp-phage (as a control with no insert), leading to much more consistent results. Although kidney samples were also analysed, experimental extraction of DNA in these kidney samples failed and we could not get biodistribution data.

Two parameters must be considered on phage biodistribution analysis. On one hand, comparison of amount of phage among organs, this is quantity of phage present in a normalized amount of each organ. It is expected that phage tend to accumulate in higher amounts in liver than in other organs. On the other hand, relative amounts of CXXXC-Tet-phage and fd-Amp-phage in the same organ denoting enrichment over a control sequence.

All results showed are expressed in "phage particles/60 ng of DNA" units.

CFLFC-phage (Figure 3.7) was enriched in the brain 1.5-fold orders over fd-Amp ($1{,}7{\cdot}10^3{\pm}4.6{\cdot}10^2$ vs $1.1{\cdot}10^2{\pm}41$) and 3-times in the liver ($2.7{\cdot}10^5{\pm}7.7{\cdot}10^4$ vs $9.7{\cdot}10^4{\pm}2.0{\cdot}10^4$). In all other organs, FLF was in lesser amount than fd-Amp: in lung ($2.0{\cdot}10^2{\pm}1.5{\cdot}10^2$ vs $4.4{\cdot}10^3{\pm}2.1{\cdot}10^3$), in pancreas ($2.5{\cdot}10^2{\pm}1.8{\cdot}10^2$ vs $7.2{\cdot}10^2{\pm}4.7{\cdot}10^2$), in spleen ($1.7{\cdot}10^4{\pm}2.3{\cdot}10^4$ vs $1.4{\cdot}10^4{\pm}9.3{\cdot}10^3$) in muscle ($1.4{\cdot}10^2{\pm}87$ vs $5.3{\cdot}10^2{\pm}1.7{\cdot}10^2$) and in fat ($1.8{\cdot}10^3{\pm}3.1{\cdot}10^2$ vs $6.0{\cdot}10^3{\pm}1.8{\cdot}10^3$). FLF-phage accumulates mainly in the following organs: liver ($2.7{\cdot}10^5{\pm}7.7{\cdot}10^4$) > spleen ($1.7{\cdot}10^4{\pm}2.3{\cdot}10^4$) > fat ($1.8{\cdot}10^3{\pm}3.1{\cdot}10^2$) > brain ($1.7{\cdot}10^3{\pm}4.6{\cdot}10^2$) > pancreas ($2.5{\cdot}10^2{\pm}1.8{\cdot}10^2$) > lung ($2.0{\cdot}10^2{\pm}1.5{\cdot}10^2$) and > muscle ($1.4{\cdot}10^2{\pm}87$).

CRWEC-phage (Figure 3.8) was enriched in the brain 1.5-fold orders over fd-Amp ($1.8{\cdot}10^3{\pm}3.0{\cdot}10^2$ vs $1.3{\cdot}10^2{\pm}55$) and 2-times in the liver ($1.9{\cdot}10^5{\pm}7.2{\cdot}10^4$ vs $9.7{\cdot}10^4{\pm}2.0{\cdot}10^4$) and 3.2-fold orders in spleen ($8.2{\cdot}10^5{\pm}5.8{\cdot}10^5$ vs $2.5{\cdot}10^4{\pm}1.2{\cdot}10^4$) and 2-times in pancreas ($1.2{\cdot}10^3{\pm}1.6{\cdot}10^3$ vs $6.3{\cdot}10^3{\pm}7.8{\cdot}10^2$). In all other organs, RWE was in lesser amount than fd-Amp: in lung (88±69 vs $6.9{\cdot}10^3{\pm}2.4{\cdot}10^3$), in muscle ($3.9{\cdot}10^2{\pm}1.7{\cdot}10^2$ vs $5.4{\cdot}10^2{\pm}2.5{\cdot}10^2$) and in fat ($2.8{\cdot}10^3{\pm}1.7{\cdot}10^3$ vs $3.2{\cdot}10^3{\pm}2.0{\cdot}10^3$). RWE-phage accumulates mainly in the following organs: spleen ($8.2{\cdot}10^5{\pm}5.8{\cdot}10^5$) > liver ($1.9{\cdot}10^5{\pm}7.2{\cdot}10^4$) > fat ($2.8{\cdot}10^3{\pm}1.7{\cdot}10^3$) > brain ($1.8{\cdot}10^3{\pm}3.0{\cdot}10^2$) > pancreas ($1.2{\cdot}10^3{\pm}1.6{\cdot}10^3$) > muscle ($3.9{\cdot}10^2{\pm}1.7{\cdot}10^2$) and > lung (88±69).

CNSQC-phage (Figure 3.9) was enriched in the brain 1-fold orders over fd-Amp ($1.3{\cdot}10^3{\pm}3.9{\cdot}10^2$ vs $1.2{\cdot}10^2{\pm}69$) and 1,3-times in the liver ($1.0{\cdot}10^5{\pm}5.5{\cdot}10^4$ vs $7.8{\cdot}10^4{\pm}3.1{\cdot}10^4$), 1.5-times in spleen ($2.0{\cdot}10^4{\pm}2.8{\cdot}10^4$ vs $1.3{\cdot}10^4{\pm}6.2{\cdot}10^3$), 3.8-times in pancreas ($3.0{\cdot}10^3{\pm}4.5{\cdot}10^3$ vs $7.9{\cdot}10^2{\pm}2.7{\cdot}10^2$), 1,6-times in muscle ($9.3{\cdot}10^2{\pm}2.3{\cdot}10^2$ vs $5.7{\cdot}10^2{\pm}49$) and 4-times in fat ($8.1{\cdot}10^3{\pm}6.6{\cdot}10^2$ vs $2.0{\cdot}10^3{\pm}4.1{\cdot}10^2$). Only in lung RWE is present in lesser amount than fd-Amp ($1.5{\cdot}10^3{\pm}1.3{\cdot}10^3$ vs $5.7{\cdot}10^3{\pm}3.0{\cdot}10^3$). RWE-phage accumulates mainly in the following organs: liver ($1.0{\cdot}10^5{\pm}5.5{\cdot}10^4$) > spleen ($2.0{\cdot}10^4{\pm}2.8{\cdot}10^4$) > fat ($8.1{\cdot}10^3{\pm}6.6{\cdot}10^2$) > pancreas ($3.0{\cdot}10^3{\pm}4.5{\cdot}10^3$) > lung ($1.5{\cdot}10^3{\pm}1.3{\cdot}10^3$) > brain ($1.3{\cdot}10^3{\pm}3.9{\cdot}10^2$) and > muscle ($9.3{\cdot}10^2{\pm}2.3{\cdot}10^2$).

These three selected sequences CFLFC, CRWEC and CNSQC behave quite similarly in terms of biodistribution. While having outstanding affinity to the brain (over control sequence: fd-Amp), it is known that phage also accumulates in greater quantities in several other organs. In terms of biodistribution, and taking into account those two parameters, CFLFC seems a better candidate than CRWEC, which seems to be better candidate than CNSQC.

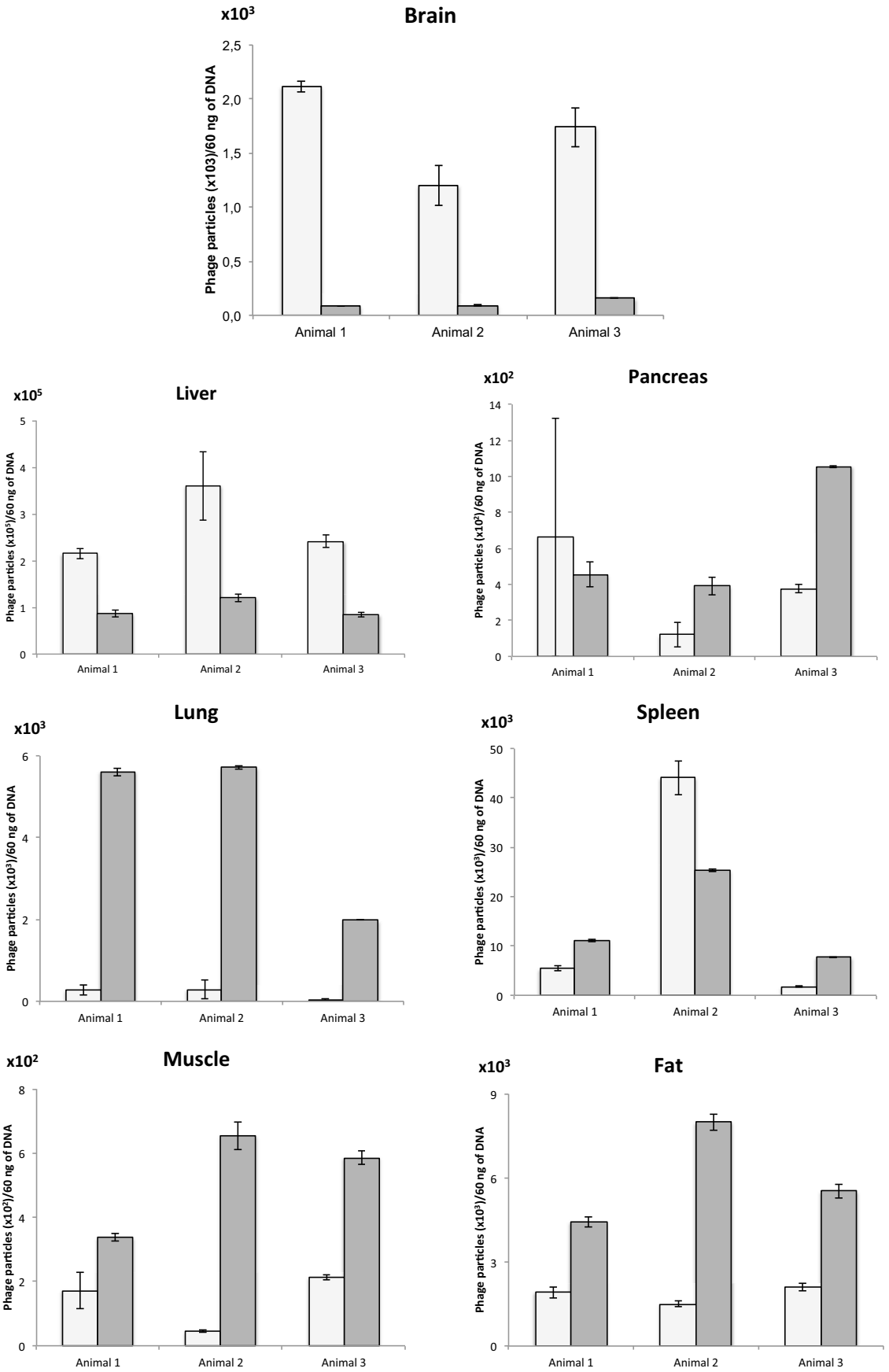


Figure 3.7. Figure showing biodistribution results obtained by q-PCR analysis. Blue bars correspond to CFLFC-Tet-phage and red bars to fd-Amp-phage (negative control not displaying insert) in all analysed organs. Results are expressed in "phage particles/60 ng of DNA" units. Light grey stands for CFLFC-phage, while dark grey for fd-Amp-phage.

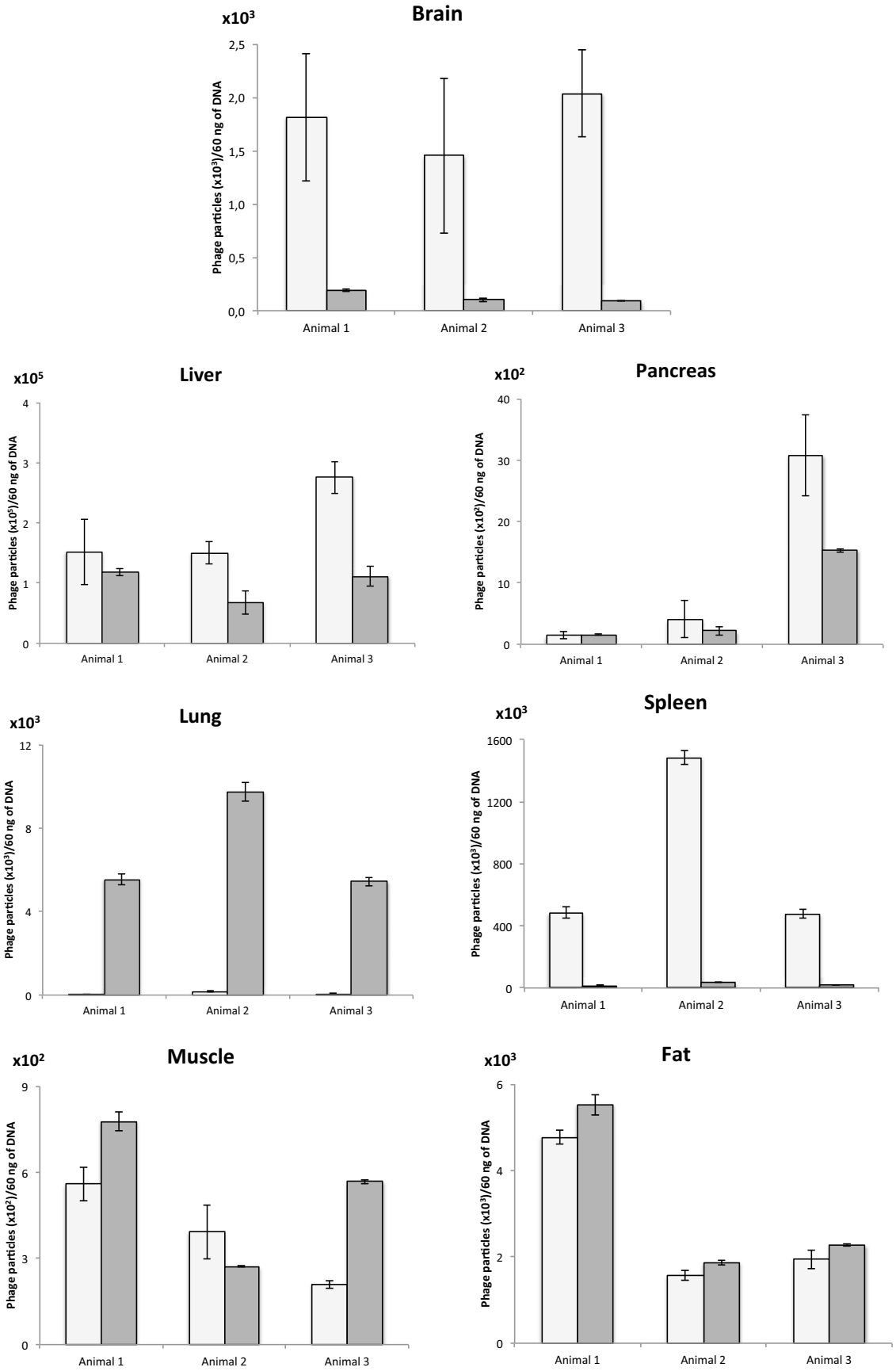


Figure 3.8. Figure showing biodistribution results obtained by q-PCR analysis. Blue bars correspond to CRWEC-Tet-phage and red bars to fd-Amp-phage (negative control not displaying insert) in all analysed organs. Results are expressed in "phage particles/60 ng of DNA" units. Light grey stands for CRWEC-phage, while dark grey for fd-Amp-phage.

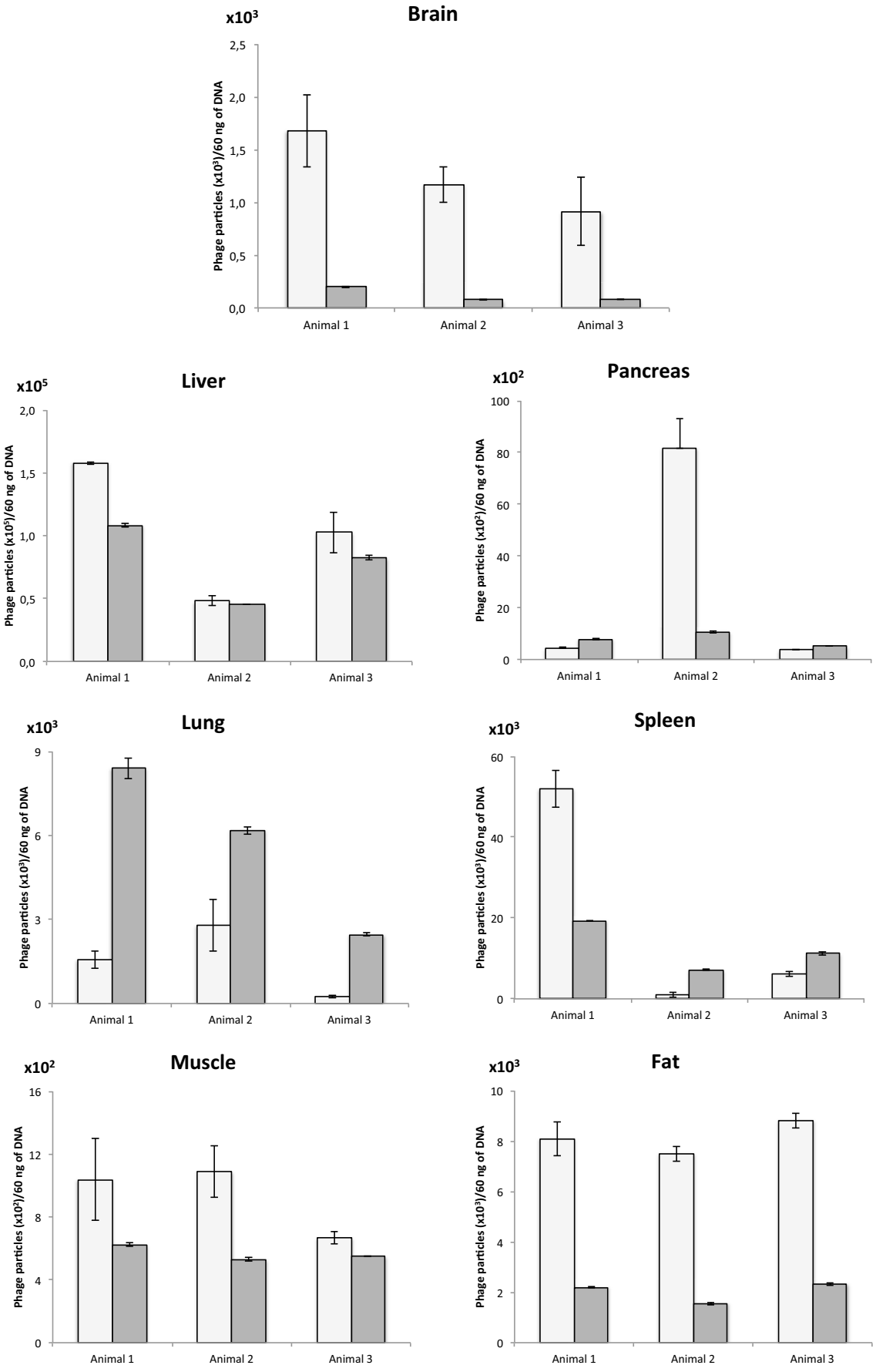


Figure 3.9. Figure showing biodistribution results obtained by q-PCR analysis. Blue bars correspond to CNSQC-Tet-phage and red bars to fd-Amp-phage (negative control not displaying insert) in all analysed organs. Results are expressed in "phage particles/60 ng of DNA" units. Light grey stands for CNSQC-phage, while dark grey for fd-Amp-phage.

3.5.2. Localization by immunofluorescence

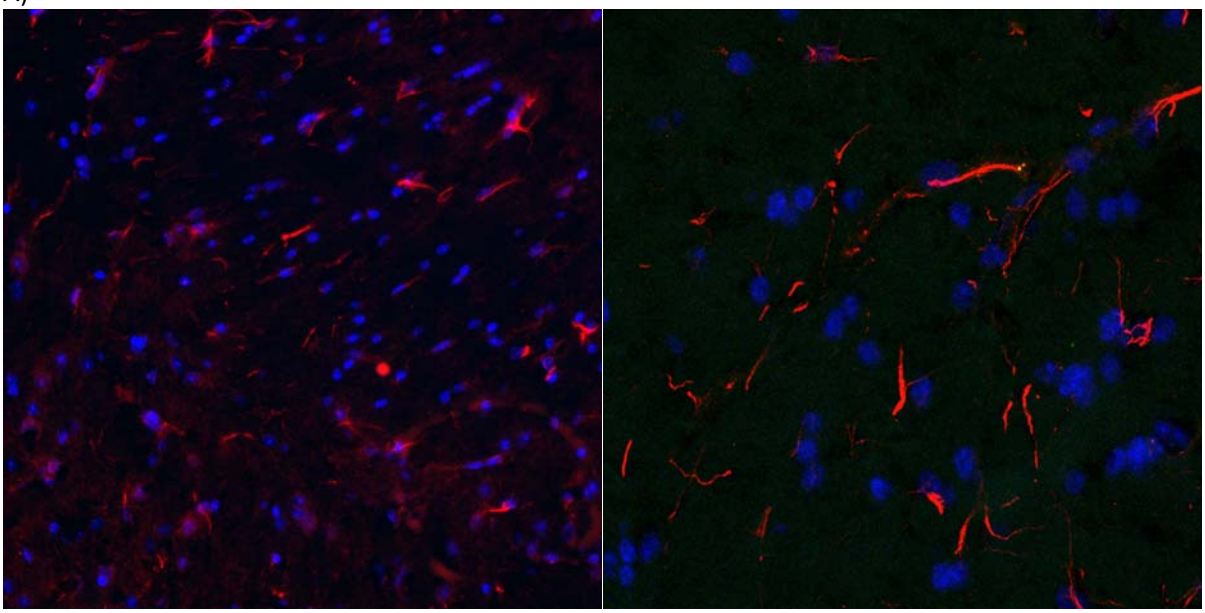
To track phage on the brain, immunofluorescence experiments of mice brains was performed after single phage injection. Each single phage was injected ($5 \cdot 10^9$ TU) to a BALB/c mouse (tail vein injection) for 6 hours. Mice were perfused and organs collected, fixed, sliced with cryostat and mounted for immunofluorescence staining.

Brain slices were immunostained with anti-phage antibody and anti-CD31, an endothelial cells marker to see whether phage co-localize with endothelial cells, meaning probably they are retained on the brain capillaries, or not. DAPI staining was also included to track cell nucleus.

Immunofluorescence experiments are a basic complement for biodistribution assay. While biodistribution determines amount of phage on each organ, it is not able to further localize phage position within the organ. Immunofluorescence experiments allow monitoring phage fate within the screened organ. Thus, although biodistribution experiments identify phage in the brain, this cannot be directly postulated as a BBB-shuttle since they might be stuck in the brain capillaries. As aforementioned, phage found in the brain can have three distinct fates. First, it may cross the BBB reaching the brain parenchyma. Second, it may internalize endothelial cells forming the BBB. Finally, it may interact with any receptor or protein in the surface of blood vessel not crossing to the brain parenchyma but getting stuck into the capillary. Only the first option fits the definition for a BBB-shuttle candidate. Then, combination of biodistribution and immunofluorescence experiments is a valuable tool to screen our lead peptides.

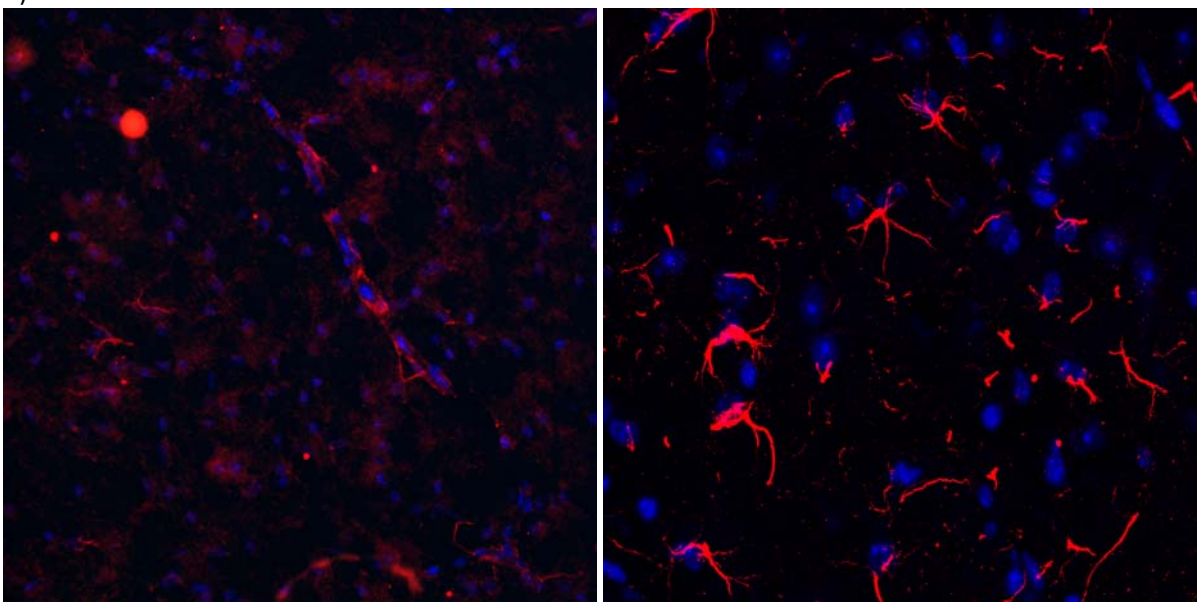
We failed on the attempt to determine co-localization since immunostaining with CD31 marker did not work properly probably due to poor conservation of the primary antibody. Then, co-localization was not possible to be assessed. However, since anti-phage antibody staining worked nicely, we took confocal microscopy pictures of the stained tissues, which clearly showed phage shapes localization.

A)



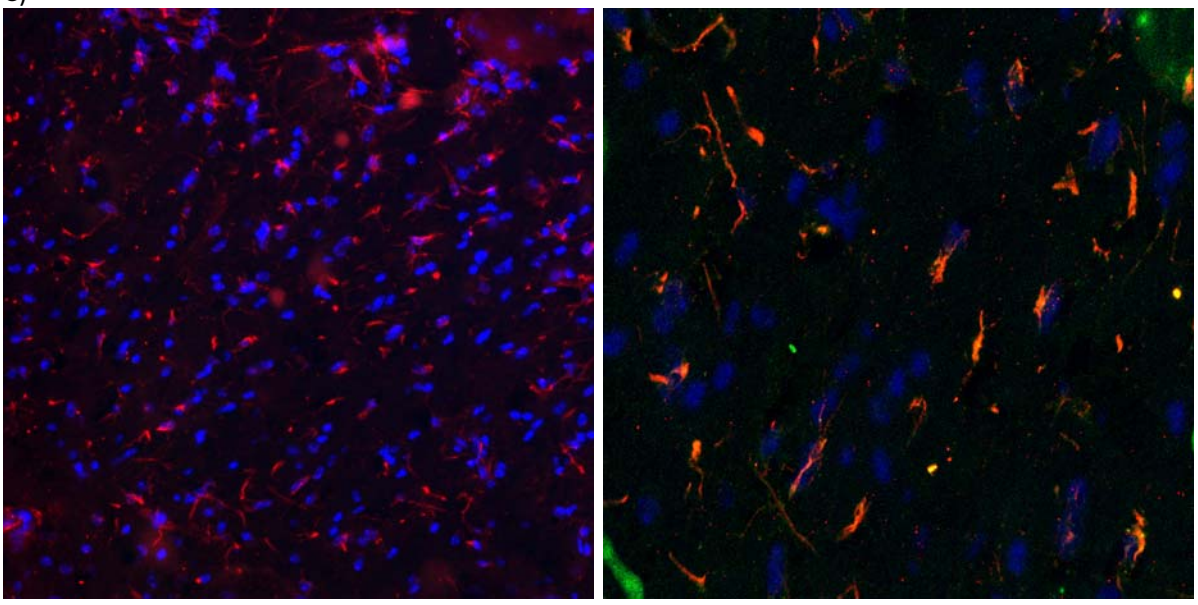
Low magnification High magnification

B)



Low magnification High magnification

C)



Low magnification High magnification

Figure 3.10. Confocal microscopy images from immunostained brain sections of mice injected with FLF-phage (A), RWE-phage (B) and NSQ-phage (C). Co-localization could not be observed since CD31 staining did not work properly in these experiments and was not acceptably observed on confocal microscopy (green). Red signals correspond to anti-phage antibody staining. Blue spots correspond to cell nucleus stained with DAPI while red traces correspond to localization of phage particles.

Phage shapes and its localization suggest phage are probably not stuck in blood vessels. DAPI staining of cell nucleus gives an idea of how big a blood vessel is. In case phage were stuck in the endothelial cells, all staining would match this vessel shapes (although brain slice would cut several of them). One example of very nice vessel was found in RWE.

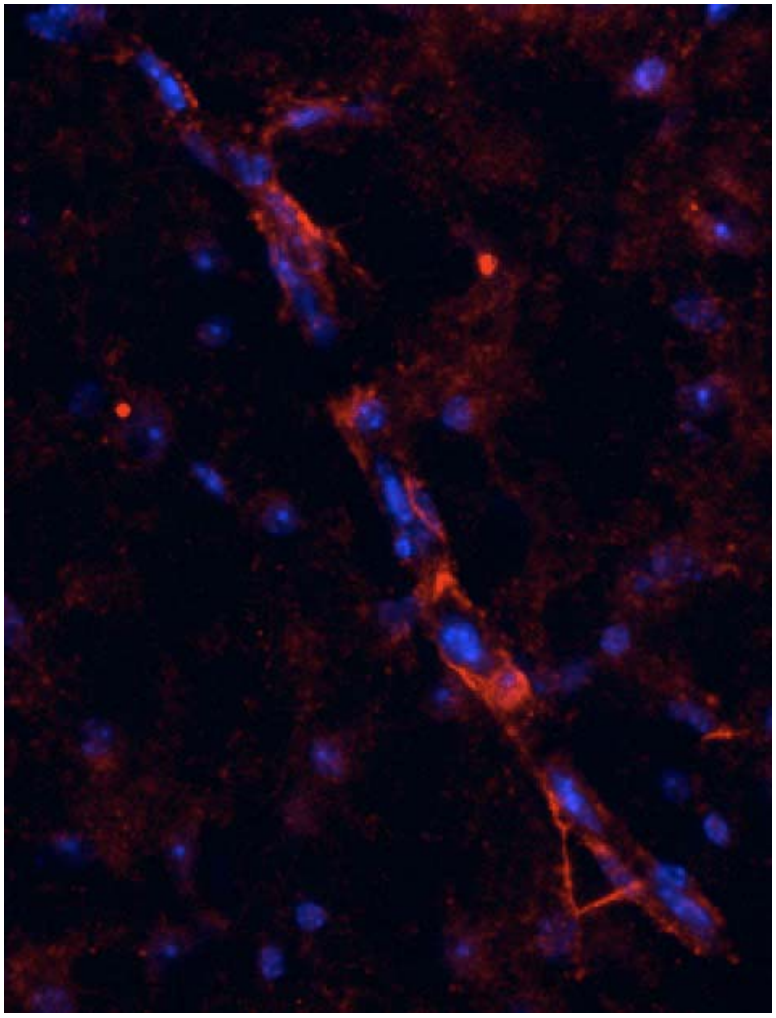


Figure 3.11. Low magnification confocal microscopy image of tissue sample of RWE-phage injected mice clearly presenting a blood capillary appearing in parallel to the slide cutting direction. Moreover, at the bottom-right region we can see a clear triangular staining corresponding to phage particles clearly out the brain capillary. We hypothesize this shape could correspond to astrocytes or perycites. Anyway, additional co-localization experiments with astrocytes and perycites markers need to be performed to further validate this hypothesis.

However, in Figure 3.11 it can be appreciated phage signal (red) forming a triangle out of the vessel domain (area highlighted). This might be a hint that explored phage are able to cross the endothelial membrane and localize to astrocytic feet on the astrocytes embedding brain endothelial cells forming brain capillaries.

All phage images show phage shapes most likely being localized in astrocytes than in endothelial cells. However, to further confirm this assumption, complementary immunofluorescence staining should be performed. At least two distinct staining experiments should be performed. First, we should repeat the endothelial cells co-localization with anti-phage antibody and any endothelial cell marker such as CD31. Second, perform co-localization experiments with other cells lines such as astrocytes, perycites or even neurons would be required to confirm phage localization in case they cross the brain capillary endothelial cells barrier.

3.6. Peptide synthesis of selected phage inserts.

To further validate the ability to target the brain of those selected peptide inserts, peptides were synthesized by solid-phase peptide synthesis with Fmoc/*t*Bu strategy. In order to obtain *C*-terminal amide peptides, Rink AM resin was chosen. The synthesis was performed at a 200 µmol scale. Oxima®, COMU and DIEA were used for coupling reactions. All peptides were *N*-terminus acetylated. Coupling reactions were monitored by the ninhydrin Kaiser test,[6] the chloranil test[7] or the *p*-nitrophenyl test (De Clercq test).[8] Peptides were cyclized on solid-phase mediating an intramolecular disulphide bond by means of two cysteines (Trt protected) constraining the selected sequences. The cyclization reaction was performed by addition of I_2 in DMF.[9-11] Reaction of iodine with the divalent sulphur atom leads to an iodosulfonium ion which is then transformed to the sulfenyl iodide and the trityl cation. The disulphide bond is then formed by disproportionation of two sulfenyl iodides or by reaction between the electrophilic sulphur atom of R^1-S-I and the nucleophilic S-atom of a second R^1-S-Trt molecule as shown in Figure 3.12.

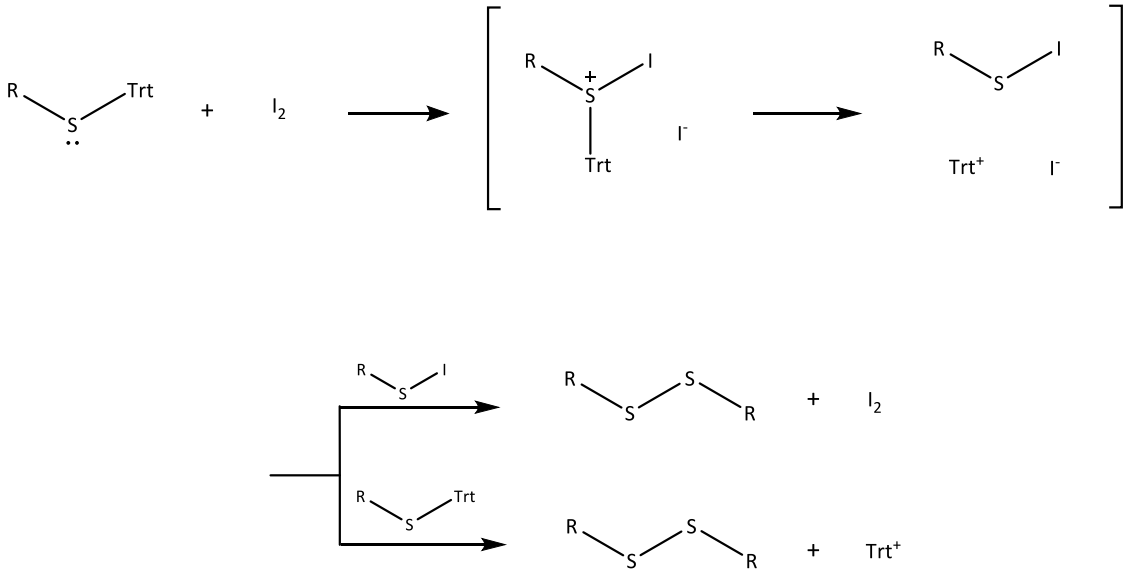


Figure 3.12. Proposed mechanisms for disulphide bond formation of S-Trt cysteines. On one hand, the mechanism based on a disproportionation reaction of two sulfenyl iodides. On the other hand, polar reaction.

This on-resin oxidation greatly facilitates the removal of the excess of reagent and soluble side-products by simple filtration and extensive washing to remove.

Cleavage of the peptide from the resin and complete removal of side-chain protecting groups was accomplished by acidolytic treatment with a TFA solution containing appropriate scavengers: TFA/TIS/H_2O (95:2.5:2.5). Crude products were analysed by reverse phase high-performance liquid chromatography (RP-HPLC), coupled either to a UV or mass spectrometry (MS) detector and by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF). Peptides were purified by RP-HPLC at semi-preparative scale and the pure product was fully characterized by RP-HPLC, HPLC-MS and MALDI-TOF. Peptides crudes were obtained with purities ranging from 35-60%. After HPLC semi-preparative purification peptides were obtained in purities higher than 95%.

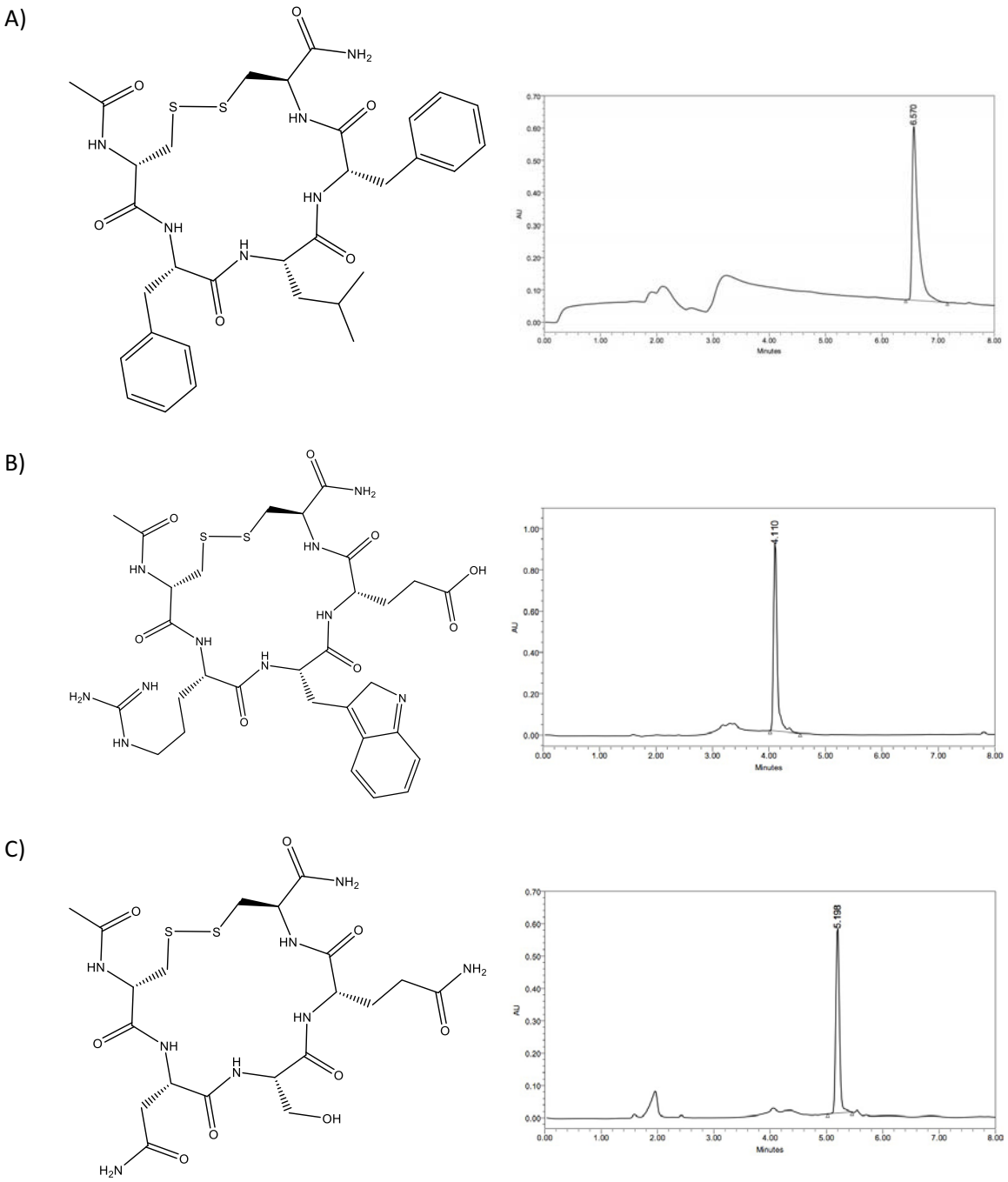


Figure 3.13. Chemical structures and HPLC chromatograms obtained after purification of peptide curdes: A) Ac-CFLFC-NH_2; B) Ac-CRWEC-NH_2; C) Ac-CNSQC-NH_2.

Competition experiments between peptides and phage displaying the same peptide insert should be performed to validate affinity. Competition would confirm that affinity is due to the presence of selected insert and not any other phage protein.

Other experiments such as fluorescent labelling of peptides would be desirable to further validate the ability of the peptide insert (with no phage) to target and eventually internalize the brain. This strategy requires fluorescent labelling of peptides during peptide synthesis and subsequent injection *in vivo*. Examples of that strategy can be found in the bibliography.[12]

Bibliography

1. Smith, G.P. & Petrenko, V.A. Phage Display. *Chem. Rev.* **1997**, 97, 391-410.
2. Arap, W., Kolonin, M.G., Trepel, M., Lahdenranta, J., Cardo-Vila, M., Giordano, R.J., Mintz, P.J., Ardelt, P.U., Yao, V.J., Vidal, C.I., Chen, L., Flamm, A., Valtanen, H., Weavind, L.M., Hicks, M.E., Pollock, R.E., Botz, G.H., Bucana, C.D., Koivunen, E., Cahill, D., Troncoso, P., Baggerly, K.A., Pentz, R.D., Do, K.A., Logothetis, C.J. & Pasqualini, R. Steps toward mapping the human vasculature by phage display. *Nat. Med.* **2002**, 8, 121-127.
3. Web site: http://www.biosci.missouri.edu/smithgp/PhageDisplayWebsite/PhageDisplayWebsiteIndex.html, 11th November 2013.
4. Triguero, D., Buciak, J. & Pardridge, W.M. Capillary Depletion Method for Quantification of Blood-Brain Barrier Transport of Circulating Peptides and Plasma Proteins. *J. Neurochem.* **1990**, 54, 1880-1888.
5. Boado, R.J. & Pardridge, W.M. A One-Step Procedure for Isolation of Poly(A)+ mRNA from Isolated Brain Capillaries and Endothelial Cells in Culture. *J. Neurochem.* **1991**, 57, 2136-2139.
6. Kaiser, E., Colescott, 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.
7. Christensen, T. Qualitative test for monitoring coupling completeness in solid-phase using chloranil. *Acta Chem. Scand. B. Org. Chem. Biochem.* **1979**, 33, 763-766.
8. Madder, A., Farcy, N., Hosten, N.G.C., De Muynck, H., De Clercq, P.J., Barry, J. & Davis, A.P. A novel sensitive colorimetric assay for visual detection of solid-phase bound amines. *Eur. J. Org. Chem.* **1999**, 1999, 2787-2791.
9. Kamber, B. & Rittel, W. Eine neue, einfache Methode zur Synthese von Cystinpeptiden. *Helv. Chim. Acta* **1968**, 51, 2061-2064.
10. Kamber, B., Hartmann, A., Eisler, K., Riniker, B., Rink, H., Sieber, P. & Rittel, W. The Synthesis of Cystine Peptides by Iodine Oxidation of S-Trityl-cysteine and S-Acetamidomethyl-cysteine Peptides. *Helv. Chim. Acta* **1980**, 63, 899-915.
11. Goodman, M., Felix, A., Moroder, L. & Toniolo, C. Synthesis of peptides and peptidomimetics. Georg Thieme Verlag, Stuttgard; **2003**. pp. 106-108.
12. Bertrand, Y., Currie, J.C., Demeule, M., Regina, A., Che, C., Abulrob, A., Fatehi, D., Sartelet, H., Gabathuler, R., Castaigne, J.P., Stanimirovic, D. & Beliveau, R. Transport characteristics of a novel peptide platform for CNS therapeutics. *J. Cell. Mol. Med.* **2010**, 14, 2827-2839.

Chapter 4

Beyond phage display limits: Mass spectrometry tools for identification and quantification of chemically synthesized peptide libraries across BBB models

4.1. Scope

High-throughput screening techniques had been widely used to explore huge amounts of molecules to address distinct aims such as drug discovery or drug delivery.[1-6]

Generally speaking, three main focuses must be defined to describe any high-throughput screening technique: library of choice, model to test our library to find hits and an appropriate identification tool. Each focus will have a strong influence on others.

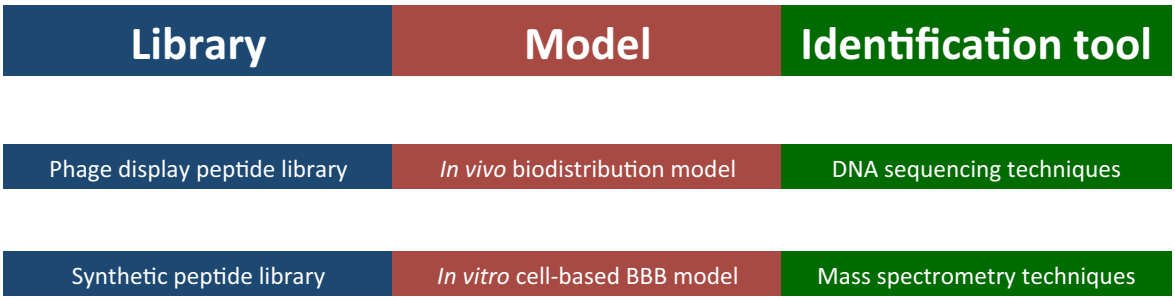


Figure 4.1. In the first row, key focuses describing any high-throughput screening technique are shown. In the rows below, specific focus on HTS methodologies used on the present thesis are disclosed.

In this chapter, we discuss the setting up of a new high-throughput screening technique combining peptide library synthesis, *in vitro* BBB models and advanced mass spectrometry techniques. We also intended to extend this very same methodology to *in vivo* mice models

although we conclude some library features should be redesigned to better face this approach, as will be discussed below.

4.2. Peptide library design and synthesis

We designed a peptide library on the basis of both synthesis feasibility and mass spectrometry detection.

4.2.1. Synthetic methodologies for library construction

Mix and split was the method of choice for peptide library synthesis. Described independently by some research groups, mix and split methodology is performed on solid-phase peptide synthesis.[7-9]

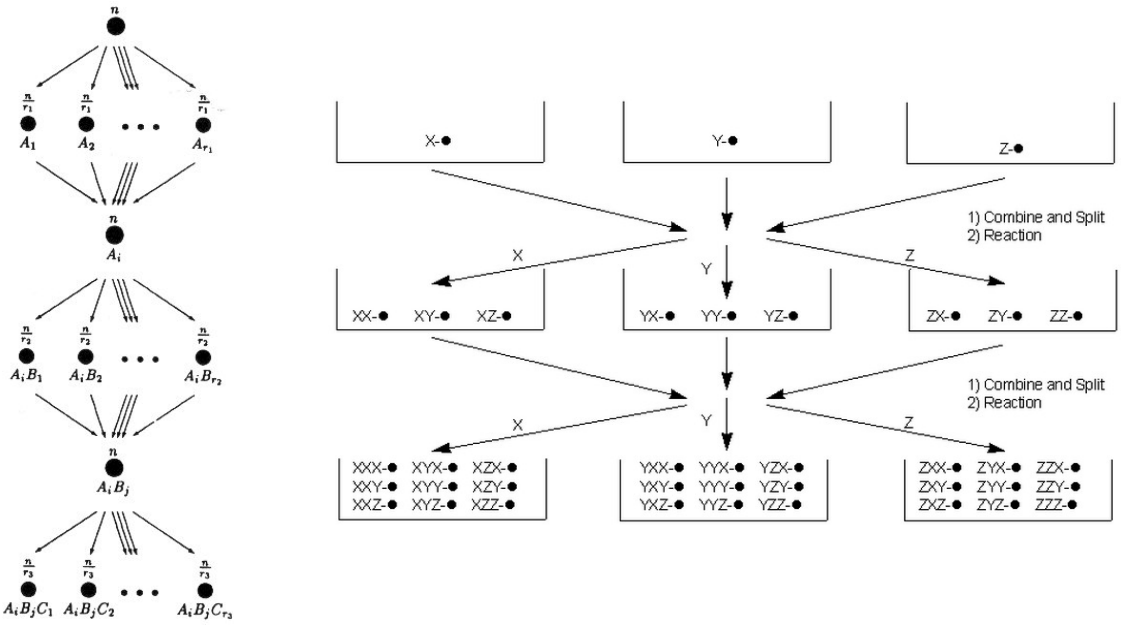


Figure 4.2. Mix (or combine) and split synthetic methodology scheme. This methodology is an iterative synthetic protocol that allows introduction of variability on a peptide. By equally dividing the polymeric resin beads into reactors, a distinct building block can be safely coupled to completion on each reactor. Resin is mixed again and deprotection step is usually performed in the initial pool. This workflow is repeated as many times as required to achieve a large variety of highly redundant peptides. Libraries synthesized with this methodology are usually called one-bead one-compound libraries since at the end of the synthesis (before cleavage), each resin bead carries several copies of a single peptide version.

General protocol begins by equally dividing polymer beads into a series of reaction pools. Resin on each vessel is treated with a single amino acid allowing coupling reaction to go to completion. Then, resin beads from all vessels are repooled. On each mix and split cycle, amino acid variability is introduced. Consecutive mix and split cycles and constant amino acid variability introduction leads to peptide libraries with m^r members, where m is the number of splitting, coupling and mixing steps performed and r the number of building blocks used to introduce variability on each step of the synthesis.

There are other methods to synthesize analogous peptide libraries, such as the use of equimolar mixtures of amino acids or equireactive mixtures.[10] However, equimolar mixtures can lead to biased coupling rates and determining equireactivity can be problematic. Several factors such as coupling reaction thermodynamics and kinetics make these calculations quite unreliable. Furthermore, amino acids are randomly coupled in reactive positions of resin beads. It can be expected that more accessible reactive positions will be filled by the most favoured reactions. This

may hinder and weaken less favoured reactions leading to bias coupling reactions and misrepresented amino acids and peptide sequences within the library.

Mix and split methodology allow recoupling whenever it is necessary without affecting amino acid variability. This feature makes this methodology very convenient for synthesize high-quality libraries.

The uniqueness of mix and split methodology is the synthesis of one-bead one-compound libraries. A unique peptide sequence is synthesized on each resin bead. However, our screening assays require peptides to be cleaved from polymer beads, thus not taking advantage of one-bead one-compound concept.

One-bead one-compound library synthesis require at least as many resin beads as sequences are planned to be synthesised. However, steps performed on mix and split methodology inevitably introduces an experimental error, especially in the split step. So that, larger number of resin beads will be required to cover all peptide sequence range.

Hurdle 1. Number of beads (n) required for reliable mix and split library synthesis.

Zhao *et al*. addressed this question from a statistical point of view.[11, 12] They described two criteria depending on the selected methodology to identify compounds activity from mix and split combinatorial libraries. On one hand, chemical-encoding strategies, where peptides are assayed still linked to resin beads; the peptide has to be present at least in one bead to be present in the library. On the other hand, on what they call iterative resynthesis methodologies, peptides are cleaved from the resin to perform the assay.

Ideally, in the later situation, each compound needs to be represented in equimolar amounts for assay consistency. It was described in previous reports that true equimolarity requires more beads than is practical for combinatorial experiments.[12] Then, Zhao *et al*. discussed the limits for practical experimental execution.

It is assumed that all compounds will be present in nearly equal amounts. They set out to determine the sample size n required such that, with 99% confidence, each compound will be present on at least $(1 - L)\cdot 100\%$ of the ideal number of beads, where $0 < L < 1$. The closer to zero is L, the closer to equimolarity will be the mixture of peptides within the library. To perform biological assays we may want all compounds simply be present in concentrations large enough to exert independent measurable effects. They suggest to use $L = 0.7$. This value ensures that all compounds will be present to at least 30% of the ideal amount. Then, n is calculated as follows:

$$n = v\, z^2_{(0.01/R)}/L^2$$

where, v and z^2 are

$$z = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}$$

$$t = \sqrt{-2 \ln (p)}$$

$$p = 0.01/R$$

$$R = r_1 r_2 \dots r_m$$

$$v = ((r_1 r_2 \dots r_m) - (r_1 + r_2 + \cdots + r_m) + (m - 1)$$

where c_0 is 2.515517, c_1 is 0.802853, c_2 is 0.010328, d_1 is 1.432788, d_2 is 0.189269 and d_3 is 0.001308.[12] m is the number of splitting and mixing steps used in the synthesis. R is

the total number of generated compounds. r_j is the number of pools in the j^{th} splitting step ($1 < j < m$). Since we want to introduce the same variability on each amino acid coupling. $r_1 = r_2 = \ldots = r_m = r$. Then we have,

$$R = r^m$$

and

$$v = r^m - mr + (m - 1)$$

4.2.2. Size and variability: Library design

Mix and split methodology enables to chemically synthesize highly redundant peptide mixtures consisting on permutations of amino acids. Since phage display enables to screen similar peptide libraries (although displayed on phage) we focus our library design on a chemical space not available for phage display: non-natural amino acids. Hence, D-amino acids were used for library construction.

■ **Evidence 1.** Peptides selected from screening of the designed peptide library will be protease resistant and will not need further tuning to improve this specific pharmacokinetic property. ■

Phage display has been massively used for the screening of peptide libraries based on a variety of scaffolds introduced by specially designed oligonucleotide libraries ligated into DNA vectors.[13]

On certain synthetic peptide libraries, variability can be easily introduced during the synthetic protocol. However, synthetic parameters must be carefully designed to make the protocol feasible in a laboratory scale while introducing desired library complexity. Furthermore, in the present case, library design considers the use of mass spectrometry for identification.

The peptide library scaffold was simple to specially emphasize the screening focus on variability. Peptides within the library were *C*-terminus amidated and *N*-terminus acetylated, thus allowing to introduce the acetyl-based isotopically labelled tag described in section 1.5 ($CD_3{}^{13}CO$-). *C*-terminus and *N*-terminus are candidate positions for cargo coupling without affecting the overall shuttle transport capabilities. For this reason, acetylation and amidation of peptide-shuttle candidates was considered.

However, this design may worsen peptide performance in MS identification by restricting suitable places for charge stabilization on ionization process. Thereby, we introduced an arginine residue always located as the *N*-terminus amino acid in all peptides synthesized. On one hand, this ensures at least one position for charge stabilization in MS ionization. On the other hand, this singular Ac-D-Arg- fingerprint is extremely useful when identifying peptides from our library within complex mixtures comprising endogenous peptides from cellular or *in vivo* assays.

As mentioned, length (*m*) and amino acid variability (*r*) determine the number of peptides present in the library, $R = r^m$. Although the larger the library, the broader chemical space can be screened, synthetic experimental experience warns to restrict resin amount for handling feasibility. To assess library size, combinations of *r* and *m* values were evaluated and converted to amount of resin (g) required to synthesize the library.

Hurdle 2: Bead size and density calculation.

To assess library size we first review the resin density-value in terms of beads/cm^3. Commercial supplier provides resin characterization data. Purchased batch of Rink Amide-AM resin had a resin density (ρ) of 0.6 g/cm^3 and a diameter size ranging from 100-200 mesh.

Mesh scale units are used to determine the particle size distribution of a granular material. 100-200 mesh resin indicate that resin beads will pass through 0.149 mm sieve (larger pieces won't fit through this mesh) but will be retained by 0.074 mm mesh (pieces smaller than this will pass through the mesh), respectively. This description establishes a range of particle sizes comprised between 0.149 and 0.074 mm.

To calculate the number of beads (n) that fit a volume (cm^3) we made some assumptions.

■ **First assumption.** It was considered that resin beads are ideal spheres which volume is:

$$V = \frac{4}{3}\,\pi\, r^3$$

Packaging density (η) is the fraction of volume filled by spheres. Packing density depends on spheres arrangement. The packing densities for several types of spheres packings had been described: loosest possible (0.0555); tetrahedral lattice (0.3401); cubic lattice (0.5236); hexagonal lattice (0.6046); random (0.6400); cubic close packing (0.7405) and hexagonal close packing (0.7405).[14] ■

■ **Second assumption.** Random packaging density (η_{random}) of 64% was presumed.[15] ■

Then, for density calculation, 0.0745 mm (0.149/2 mm) was considered as the sphere radius. The volume of a single sphere was calculated. Finally, density of beads per cm^3 was calculated:

$$d = \frac{1}{V_{bead}}\,\eta_{random}$$

Resin required for starting the synthesis was determined:

$$Resin = \frac{n}{(d \cdot \rho)}$$

where ρ is density of the resin (g/cm^3). Evaluation of r and m values was performed to assess synthesis feasibility.

m	*r*	*R*	*L*	*p (99%)*	*t*	*z*	*v*	*n*	*Resin/g*
3	20	8000	0.7	$1.25{\cdot}10^{-6}$	5.2139	-4.7080	7942	359254.3	1.6
3	19	6859	0.7	$1.46{\cdot}10^{-6}$	5.1843	-4.6765	6804	303677.0	1.4
10	10	10000000000	0.7	$1.00{\cdot}10^{-12}$	7.4338	-7.0340	9999999909	1009751421694.2	4554499.7
6	10	1000000	0.7	$1.00{\cdot}10^{-8}$	6.0697	-5.6117	999945	64263704.5	289.9
5	10	100000	0.7	$1.00{\cdot}10^{-7}$	5.6777	-5.1991	99954	5513881.3	24.9
8	10	100000000	0.7	$1.00{\cdot}10^{-10}$	6.7861	-6.3609	99999927	8257451788.4	37245.4
8	7	5764801	0.7	$1.73{\cdot}10^{-9}$	6.3518	-5.9073	5764752	410551345.2	1851.8
6	7	117649	0.7	$8.50{\cdot}10^{-8}$	5.7062	-5.2292	117612	6563378.5	29.6
5	7	16807	0.7	$5.95{\cdot}10^{-7}$	5.3544	-4.8571	16776	807696.6	3.6

Table 4.1. Library size calculations set up for a variety of *m* and *r* values. All parameters required to assess the resin amount required to faithfully synthesize the peptide library are shown.

■ **Key decision 1.** Library size was finally defined by m = 5 and r = 7. Then, R = 16,807 is the total number of generated peptides. ■

■ **Key decision 2.** Seven D-amino acids were selected for library synthesis: D-Alanine, D-Arginine, D-Isoleucine, D-Glutamic acid, D-Serine, D-Tryptophan and D-Proline. ■

This reduced set of amino acids was rationally selected considering physicochemical, synthetic and statistical parameters such as propensity of each amino acid to a given secondary structure, amino acid hydrophobic properties, and their presence in protein active sites.

Hurdle 3. Basis for choosing the set of amino acids.

This decision was based on a work developed by Dr. Ignasi Belda in Prof. Giralt's lab.[16] Physicochemical parameters were considered to make a rational-based selection also regarding on synthetic performance of amino acids to minimize coupling issues. A summary of physicochemical and statistical parameters is disclosed in Table 4.2.

Second column in the table shows each amino acid tendency towards a secondary structure: α-helix and β-sheet.[17] Symbol ≈ indicate a similar value for α-helix and β-sheet structures; ≈β a medium tendency to β-sheet; and α and β high tendency for α-helix and β-sheet, respectively.

In the third column it is shown the hydrophobicity of selected amino acids in Kyte-Doolittle scale.[18] Selected amino acids include hydrophobic, hydrophilic and neutral amino acids.

Fourth column shows the percentage of amino acid appearance in an active site about the total number of selected amino acid in the sequence.[19] Alanine value is not known since Alanine scanning was the methodology to calculate this parameter. Amino acidic base includes 3 top amino acids on this parameter: Trp, Arg and Ile.

Furthermore, proline was included on this selections since is the only proteinogenic *N*-alkyl amino acid.

Amino acid	Secondary structure tendency	Hydrophobicity*	Presence in active sites (%)
Trp	≈β	11	21
Arg	≈	20	13.3
Ala	α	7	-
Glu	α	17	3.6
Ser	≈	10	1.12
Ile	β	1	9.6

Table 4.2. Summary of considered parameters for reduced amino acid set selection. *Hydrophobicity in Kyte-Doolittle scale.

Here, we briefly summarize most important contributions of all selected amino acids. Tryptophan is a fluorescent amino acid described to play an important role on many potential hot spots (hydrophobic, H-bonds, π-π, π-cation, and so on) being important for molecular recognition. Arginine is also described to have many potential interactions (electrostatic, H-bond) and playing a crucial role on molecular recognition; moreover having a pI = 10.76 displays a positive charge in physiological pH. It is also described as a crucial residue for peptides able to cross membranes. Alanine is a highly abundant residue in natural systems. Since alanine has a small side chain it might act as spacer for distinct interaction sites on peptides. Glutamic acid contributes on negative charges and

show better synthetic performance than aspartic acid. Serine is a polar non-charged amino acid, which behaves better on SPPS than threonine. Isoleucine is an aliphatic, β-branched amino acid. Its side chain is significantly bulky. It has a high tendency to β-sheet structure and has a higher presence in active sites than valine or leucine. Finally, proline is the only *N*-alkyl amino acid. It has a tendency to facilitate β-turns and poly-Pro produce particular structures. It has been described to play a crucial role on hydrophobic-hydrophobic interactions.

4.2.3. Mix and split library synthesis

Synthetic workflow consists on 5 rounds of mix and split. Each round involves division of resin from a large vessel into 7 smaller reactors, coupling of amino acid building blocks, Fmoc deprotection and mixing again in the large vessel. Once five residues were coupled, a final coupling step was performed in the large vessel to introduce D-Arginine as the *N*-terminus amino acid in all peptides. Finally the whole library was split to perform acetylation with acetic acid or isotopically labelled acetic acid to achieve the light and heavy versions of the library.

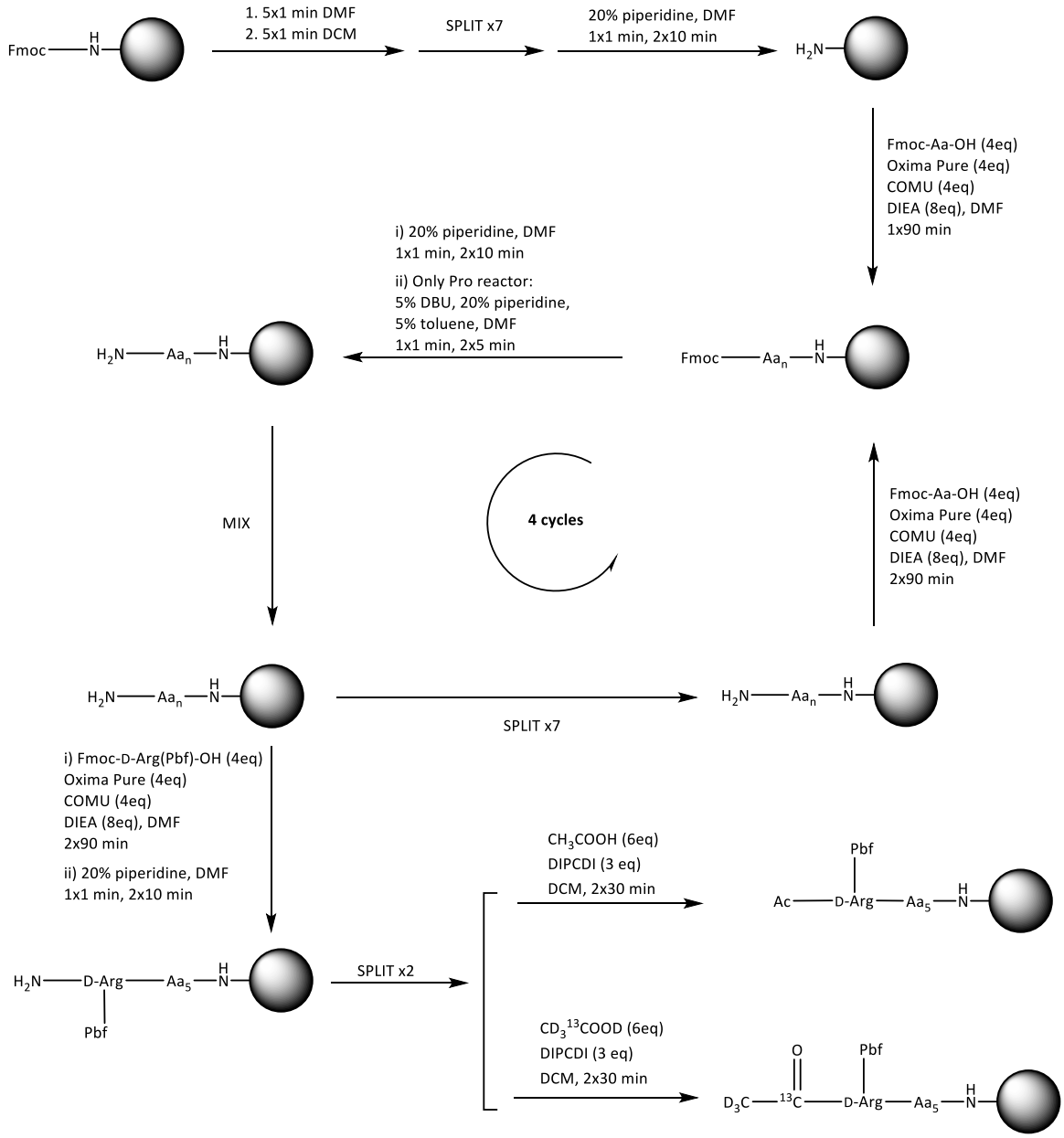


Figure 4.3. Ac-D-Arg-library synthetic workflow.

4.2.4. Library validation

Once library synthesis was performed, additional validation stage was carried out to monitor and validate library quality.

During library synthesis process, colorimetric tests used in regular SPPS were not performed. Basically two reasons drove this decision: first, colorimetric tests spend some beads that would slightly bias library variability; second, and most important, every bead contains a peptide with distinct synthetic complexity. Then, by choosing random beads we would never have a complete picture of the whole synthesis discouraging the use of these colorimetric tests.

Before whole library cleavage, single peptide analysis was performed by cleavage of single beads. Then, once library was cleaved from the resin, high-resolution mass spectrometry analysis was performed by direct infusion of the library on the FT-ICR MS.[20]

4.2.4.1. Single bead peptide cleavage and MALDI-TOF tandem analysis

Because library was synthetized using mix and split protocol, each resin bead should contain a single sequence in all its active positions. We randomly selected 15 beads and cleaved the peptides on them in independent vessels. Since peptides were acetylated in *N*-terminus Edman degradation was not possible for our library validation, so cleaved peptides were evaluated by tandem mass spectrometry. Tandem MALDI-TOF, MS/MS, was the chosen technique as it has far enough sensitivity for the analysis carried out. Peptide fragmentation was induced by collisions with residual gas, and bond breakage mainly occurred through the lowest energy pathways, this is cleavage of the amide bonds. This leads to *b*-ions when the charge is retained by the amino terminal fragment or *y*-ions when it is retained by the carboxy-terminal fragment.

Identified peptides must have the Ac-D-Arg-library fingerprint and only one sequence can be assigned to each spectra. Furthermore, peptides identified must fulfil appropriate peptide length (m=6) and exclusive presence of building blocks used for library synthesis. All cleaved peptides were manually assigned and disclosed in Table 4.3. As an assignation example, MS/MS spectra of Ac-RWPRER-NH_2 is detailed in the Figure 4.4.

ID	$[M+H^+]$	Sequence
1	786.49	Ac-RASEWP-NH_2
2	824.52	Ac-RIREIP-NH_2
3	770.51	Ac-RAIWSP-NH_2
4	687.44	Ac-RIAAES-NH_2
5	756.57	Ac-RERIAA-NH_2
6	901.62	Ac-RIIWWS-NH_2
7	843.57	Ac-RWPSWA-NH_2
8	828.48	Ac-RWEAEP-NH_2
9	1001.70	Ac-RWWWAR-NH_2
10	940.65	Ac-RWPRER-NH_2
11	839.72	Ac-RRWIPA-NH_2
12	943.63	Ac-RWWIEI-NH_2
13	886.54	Ac-RIEWEI-NH_2
14	739.49	Ac-RIPESP-NH_2
15	796.48	Ac-RIPWIA-NH_2

Table 4.3. Fifteen peptides sequences were identified after picking fifteen random beads from the reactor vessel. Peptides were cleaved from each bead and were individually analysed by MALDI-TOF MS/MS.

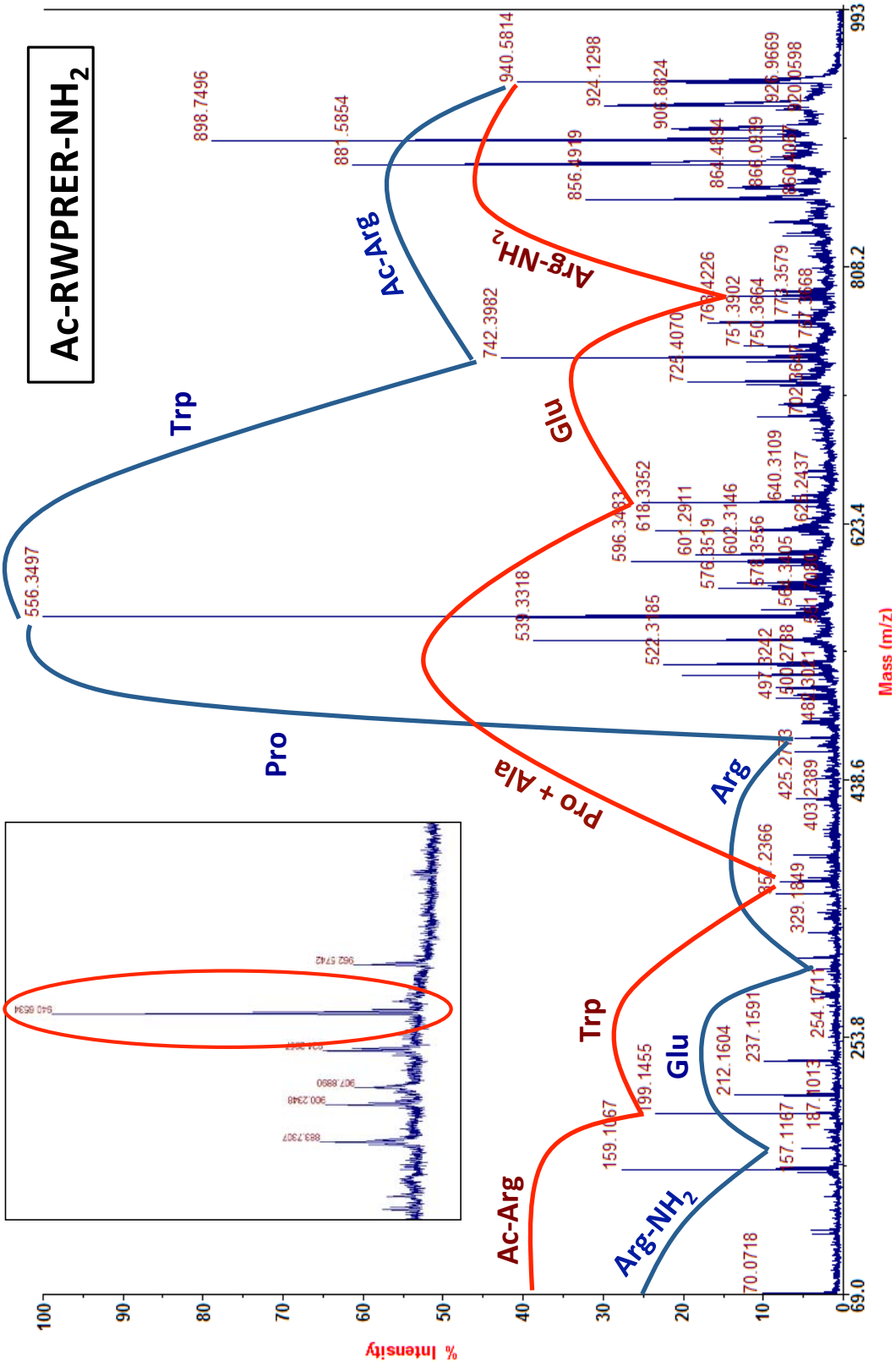


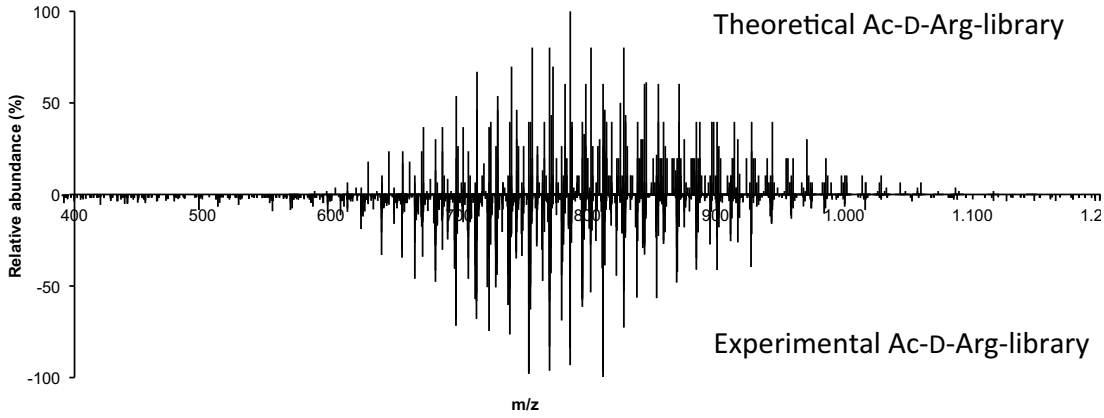
Figure 4.4. MALDI-TOF spectra of a specie with [M+H$^+$] = 940.65 is shown in the upper-left spectra. MALDI-TOF MS/MS spectra assignment is detailed for Ac-RWPRER-NH_2 as a case example, red corresponds to *b*-ions while blue corresponds to *y*-ions.

4.2.4.2. High-resolution mass spectrometry library analysis

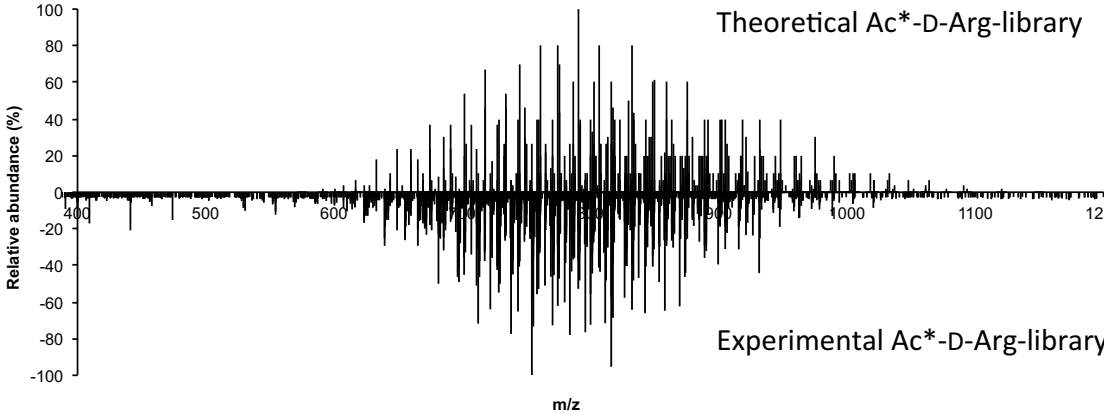
Combinatorial libraries can be characterized in detail by high resolving power mass spectrometry. Use of FT-ICR instruments for analysis of large combinatorial peptide libraries was early proposed by P. A. Demirev and R. A. Zubarev.[21] Some library synthetic validation attempts had been successfully reported.[22] Palmblad *et al.*[20] revised the use of mass spectrometry using state-of-the-art FT-ICR instruments with magnets up to 15 Tesla. Dynamic range is critical for direct analysis (no liquid chromatography coupled) of complex mixtures and depends quadratically with the equipment magnetic field.[23] Quality control on the library synthesis can be assessed by comparison between the predicted and measured library mass spectra. The correlation will never be perfect, but deviations are easy to detect. Parameters such as peptides with molecular weight lower or larger than the expected from the library design are extremely useful to evaluate the library quality. So, these techniques can be used to rapidly quality control combinatorial synthesis.

We used a FT-ICR with 7T equipment to assess quality control of our hexapeptide library. Experimental and theoretical spectra were compared in the Figure 4.5.

A)



B)



C)

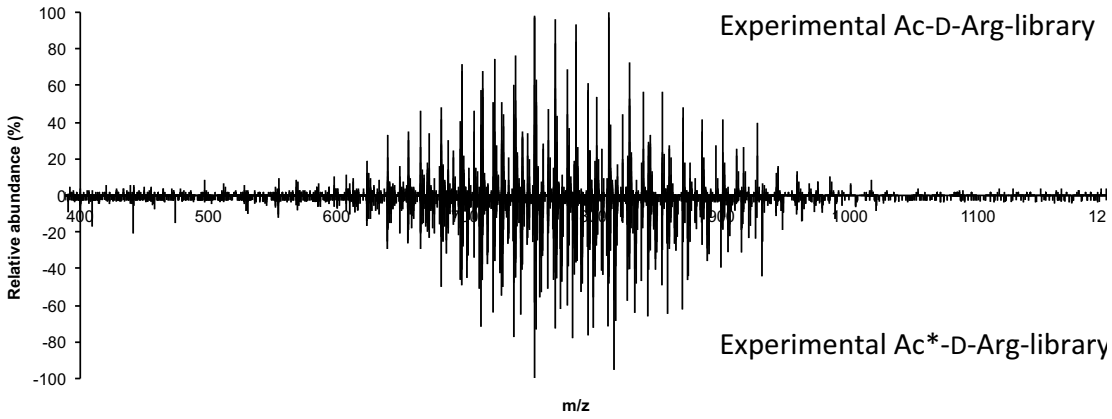


Figure 4.5. A) Comparison of relative abundance of the light-library peptides in theoretical and experimental spectra, up and upside down respectively. B) Comparison of relative abundance of the heavy-library peptides in theoretical and experimental spectra, up and upside down respectively. C). Comparison of relative abundance of the light-library peptides and heavy-library peptides in experimental spectra, up and upside down respectively. Ac* stands for $CD_3{}^{13}CO$-.

Experimental spectra were acquired by injection of a small aliquot of Ac-D-Arg-library dissolved in H_2O/MeCN (1:1). Theoretical spectra were built by assuming the same ionization and detection performance of all peptides in the library.

Although this quality control methodology is well established, to get the most of the technique libraries analysed should fulfil some requirements when being analysed by mass spectrometry. This is, for example, ionization ability of peptides within the library. Since peptides were acetylated and amidated, they might suffer from low ionization and thus poor MS detection. Despite having always an arginine residue located at the *N*-terminus amino acid position of all peptides in the library, experimental spectra may differ from the theoretical spectra where all peptides are considered to be equally ionized.

However, results showed nice agreement between theoretical and experimental library. Molecular weight of peptides in the library fit nicely molecular weight range predicted in theoretical spectra thus not evidencing synthetic problems.

4.3. Identification approach: From a general overview to key decisions

Mix and split libraries complexity lies on peptide similarities. To analyse such mixtures, combination of mass spectrometry techniques need to be considered.

As was already described, Ac-D-Arg-library is comprised by hexapeptides. In all peptides there is an arginine always located at the *N*-terminus. Mix and split method performed throughout the synthesis in 5 remaining residues leads to a combinatorial library comprising all scrambled versions involving the chosen amino acids.

$$Ac - \text{D-Arg} - X - X - X - X - X - NH_2; \qquad X = \begin{cases} \text{D-Ala} \\ \text{D-Arg} \\ \text{D-Ile} \\ \text{D-Glu} \\ \text{D-Ser} \\ \text{D-Trp} \\ \text{D-Pro} \end{cases}$$

Figure 4.6. General structure of the synthesized mix and split peptide library.

Thereby, we are facing a mixture of extremely similar peptides, which are particularly redundant and share similar properties. Therefore they are extremely difficult to differentiate. Still, it is a peptide mixture and the same basis of proteomic analysis can be applied on those mixtures.

Our aim was to be able to set up a methodology using mass spectrometry to identify specific sequences crossing *in vitro* assays by revising state-of-the-art mass spectrometry techniques for identification of synthetic peptide mixtures. By 1990s this approach was already envisaged but mass spectrometry tools were not as much developed as are nowadays.

A stringent assay such as *in vitro* BBB cellular assay should enormously simplify or at least reduce the amount of peptides within the library. Coupled chromatographic and mass spectrometric techniques are a widely used strategy to fractionate complexity of a mixture. RP-HPLC separation is based on differential partitioning between the mobile and stationary phases. Subtle differences in compounds partition coefficient result in differential retention on the stationary phase and thus defining the separation of each molecule in mixtures.

However, from such similar mixtures, large enough retention time differences to separate peptides by chromatographic techniques are not expected from each individual in RP-HPLC, even working with isocratic gradients. In many cases, similar molecules will be ionized and analysed in the mass spectrometer together. Mass spectrometry techniques provide *m/z* information of molecules ionized thus identifying all molecules with characteristic *m/z* signals on peptide ions spectra, also called MS1. High resolution and enough precision (10 ppm) displayed by this advanced techniques unequivocally permits to assign signals to molecular compositions.

Molecular masses mostly represent a single amino acidic peptide composition, but in some cases combination of pairs, trios or larger groups of amino acids may have the same overall mass (considering 10 ppm precision). So that, it is important to do not assume a same molecular mass implies one single amino acid peptide composition (from now on peptide family).

Peptide composition	a.m.u.	Peptide composition	a.m.u.	ppm
A2-E1	271.11681	S2-P1	271.11682	0.04
E2-I2	484.25330	A2-R1-W1	484.25463	2.75

Table 4.4. Table showing amino acid combinations sharing molecular weights not distinguishable with 10 ppm precision.

Each peptide family contains a diverse number of specific peptide sequences (family members) depending on its amino acid variability: from one single member when all amino acids are the same, to 120 members if all of them are different. Family members can be calculated following the combinatorial rule:

$$P_n^{n_1.n_2....n_n} = \frac{n!}{n_1!\,n_2!\,...\,n_n!}$$

where $n_1, n_2, \ldots n_n$ are element repetitions within the family and $n = n_1 + n_2 + \cdots + n_3$.

Families	Family member
5	1
4-1	5
3-2	10
3-1-1	20
2-2-1	30
2-1-1-1	60
1-1-1-1-1	120

Table 4.5. Number of family members among possible families comprised in the Ac-D-Arg-library.

Specific family members cannot be identified using solely MS1 spectrometric techniques. Advanced mass spectrometry techniques such as Linear Trap Quadrupole-Orbitrap (LTQ-Orbitrap) allow performing MS/MS analysis continuously with MS1 recording data. Then, we hypothesize why not to use this available tool in LTQ-Orbitrap (and also in several MS equipment) to elucidate specific sequences within Ac-D-Arg-library.

4.3.1. Approach inspired on bottom-up proteomic identification

Shotgun proteomics or bottom-up proteomics techniques rely on identifying proteins and mixture using a combination of high-performance liquid chromatography and mass spectrometry. This proteomics approach has been largely used (around 95%) over other techniques such as top-down proteomics, also focused on analyse complex protein mixtures.

Bottom-up approach involves enzymatic protein digestion, and peptide identification by database theoretical information cross correlation with acquired experimental MS/MS data.[24]

Several softwares had been developed to address protein identification, such as: SEQUEST,[25] Mascot,[26] MassMatrix,[27] OMSSA,[28] and X!Tandem,[29] among others.

SEQUEST proteomics software was a tool described in 1994 that automatically perform peptide/protein sequencing. The algorithm cross correlates experimental mass spectrometry fragmentation pattern of targeted peptides and MS/MS generated theoretical spectra selected from amino acids sequences in a protein database.

Although our peptide mixtures were not enzymatically digested, but was indeed an extremely complex mixture, we decided to initially explore this approach. We generated a "*.fasta*" database comprising all members of the Ac-D-Arg-library and we injected our sample to LTQ-Orbitrap.

Unfortunately, our library design unfit completely this workflow identification approach for some reasons which eventually lead us to discard this approach:

1. When performing MS/MS analysis, precursor ions with equal mass and *m/z* signal (members of the same family) will be eventually fragmented together resulting in complex and messy MS/MS spectra. This leads to chimeric spectra assignment because this MS/MS spectrum contains overlapped information from all individual MS/MS spectra corresponding to each precursor ion fragmented together. This would not be a problem in case only one member of the family crosses the biological assay. However, this scenario is unlikely.

2. Scrambled controls used as validation sequences for those above stated algorithms are indeed present within the library, thus disrupting typical validation strategies reported for protein identification.[30]

Due to those significant drawbacks, to verify output data of used algorithms we needed to manually confirm the assignation of all sequences found. In most cases we realized we could not ensure the assignation leading to very poor identification results (data not shown).

4.3.2. Proposed novel identification strategy

The identification strategy was directed through less standard protocols to elucidate targeted families rather than high-throughput screen all signals within the sample. This approach better suit redundant and complex mixtures such as those we are analysing. Although this approach entails loosing some amount of information on the way, the final acquired data is high quality. Since our aim points towards finding one or few excellent BBB-shuttles rather than map all molecules within the sample, this technique should fit nicely our expectations.

We designed an analytical MS strategy comprising two mass spectrometric levels.

4.3.2.1. First MS-identification level: Shotgun-like approach

First level of identification consists on injecting the sample into an LTQ-Orbitrap (coupled to LC) mass spectrometer. This first step only focus on MS1 spectrum data and shows a bright picture of the whole precursor ions (in this case should be called peptide ions or molecular ions because no latter fragmentation is performed) present within the sample.

Even though by using solely MS1 spectrometric techniques specific peptide sequences (family members) cannot be identified, MS1 spectra information allows identifying specific molecular compositions (peptide families) present in the sample. A *perl* programed computer program called *Bibliopepfinder*, developed by Xavier Arroyo in Prof. Giralt's lab, converts MS1 LTQ-Orbitrap output data to potential peptide families corresponding to MS1 spectra peaks.

See Appendix 1
Bibliopepfinder script.

Thereby, this first step results in LTQ-Orbitrap MS1 precursor ions list, which only contains information on the composition of peptides present in the sample. At this point, we focus all efforts to target some selected families since *second identification level* cannot face all mixture diversity at a time.

4.3.2.2. Second MS-identification level: Targeted mass spectrometry

In the next step, *second level of identification*, we use the well-established SRM technology, nowadays extensively used for targeted proteomic analysis.[31] For this purpose triple quadrupole (QqQ) equipment was used taking advantage on the unique sensitivity of this technique.

The combination of m/z setting for the first and third quadrupole of QqQ is referred to as transition. To filter the desired peptide parent ion, the m/z value of the first quadrupole is determined by the mass and the most abundant charge state of the target peptide. In the third quadrupole, one fragment ion (or daughter ion) is selected and thus a transition defined.

In contrast to classical proteomic SRM approach, which only monitors a reduced set of transitions from a selected number of peptides, our analysis aims to explore in depth all sequences from selected compositions (peptide families) by monitoring the complete transition set of each sequence.

Transitions corresponding to all *b* and *y* ions from singly and doubly charged precursor ion were monitored. Then, if enough transitions that define a sequence are found in the resulting spectrum, we can assure that this specific peptide sequence is present within the *m/z* LTQ-Orbitrap MS1 signal, and thus crossed the *in vitro* assay.

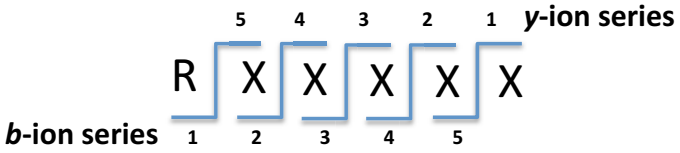


Figure 4.7. Peptide ion series monitored for all selected transitions defining a specific sequence. To ensure the presence of each targeted sequence signals confirming each transition must appear on SRM spectra.

QqQ equipment allows monitoring around 120 transitions per injection with a dwell time of 20 ms (further discussion on dwell times in section 4.3.2.3). Thereafter, few families can be screened because there is no enough sample to recurrently inject (approximately 4-5 injections could be performed per sample). For this reason, it is of paramount importance to have an appropriate or,

at least, rational selection criteria to choose the most promising families within MS1 LTQ-Orbitrap list.

Our understanding of the most promising families is to take the maximum benefit from our few injections by increasing chances to find positive sequences and from those, find the best ones. That means, for example, to try to diversify our options instead of gambling in only one sequence. Or betting in families present in all triplicates instead of those which are not.

Hurdle 4. Family selection.

We formulate two requirements all candidate families have to accomplish. First, families have to be present in all triplicates. Second, its molecular peak must fulfil typical mass spectrometry shape of peptide species in the MS1 LTQ-Orbitrap spectra: molecular and isotopic peaks must have a consistent profile and, on the other hand, it must be confirmed that the molecular peak is indeed a molecular peak and not an isotopic peak from another signal.

Then, we discuss two additional selection requirements that may lead to distinct families selection. To make the best decision, we should weight how these requirements affect the subsequent analysis. Here, we analyse their strengths and drawbacks.

■ **Key decision 3 (KD3).** Should we prime consistency among triplicates or should we shoot for the most intense family-peaks although being only intense in one triplicate?

- Option KD3-A: to shoot for the most intense family-peak although relative intensity is no consistent among samples might be advantageous in the sense we have more chances to see better and more intense transitions signals. However, family selection might be inconsistent due to unreliable performance of cell-based *in vitro* assay triplicates.
- Option KD3-B: to prioritize peaks which are consistently more intense makes the choice much more robust and avoids possible false positive results. Nevertheless, family selected signals might be not intense enough and transitions spectra poorer for sequence identification/validation.

Once weighted both options, option KD3-A was considered as the most appropriate because we can better measure up consistency on candidate selection to avoid false positive. In case transitions spectra would be too poor for any identification, we would shift to option KD3-B.

To display option KD3-A, we need a consistent way to weight up family ranking among triplicates. First, families not present in all triplicates are deleted from the candidate selection list. Second, all remaining families on each triplicate are ranked by their intensity values. We end up generating three lists, one for each triplicate. Finally, the most intense peaks from each list are candidates for the final ranking. Position on each triplicate list for each family candidate are added up to get a numeric score that rates its presence in weighted prominent positions. The lower the value, the better the candidate. (Examples are shown in section 4.4.3.3: Table 4.8 and Table 4.10.) ■

■ **Key decision 4 (KD4).** Should families with more than 30 members be considered and screened?

- Option KD4-A: Yes, it is worth trying to monitor as many molecules as possible. Monitoring families with several members is less transitions consuming (most of them are redundant) than some few-member families. Then, we can screen much more sequences with a single injection.
- Option KD4-B: They should be discarded because manually checking more than 30 spectra that may have several peaks is not realistic.

Option KD4-B was finally preferred. We consider that is especially convenient to focus our effort on screening a variety of families rather than restricted family candidates, which risks our finding possibilities. ■

4.3.2.3. Expectations on Ac-D-Arg-library behaviour on ionization and fragmentation processes

Ionization process is imperative for detection. Mass spectrometry techniques are blind to peptides or fragments that are not able to incorporate any charge. For this reason, our library incorporates an arginine residue (always being the *N*-terminus residue) to ensure all peptides within the library can be seen at least singly charged by MS. Other arginine residues might be found in peptides since arginine was one of the amino acids used for introducing variability into the library and thus other charges can be incorporated in those peptides.

Ac-D-Arg-library is *N*-terminus acetylated and *C*-terminus amidated. As stated by Z. Zhang the number of arginine residues present in the peptide suggests how many charges are expected for precursor ions. Though, it is important to point out that not all arginine residues have to be necessarily charged when the molecule is ionized. Still, small populations with higher charges than those expected cannot be discarded, although unlikely.

Then, Ac-D-Arg-library permits all peptides in the library to be (but not necessarily) ionized and protonated and thus seen in MS1 spectra. This means they are designed to have the chance to be protonated but it may happen that for ionization competition reasons some peptides would not be present in the spectra, although being present in the sample.

Similarly, all generated fragmentation ions may not be seen in MS/MS spectra. Collision induced dissociation (CID) is nowadays the gold standard fragmentation technique. CID peptide fragmentation mostly occurs on amide bonds, mainly producing *b* and *y* fragment ions.[32, 33] Among several fragmentation ions that could be monitored, we consider *b* and *y* ions as the most convenient for throughput screening analysis of members from selected families. Of course, for exhaustive analysis of each sequence *a*, *b*, *c* and *x*, *y*, *z* ions could be monitored to increase the chances to monitor each amino acid position within the sequence with other fragmentation techniques, but this would limit the amount of sequences screened because much more transitions would be spend on each sequence analysis.

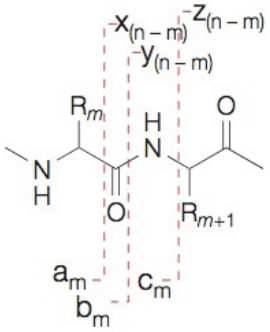


Figure 4.8. General backbone fragmentation pattern of peptides (Roepstorff-Fohlmann-Biemann nomenclature).[34, 35]

De novo prediction of peptide MS/MS spectra is an extremely challenging task, still not completely achieved. Several factors play a role usually difficult to quantify and predict. However, some successful attempts had been reported. Z. Zhang[36] developed a kinetic model based on the *mobile proton* model for peptide fragmentation prediction that quantitatively simulate the low-energy collision-induced dissociation (CID). Z. Zhang shown that the model is able to predict peptide CID spectra with reasonable accuracy in fragment ion intensities for both singly and doubly charged peptide precursor ions up to 2,000 atomic mass units.

Factors affecting SRM transition efficiency should be adjusted to achieve better sensitivity and accuracy results on QqQ. Those factors are mainly referred to ionization and fragmentation conditions.

First, ionization conditions rely on precursor charge state. Predominant charge state should ideally be targeted for better sensitivity. In the present case, this is addressed when selecting families on the first level of identification. However, ionization devices and experimental conditions (background, solvents or flow rate) can influence charge state distributions and intensities. Second, ion source parameters such as declustering potential or fragmentor voltage affect transition efficiency. Default parameters were used in performed QqQ experiments. Although this might not be optimal because those parameters are usually adjusted by default for proteomic experiments, we had no enough tools and sample amounts to optimize conditions for our Ac-D-Arg-library.

Among fragmentation conditions that should be considered, collision energy strongly influences fragmentation signals. By increasing collision energy, increasing amounts of precursor ions are fragmented until this phenomenon is compensated by secondary fragmentation effects. The optimal collision energy is in most cases linearly correlated with the precursor mass for a given charge state. However, large deviations can be encountered. As for ionization conditions, default fragmentation conditions were used.

Fragment ion type is another crucial parameter for better assign target peptides. As described above, CID is the fragmentation technique on QqQ. Within our library, particularly *y*-ions are unlikely to be seen as a whole series. On the other hand, *b*-ions series are more likely to clearly appear in the spectra, *b*-ions correspond to the *N*-terminal fragmentation series where an arginine is located and charge is expected to be present. However, in case there is another arginine residue in the peptide or any other charge stabilized elsewhere, *y*-ion series can be seen. So, we expect to base the assignments basically on *b*-ion series and complement them by *y*-ion series.

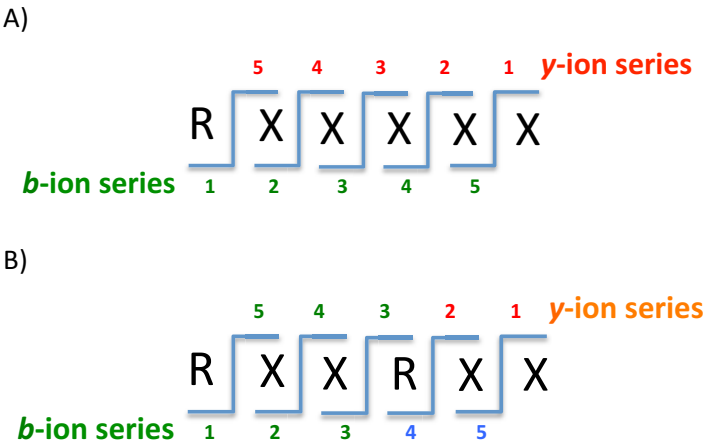


Figure 4.9. A) As shows the fragmentation pattern of a random peptide in the library (X = A, I, E, S, W, P, but not R). *b*-ion series is expected to be seen while *y*-ion series are not; B) shows the fragmentation pattern of a random peptide in the library (X = A, I, E, S, W, P, but not R) containing an extra R residue in position Aa_4. In this case, y_5, y_4, y_3 transitions from *y*-ion series are expected to be present in the spectra while y_2 and y_1 are not. *b*-ion series are still expected to be present in the spectre and b_4 and b_5 might be seen double charged.

This assignment reasoning completely differs from the one usually performed in proteomic approaches. Since bottom-up proteomic strategy involves use of enzymes (usually trypsin) to digest the sample, digested peptides contain an arginine or lysine in its *C*-terminus. Thus, positive charge is ensured on *C*-terminus end and *y*-ion series are usually crucial for assignment. Furthermore, peptide *N*-terminus is usually protonated and *b*-ions also likely to be found.

On SRM analysis a compromise must be considered between number of transitions analysed and signal sensitivity. High-sensitivity measurements rely on sufficient signal accumulation of target transition. In SRM mode, the QqQ instrument cycles through the list of transitions. Signal accumulation will depend on the time spend for the equipment to record each transition signal. This time is called dwell time and can be defined by the user. Ideally, 8 measurements should be performed per chromatographic peak. Then, we have to meet a compromise between dwell time and number of transitions according to cycle time, which depends on chromatographic peak width.

$$Cycle\ time = dwell\ time \cdot transitions$$

High-sensitivity requires long enough dwell time to accumulate desired signal. Dwell-time settings range between 10 ms for good sensitivity and 100 ms for excellent sensitivity. It must be noted that three extra milliseconds are spend on each transition change due to equipment settings.

The cycle time might be increased to analyse a higher number of transitions at a desired dwell time. However, in this scenario, accuracy would be diminished. If not enough data points are recorded, a poor chromatographic trace is obtained and peak reconstruction is biased. Neither identification nor quantification can be faithfully assigned.

Accuracy on SRM experiments is shown in Figure 4.10 taken from Lange *et al*.[37] For distinct dwell and cycle times settings it can be easily understood the importance of experiment accuracy and sensitivity.

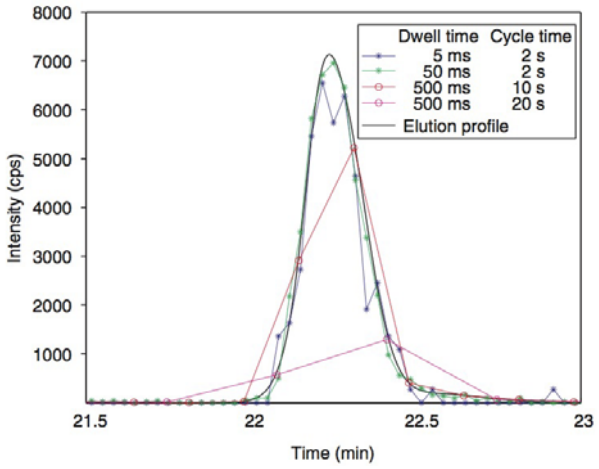


Figure 4.10. Illustration of Lange *et al.* on the effect of quantitative accuracy as a function of dwell time and cycle time. Although this effect is especially problematic on quantification it also does affect on identification accuracy.[37]

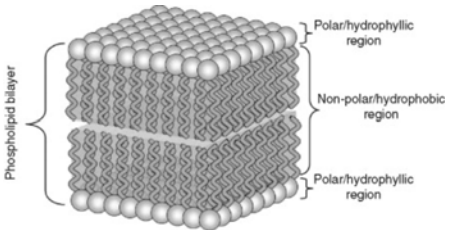
To avoid loss of sensitivity in experiments were large amount of transitions are measured in proteomics field, two approaches had been considered. First strategy consists on restrict the acquisition of particular transitions to a window around the elution time of the corresponding peptide. This strategy is called scheduled-SRM. It has been described that this strategy allows increasing the monitored transitions between 5-20 times. This will depend on gradient length, elution peak width and reproducibility achieved on chromatography.[38] Since monitored peptides in our samples showed extremely similar retention times, we did not considered using this strategy.

Second, dwell times can be individually adjusted depending on concentration and expected intensity of each targeted peptide and/or transition. Transitions monitored from concentrated peptides can support shorter dwell times, while lower concentrations peptides or low intensity transitions will require larger dwell times. However, in our case, since neither peptide concentration nor relative proportions among targeted peptides is not known this strategy was not contemplated.

4.4. Libraries evaluation

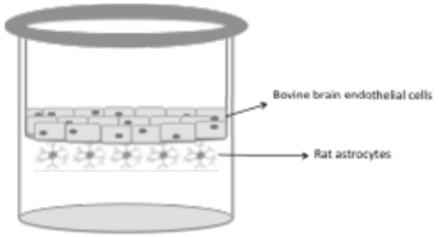
4.4.1 BBB-mimicking assays to test Ac-D-Arg-library

Parallel artificial membrane permeability assay (PAMPA)



***In vitro* cell-based BBB model**

Co-culture of
bovine brain endothelial cells
and rat astrocytes



***In vivo* assays**

BALB/c mice

Figure 4.11. Schematic illustration on BBB-mimicking assays used on the present thesis to test Ac-D-Arg-library.

4.4.1.1. Parallel artificial membrane permeability assay (PAMPA)

PAMPA assay basis has been already discussed in the Introduction (1.5.5).[39] A saturated solution of Ac-D-Arg-library was prepared in water (neither commercial pION PAMPA buffer nor cosolvents were used). Sample was sonicated to make sure solution was saturated and pH was adjusted to 7.0. To eliminate non-water soluble peptides, sample was centrifuged and precipitate was removed. PAMPA assay was run for 4 hours by triplicate. Acceptor (upper) and donor (lower) compartments and t_0 aliquot were kept for analysis. A blank sample was run in parallel.

4.4.1.2. *In vitro* cell-based BBB assay

In vitro BBB cell-based assay basis has been already discussed in the Introduction (1.5) and in section 2.4. A 10% of saturated solution of Ac-D-Arg-library was prepared in Ringer/HEPES buffer (buffer required for cell-based BBB model) and pH was adjusted to 7.0. To eliminate non-water soluble peptides sample was centrifuged and precipitate was removed. Lucifer yellow was added as a negative control. *In vitro* cell model was run for 2 hours at 37ºC by triplicate. Donor (apical) and acceptor (baso-lateral) compartments and t_0 aliquot were kept for analysis. TEER measures (previous to the assay) of brain endothelial cell *in vitro* membranes were 125 ± 18 $\Omega \cdot cm^2$. A blank sample was run in parallel.

4.4.1.3. *In vivo* assay: BALB/c mice

Discussion of *in vivo* BALB/c mice assays is discussed latter on this chapter (*see* 4.4.5). A 100% saturated solution of Ac-D-Arg-library was prepared in PBS buffer and pH was adjusted to 7.0. To eliminate non-water soluble peptides sample was centrifuged and precipitate was removed.

2% (w/v) of blue evans was added as a negative control.[40, 41] BALB/c mice were injected (tail vein injection) with 200 µL of sample. Mice were perfused after 6 hours of injection and organs and tissues collected: brain, lung, heart, pancreas, spleen, liver, kidney, muscle and fat. Organs and tissues were chopped and stored at -80ºC. A blank sample was run in parallel.

■ **Evidence 2.** Peptide library was always assayed in aqueous solutions with no cosolvent to improve peptide solubility. Whereas loosing a significant amount of peptides (those which are not soluble in aqueous medium) we are selecting those peptides soluble in aqueous media, which will not require any tuning step to improve their solubility properties; an important feature for any drug candidate. ■

Even so, injection in FT-ICR MS showed nice agreement on library profile when dissolved in H_2O/ MeCN (1:1) or dissolved in H_2O. By this means it can be concluded that few peptides within the library are loosen in this step.

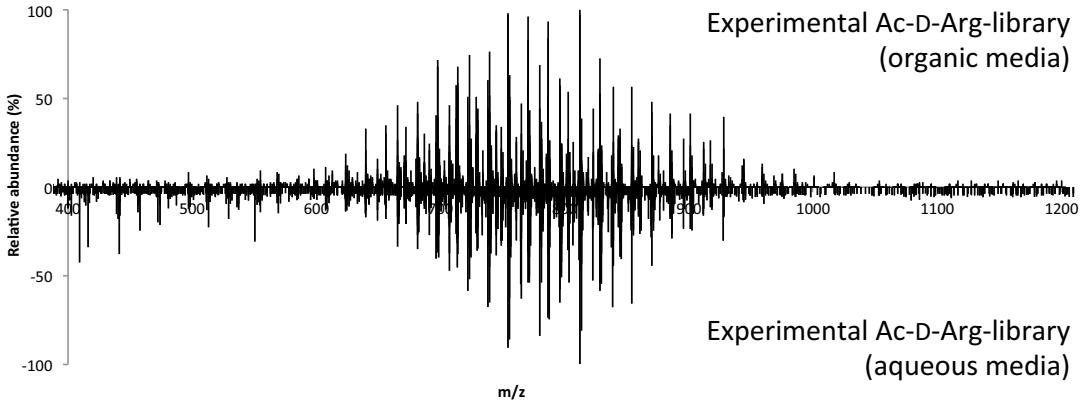


Figure 4.12. Comparison of relative abundance of peptide library dissolved in H_2O/MeCN (1:1) and in H_2O; up and upside down, respectively.

4.4.2. Sample cleaning

A cleaning protocol was required to clean up salts from *in vitro* cell-based assay samples, as well as endogenous cell peptides and proteins that would be harmful for analytic equipment performance.

Despite it might not be necessary for equipment caring (water instead of pION buffer was used for sample preparation); PAMPA assay samples were also cleaned with the same mini-C_{18} standard desalting protocol for sample consistent management.

Acceptor well (baso-lateral) and t_0 samples from *in vitro* cell-based BBB assay were desalted with mini-C_{18} columns with a standard desalting protocol. Peptides present in samples were retained within mini-C_{18} columns and water washes are performed to remove salts. Eluted peptides were ready for HPLC-MS analysis. Similarly, acceptor well and t_0 PAMPA assay samples were cleaned as *in vitro* cell-based BBB model samples.

In vivo assay samples (BALB/c mice) required further steps. Brain, as well as liver and kidney (taken as control organs), were sonicated with a lytic buffer (4% SDS, 0.1M HEPES, pH = 8.5) to lysate all cells and release all library peptides retained. Three different methods were reviewed to get rid out of as many proteins as possible to clean the samples:

- Microfiltration with 10KDa spinfilters (150 µL of sample per filter).
- Protein precipitation with acetone (5 volumes) at -20ºC o/n.
- Protein precipitation with methanol (5 volumes) at -20ºC o/n.

Finally, a standard desalting protocol was performed with C_{18} desalting columns.

Blank samples were cleaned using the very same protocol described for both *in vitro* cell-based BBB model, PAMPA assay and *in vivo* BALB/c mice assayed samples.

4.4.3. First MS-identification level: LTQ-Orbitrap

Samples were injected to LTQ-Orbitrap equipment. After injection, a list of all peaks recorded and their basic information (*m/z*, intensity, relative intensity and charge) was extracted from equipment software.

All data was evaluated with a *perl* programed computer program *Bibliopepfinder* developed by Xavier Arroyo in Prof. Giralt's group to filter peaks that might correspond to peptides present in tested Ac-D-Arg-library. Data was analysed considering an experimental error of 10 ppm.

Bibliopepfinder provides output information for all peaks potentially corresponding to peptides of Ac-D-Arg-library: theoretical mass, experimental mass, amino acid composition (family) matching *m/z* signals, a family general number code and specific family. For example, for Glu3-Ala1-Pro1-Arg1 family would appear as E3-A1-P1-R1 and family general number code would be 3111 referring to the number of distinct amino acids present in the family.

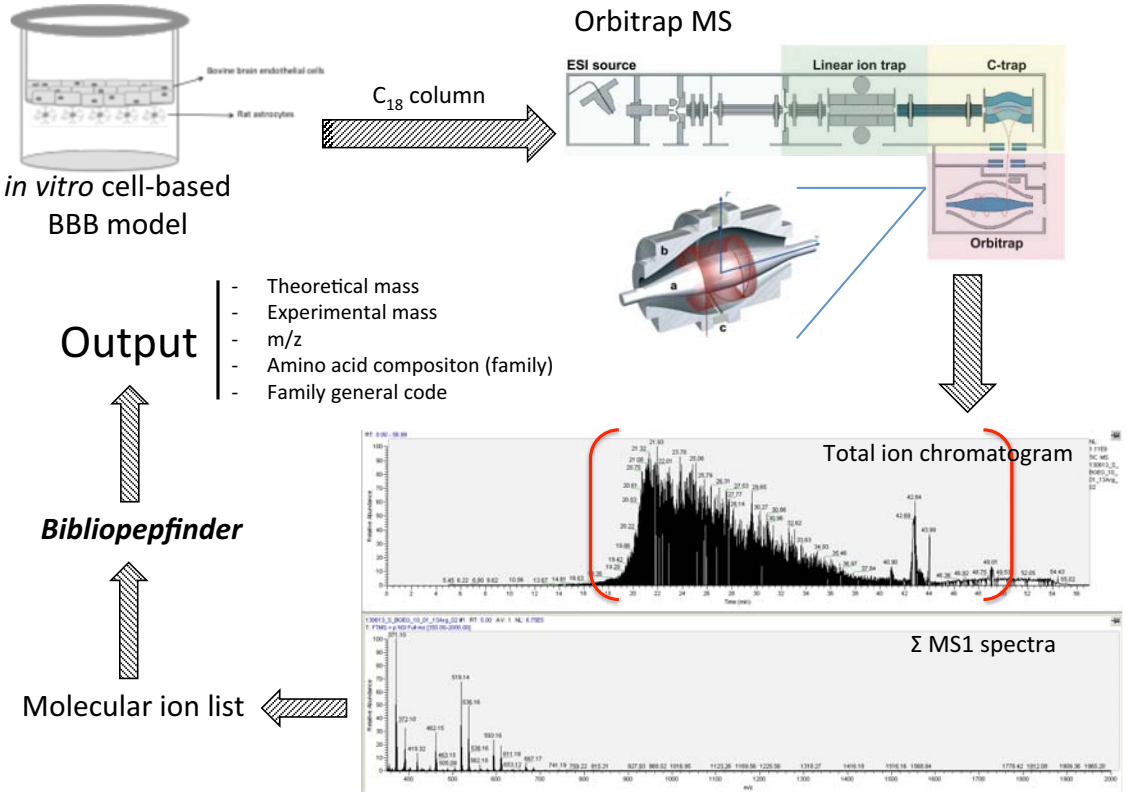


Figure 4.13. General workflow followed to obtain peptide compositions lists (peptide families). Molecular ions from all MS1 spectra recorded throughout the TIC chromatogram matching peptides masses within Ac-D-Arg-library are selected by *Bibliopepfinder*.

4.4.3.1. Parallel artificial membrane permeability assay (PAMPA)

All PAMPA families found in all triplicates were pulled and all replicates were removed. Then, signals matching library families found in blank sample (3) were subtracted from the former list. Finally, *Bibliopepfinder* assigned 187 families to 169 m/z signals in LTQ-Orbitrap spectrum, 151 m/z signals matched one single family each while 18 m/z signals could be assigned to two families each.

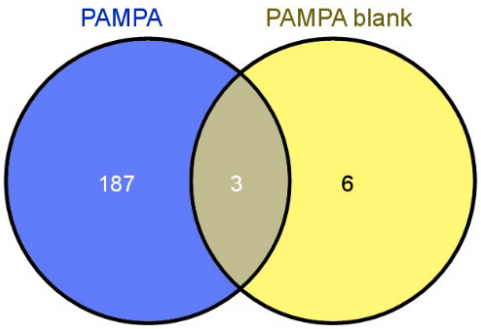


Figura 4.14. Venn diagram showing the number of overlapping sequences between Ac-D-Arg-library PAMPA assayed sample and blank PAMPA sample.

See Appendix 2
Ac-D-Arg-library families found in PAMPA assay.

4.4.3.2. *In vitro* cell-based BBB assay

LTQ-Orbitrap-analysed *in vitro* cell-based assay triplicate samples were pulled and blank signals (10) matching possible families from the library were deleted. *Bibliopepfinder* assigned 367 families to 335 m/z signals in LTQ-Orbitrap spectrum, 304 m/z signals matched one single family each while 30 m/z signals could be assigned to two families each and one to 3 distinct families.

From all those families assigned by *Bibliopepfinder*, 78% were present in all triplicates, while 15% were present in two triplicates and 7% to only one single triplicate.

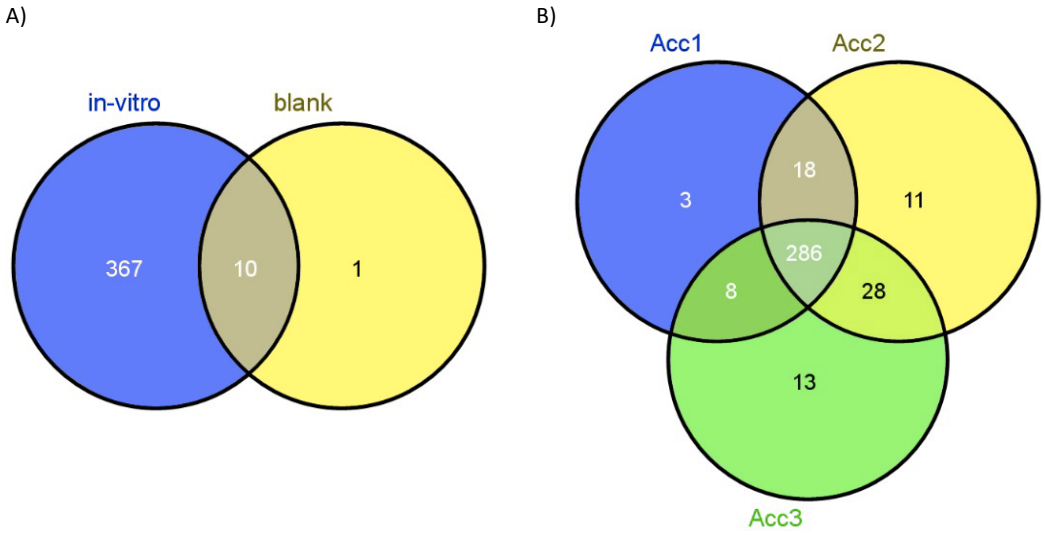


Figura 4.15. A) Venn diagram showing the number of overlapping sequences between Ac-D-Arg-library cell-based *in vitro* BBB model assayed sample and blank sample. B) Venn diagram showing the number of overlapping sequences of Ac-D-Arg-library among cell-based *in vitro* BBB model triplicates.

See Appendix 3
Ac-D-Arg-library families found in cell-based in vitro *BBB model.*

Two interpretations related with cell-based *in vitro* BBB model can be stated from this data analysis. First, it seems to show high reproducibility from well to well in a plate. Second, finding 286 over 462 possible families within Ac-D-Arg-library (62%) probably shows still too low tightness of endothelial cell layer forming the *in vitro* cell-based BBB assay.

4.4.3.3. PAMPA assay vs *in vitro* cell-based BBB assay

At this point, an extra selection criterion was considered. Since PAMPA assay only mimics passive diffusion properties of the BBB, all families detected in PAMPA samples were hypothesized to cross the BBB through this mechanism. On the other hand, *in vitro* cell-based BBB model should mimic both active transport and passive diffusion properties of the BBB. Then, it can be hypothesized that families both present in "PAMPA list" and in "*in vitro* cell-based BBB model list" cross by passive diffusion, while those only present in "*in vitro* cell-based BBB model list" cross by any other mechanism (either active transport related or paracellular flux). Thus, two lists were generated, on behalf of hypothetical *passive diffusion permeability* and *other mechanisms transport*.

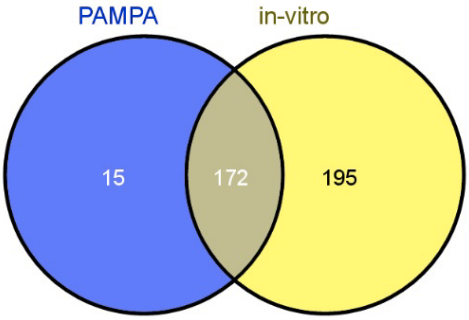


Figure 4.16. Venn diagram showing the number of overlapping sequences between cell based *in vitro* BBB model assayed sample and PAMPA assay after Ac-D-Arg-library assay.

It is noteworthy to stress that 172 families found in PAMPA assay out of 187 (92%) were also found in the *in vitro* cell-based BBB assay (not necessarily in all triplicates). This data demonstrate nice performance of this combined assay approach showing great agreement when predicting molecular passive diffusion properties.

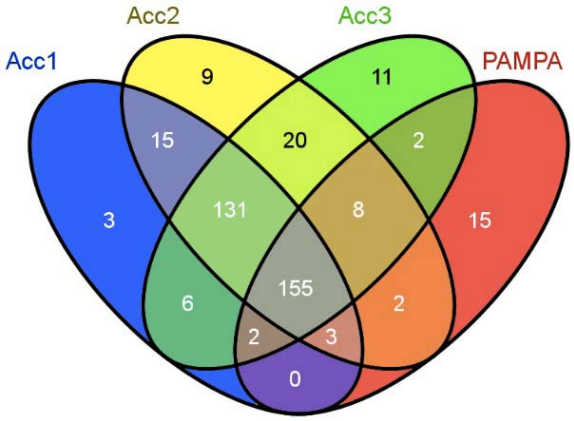


Figura 4.17. Venn diagram showing the number of overlapping sequences among all triplicates (Acc1, Acc2, Acc3) of the *in vitro* cell-based BBB model assayed sample and PAMPA assay after Ac-D-Arg-library assay.

See Appendix 4
Complete family list and information filtered on passive diffusion transport list.

See Appendix 5
Complete family list and information filtered on other mechanisms transport list.

Other mechanism transport list includes signals found in all triplicates of *in vitro* cell-based BBB assay but not in PAMPA. *Bibliopepfinder* assigned 131 families to 119 m/z signals in LTQ-Orbitrap

spectrum. 108 m/z signals matched one single family each while 10 m/z signals could be assigned to two families each, and one to 3 distinct families.

■ **Assumption 3.** For family selection procedure, families are considered as a single entity. It is assumed that the same members (sequences) within a family (composition) are present in compared lists. Then, during this selection process some sequences might be wrongly discarded. This happens when subtracting both blank sample from Ac-D-Arg-library tested sample and *in vitro* cell-based model – PAMPA assay lists. ■

Hurdle 5. False discovery rate.

False discovery rate is an important parameter to determine when analysing such complex mixtures by matching with a database. We calculated false discovery rate on the fist MS-identification level by creating a new database corresponding to a random library of hexapeptides comprising distinct amino acids set than the one analysed; only arginine is present in this decoy database to apply the same restriction parameters on matching than for Ac-D-Arg-library. Amino acid comprising decoy database are: arginine, glycine, phenylalanine, asparagine, glutamine, threonine and valine. Then, all extracted lists from LTQ-Orbitrap were matched with this decoy database in *Bibliopepfinder* to obtain a *false positive decoy* list to eventually assess the false discovery rate. False discovery rate is determined with the formula below:

$$\%\, FPR = \frac{False\ positive_{Decoy}}{Total_{Ac-D-Arg-lib}} \cdot 100$$

Where *Total* include all families found by *Bibliopepfinder* with Ac-D-Arg-library database and those found in $False\ positive_{Decoy}$.

	Ac-D-Arg-library	False library database	FDR (%)
1	315	6	1.9
2	343	12	3.4
3	335	3	0.9

Table 4.6. False discovery rate calculated for each triplicate on *in vitro* cell-based BBB model. FDR refers to false discovery rate percentage. Families identified by *Bibliopepfinder* in Ac-D-Arg-library database and decoy database are shown in the second and third columns, respectively.

False discovery rate values appear significantly low and thus lead to more trustable family selection. Furthermore, second level of identification will further address validation of those selected families since sequence confirmation requires a full set of co-eluting transitions on Q-trap (further discussion on 4.3.3.2).

Thereby, from each list some families were selected following the rational discussed in Hurdle 4. The following statements were applied:

1 - Must be in all triplicates.
2 - Its molecular peak must have a typical mass ratio profile in mass spectrometry of peptide species in the MS1 LTQ-Orbitrap spectra (example shown in Figure 4.18).
3 - Apply a score criterion based on weighted intensity ranking on all triplicates.
4 - Families with more than 30 members are automatically excluded.

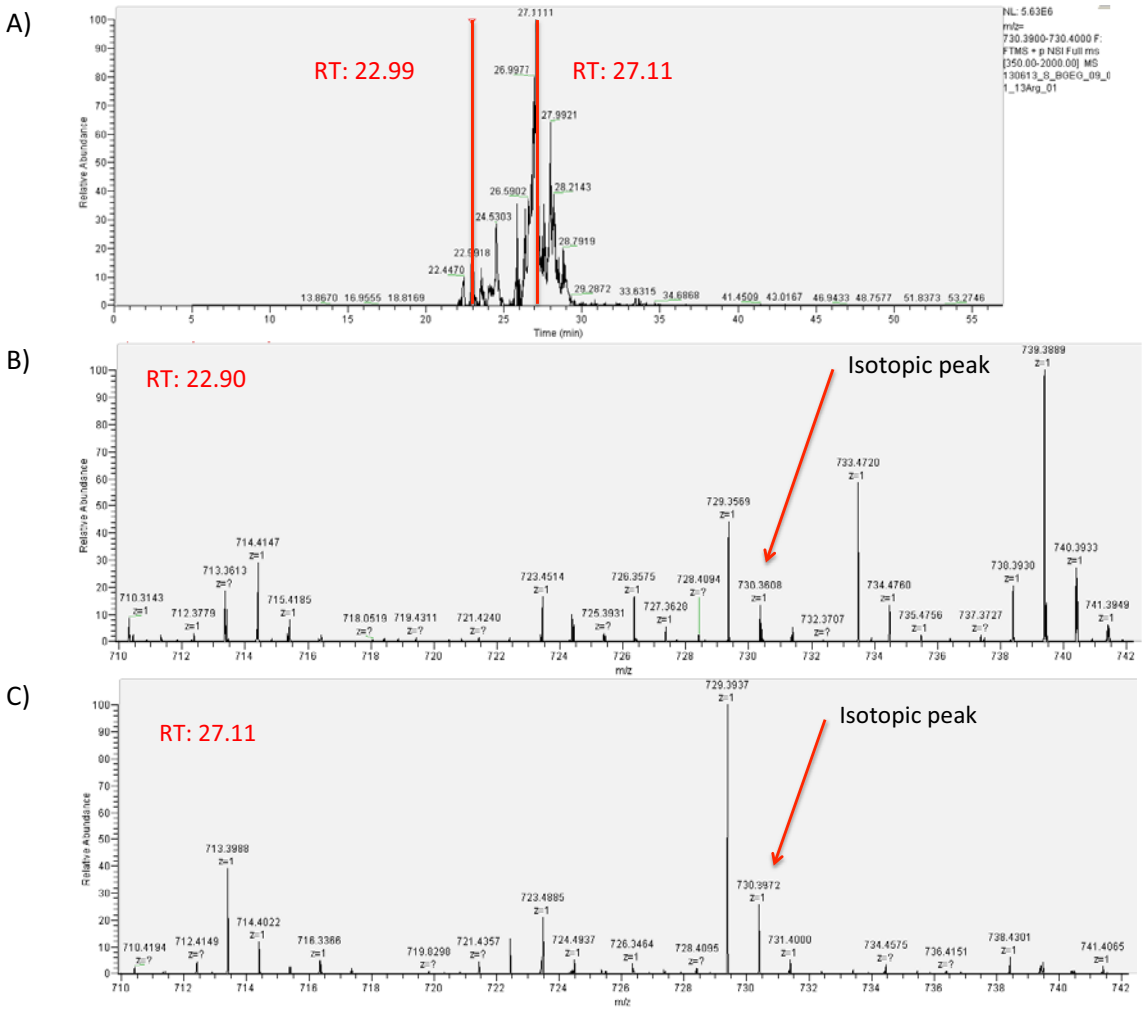


Figure 4.18. Mass list extracted from MS1 spectra of LTQ-Orbitrap contains all peaks detected. However, several peaks correspond to isotopic peaks. Since *Bibliopepfinder* is not able to discriminate between molecular or isotopic peaks on the extracted list it is required a manual checking of selected compositions. m/z signals corresponding to each composition must appear as a molecular peak with its typical isotopic envelope. The figure shows this common phenomenon (stated as *isotopic* in Tables 4.8. and 4.10.). These peptide compositions are usually discarded for further study. A) SIC spectrum (selected ion chromatogram) shows all RT where an m/z between 730.39 and 730.40 is found. B&C) MS1 spectra are shown at RT of 22.90 min and 27.11 min. Peak corresponding to m/z = 730.39 clearly corresponds to an isotopic peak.

In Table 4.9, *other mechanisms transport list* selection rational can be followed, which lead to the selection of the following families (Table 4.7):

FAMILY	FORMULA	M/Z	Members
3111	I3-A1-P1-R1	723.48	20
2211	E2-P2-A1-R1	739.37	30
2211	P2-W2-I1-R1	895.49	30
321	I4-P1-R1	765.53	5

Table 4.7. Families selected from *other mechanisms transport list*.

The same selection criterion was followed for the *passive diffusion transport list* (*see* Table 4.10) and the resulting selected families were the ones shown in Table 4.8:

FAMILY	FORMULA	M/Z	Members
321	I3-P2-R1	749.50	10
321	P3-S2-R1	681.37	10
321	A3-R2-P1	682.41	20

Table 4.8. Families selected from *passive diffusion transport list*.

FAMILY	FORMULA	ACC 1	ACC 2	ACC 3	Relative position	Reduced family	Family members	M-THEORIC	M-EXP	M/Z	Z	Comments
3111	I3-A1-P1-R1	3	2	3	2.67	311	20	722.48023	722.47936	723.48718	1	✔
2211	E2-P2-A1-R1	5	3	1	3.00	221	30	738.36599	738.37023	739.37805	1	✔
51	I5-R1	6	5	7	6.00	5	1	780.55848	780.55504	391.28534	2	✗ *poor signal*
2211	P2-W2-A1-R1	12	9	9	10.00	221	30	852.43943	852.43854	427.22709	2	✗ *poor signal*
2211	P2-W2-I1-R1	14	16	12	14.00	221	30	894.48638	894.48326	895.49108	1	✔
321	A3-R2-S1	18	13	11	14.00	311	20	671.38264	671.38362	672.39144	1	✗ *isotopic*
222	A2-R2-S2	26	18	22	22.00	221	30	687.37756	687.37856	688.38638	1	✗ *isotopic*
2211	E2-I2-R1-S1	28	21	18	22.33	221	30	786.42351	786.4232	787.43102	1	✗ *poor signal*
411	I4-P1-R1	30	22	15	22.33	41	5	764.52718	764.5263	765.53412	1	✔
3111	I3-A1-E1-R1	31	23	19	24.33	311	20	754.47006	754.47181	755.47963	1	✗ *isotopic*
222	E2-R2-W2	29	24	25	26.00	221	30	1001.48308	1001.48146	501.74855	2	
2211	A2-W2-I1-R1	34	25	23	27.33	221	30	842.45508	842.45418	422.23491	2	
2211	A2-W2-P1-R1	54	17	16	29.00	221	30	826.42378	826.42764	414.22164	2	
321	W3-R2-E1	36	27	28	30.33	311	20	1058.5198	1058.5227	530.26917	2	
321	W3-R2-A1	37	28	26	30.33	311	20	1000.51432	1000.5147	501.26517	2	
2211	E2-S2-P1-R1	8	6	82	32.00	221	30	744.34018	744.34565	745.35347	1	
321	E3-A2-R1	9	7	83	33.00	32	10	744.34017	744.34565	745.35347	1	
2211	E2-W2-I1-R1	39	30	34	34.33	221	30	958.46604	958.46376	480.2397	2	
2211	E2-P2-R1-S1	49	35	29	37.67	221	30	754.36091	754.36572	755.37354	1	
321	I3-R2-A1	20	12	87	39.67	311	20	781.52857	781.52822	391.77193	2	
2211	A2-I2-P1-R1	1	65	64	43.33	221	30	680.43328	680.4307	681.43852	1	
2211	P2-W2-E1-R1	121	11	10	47.33	221	30	910.44491	910.44438	456.23001	2	
321	P3-W2-R1	24	103	20	49.00	32	10	878.45508	878.45418	440.23491	2	
3111	I3-E1-R1-S1	23	70	65	52.67	311	20	770.46498	770.464	386.23982	2	
2211	E2-P2-I1-R1	41	98	24	54.33	221	30	780.41294	780.4135	391.21457	2	
321	I3-R2-P1	92	82	4	59.33	311	20	807.54422	807.5449	404.78027	2	
2211	I2-S2-R1-W1	38	29	123	63.33	221	30	801.44967	801.44908	401.73236	2	
222	P2-R2-W2	7	99	95	67.00	221	30	937.50342	937.502	469.75882	2	
2211	A2-I2-R1-W1	22	69	110	67.00	221	30	769.45983	769.45852	385.73708	2	
3111	A3-E1-R1-W1	15	108	118	80.33	311	20	743.37141	743.37574	744.38356	1	

Table 4.9. This table details the selection reasonement for *other mechanisms transport list*. Acc1, Acc2 and Acc3 show the position of each family in each triplicate complete LTQ-Orbitrap family list. Relative position column shows the average of Acc1, Acc2 and Acc3 columns showing the best candidate families. Reduced family column illustrates in simplified notation family composition removing the N-terminus arginine to calculate the number of family members. Finally, "observations" column comments on individually checked LTQ-Orbitrap spectra for each family specific mass.

FAMILY	FORMULA	ACC 1	ACC 2	ACC 3	Relative position	Reduced family	Family members	M-THEORIC	M-EXP	M/Z	Z	Comments
321	I3-P2-R1	9	2	8	6.33	32	10	748.49588	748.49572	749.50354	1	✔
321	S3-R2-P1	16	7	12	11.67	311	20	729.38813	729.38985	730.39767	1	✗ *isotopic*
321	A3-R2-E1	22	9	13	14.67	311	20	713.3932	713.39458	714.4024	1	✗ *isotopic*
2211	A2-P2-E1-R1	30	11	5	15.33	221	30	680.36051	680.36452	681.37234	1	✔
321	P3-S2-R1	31	12	6	16.33	32	10	680.36052	680.36452	681.37234	1	✔
321	A3-R2-P1	29	17	10	18.67	311	20	681.40337	681.40593	682.41375	1	✔
3111	I3-E1-P1-R1	20	8	37	21.67	311	20	780.48571	780.48532	781.49314	1	✗ *poor signal*
321	A3-R2-I1	38	21	24	27.67	311	20	697.43467	697.43879	698.44661	1	✗ *isotopic*
2211	I2-P2-R1-S1	3	1	91	31.67	221	30	722.44385	722.44355	723.45137	1	✔
321	R3-A2-E1	34	36	25	31.67	221	30	798.45719	798.46478	799.4726	1	✗ *isotopic*
3111	P3-E1-I1-R1	43	26	29	32.67	311	20	748.42311	748.42691	749.43473	1	
321	R3-S2-P1	35	37	26	32.67	221	30	798.4572	798.46478	799.4726	1	
222	P2-R2-S2	48	28	32	36.00	221	30	739.40886	739.4149	740.42272	1	
2211	A2-E2-P1-R1	1	49	60	36.67	221	30	712.35034	712.35571	713.36353	1	
3111	S3-I1-P1-R1	18	5	89	37.33	311	20	686.37109	686.37494	687.38276	1	
2211	P2-S2-E1-R1	2	50	61	37.67	221	30	712.35035	712.35571	713.36353	1	
411	A4-R1-S1	5	70	48	41.00	41	5	586.31865	586.32175	587.32957	1	
2211	I2-P2-E1-R1	40	24	81	48.33	221	30	764.45441	764.45618	765.464	1	
3111	P3-I1-R1-W1	7	141	4	50.67	311	20	805.45983	805.46066	403.73815	2	
321	R3-A2-S1	110	16	28	51.33	221	30	756.44663	756.44764	379.23164	2	
321	A3-S2-R1	36	19	105	53.33	32	10	602.31357	602.31713	603.32495	1	
3111	A3-E1-R1-S1	10	93	75	59.33	311	20	644.32413	644.32545	645.33327	1	
321	I3-R2-E1	45	29	104	59.33	311	20	839.53405	839.53694	420.77629	2	
3111	S3-A1-P1-R1	11	94	76	60.33	311	20	644.32414	644.32545	645.33327	1	
2211	P2-W2-R1-S1	152	30	33	71.67	221	30	868.43435	868.43421	869.44203	1	
3111	P3-A1-R1-W1	8	131	116	85.00	311	20	763.41288	763.41671	764.42453	1	
321	P3-I2-R1	6	125	146	92.33	32	10	732.46458	732.46612	367.24088	2	
3111	I3-P1-R1-S1	143	143	9	98.33	311	20	738.47515	738.4771	370.24637	2	

Table 4.10. This table details the selection reasonement for *passive diffusion transport list*. Acc1, Acc1 and Acc3 show the position of each family in each triplicate complete LTQ-Orbitrap family list. Relative position column shows the average of Acc1, Acc2 and Acc3 columns showing the best candidate families. Reduced family column illustratrates in simplified notation family composition removing the N-terminus arginine to calculate the number of family members. Finally, “observations” column comments on individually checked LTQ-Orbitrap spectra for each family specific mass.

Selection of P3-S2-R1 was a risky decision since share the same m/z with A2-P2-E1-R1 family (LTQ-Orbitrap precision set at 10 ppm). Then, here we risk our chances to find the sequences because we may be wrong on which family is responsible for that m/z signal (both could be present).

■ **Assumption 4.** Peptide selection criteria are based on peak intensity. Since signal intensity not always correlates with peptide abundance, we assume this selection might lead to biased conclusions. ■

Another *perl*-programmed script was written by Xavi Arroyo in Prof. Giralt's group to generate all permutations of selected families. Thus, from all selected families, a *.fasta* file with all those permutations was edited.

See Appendix 6
Perl program to generate permutations script.

Passive diffusion transport list

FAMILY	FORMULA	M/Z	Members
321	I3-P2-R1	749.50	10
321	P3-S2-R1	681.37	10
321	A3-R2-P1	682.41	20

I3-P2-R1	P3-S2-R1	A3-R2-P1
RIIPP	RPPPSS	RAAAPR
RIIPIP	RPPSPS	RAAARP
RIIPPI	RPPSSP	RAAPAR
RIPIIP	RPSPPS	RAAPRA
RIPIPI	RPSPSP	RAARAP
RIPPII	RPSSPP	RAARPA
RPIIIP	RSPPPS	RAPAAR
RPIIPI	RSPPSP	RAPARA
RPIPII	RSPSPP	RAPRAA
RPPIII	RSSPPP	RARAAP
		RARAPA
		RARPAA
		RPAAAR
		RPAARA
		RPARAA
		RPRAAA
		RRAAAP
		RRAAPA
		RRAPAA
		RRPAAA

Other mechanisms transport list

FAMILY	FORMULA	M/Z	Members
3111	I3-A1-P1-R1	723.48	20
2211	E2-P2-A1-R1	739.37	30
2211	P2-W2-I1-R1	895.49	30
321	I4-P1-R1	765.53	5

I3-A1-P1-R1	E2-P2-A1-R1	P2-W2-I1-R1	I4-P1-R1
RAIIIP	RAEEPP	RIPPWW	RIIIIP
RAIIPI	RAEPEP	RIPWPW	RIIIPI
RAIPII	RAEPPE	RIPWWP	RIIPII
RAPIII	RAPEEP	RIWPPW	RIPIII
RIAIIP	RAPEPE	RIWPWP	RPIIII
RIAIPI	RAPPEE	RIWWPP	
RIAPII	REAEPP	RPIPWW	
RIIAIP	REAPEP	RPIWPW	
RIIAPI	REAPPE	RPIWWP	
RIIIAP	REEAPP	RPPIWW	
RIIIPA	REEPAP	RPPWIW	
RIIPAI	REEPPA	RPPWWI	
RIIPIA	REPAEP	RPWIPW	
RIPAII	REPAPE	RPWIWP	
RIPIAI	REPEAP	RPWPIW	
RIPIIA	REPEPA	RPWPWI	
RPAIII	REPPAE	RPWWIP	
RPIAII	REPPEA	RPWWPI	
RPIIAI	RPAEEP	RWIPPW	
RPIIIA	RPAEPE	RWIPWP	

Figure 4.19. Complete set of sequences from selected families from *passive diffusion transport list* and *other mechanisms transport list* generated is shown as an example.

4.4.4. Second MS-identification level: Q-trap (QqQ)

4.4.4.1. Fragmentation pattern generation

Skyline proteomics software generates all desired transitions from all sequences selected. This allows monitoring every single transition with Q-trap to confirm the presence or not of each monitored sequence.

To confirm the presence (or not) of each target peptide sequence we monitored transitions from both singly and doubly charged precursor ion. On one hand, for singly charged precursor ion, 10 transitions were explored: b_1^+, b_2^+, b_3^+, b_4^+, b_5^+ and y_5^+, y_4^+, y_3^+, y_2^+, y_1^+. On the other hand, for doubly charged precursor ion, 20 transitions were followed: b_1^+, b_2^+, b_3^+, b_4^+, b_5^+, y_5^+, y_4^+, y_3^+, y_2^+, y_1^+ and b_1^{++}, b_2^{++}, b_3^{++}, b_4^{++}, b_5^{++}, y_5^{++}, y_4^{++}, y_3^{++}, y_2^{++}, y_1^{++}.

To validate any sequence, transitions corresponding to fragmentation of each amide bond must be found. So that, any b_1 or y_5, b_2 or y_4, b_3 or y_3, b_4 or y_2 and b_5 or y_1 must be found to unequivocally confirm the presence of this sequence in the sample. Obviously, those signals must not appear in the blank sample injected in QqQ.

In some cases, not all transitions need to be found in case there are two or more equal amino acids next to each other. For example, RIIIPA can be assigned by finding b_1 or y_5, b_4 or y_2 and b_5 or y_1. However, assignments finding fragmentation ions validating all transitions are much more consistent and robust.

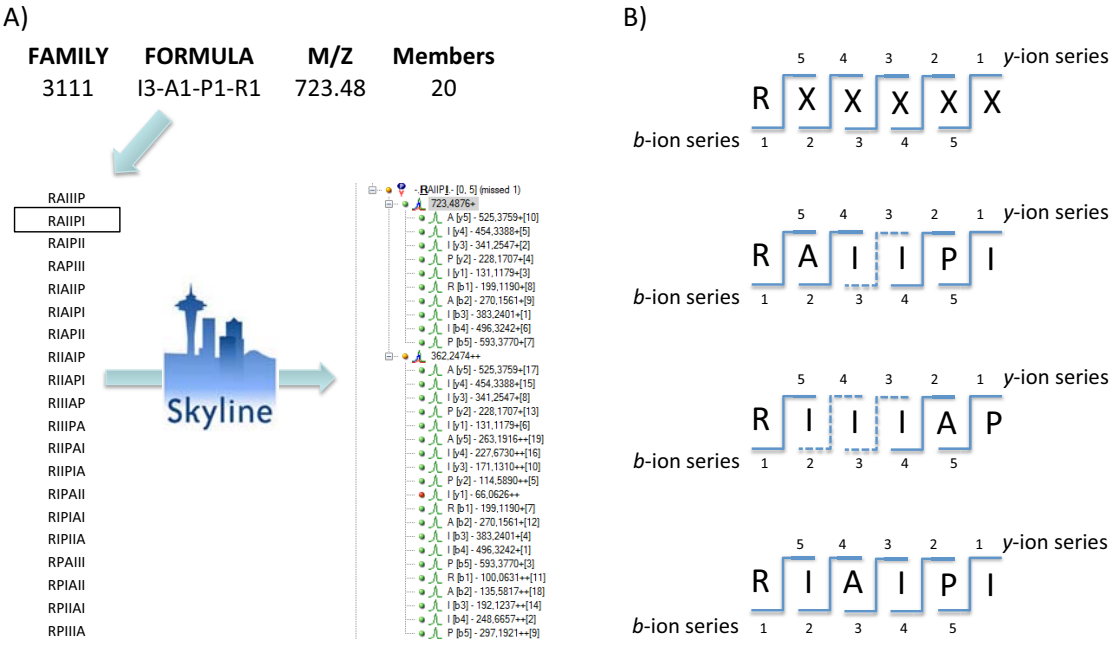


Figure 4.20. A) Workflow followed for generation of transitions. RAIIPI is shown as a case of example. B) Transitions required to fully confirming each sequence depending on their specific sequence. Full transitions profile ensures more robust identification results.

Hurdle 6. Generation of monitored transitions information for QqQ. Skyline proteomics.

Skyline is an open-source software platform that provides support for experimental design and downstream data analysis for proteomics.[42] Skyline is a vendor-neutral resource that uses ProteoWizard libraries and allows the analysis of data from all MS instrument platforms.[43]

Skyline generates proteomic peptide list from protein sequences or database entries and supports the generation of custom-built libraries. Several parameters can be monitored and personalized for creating transition lists. It allows the export of processed data in reports formats compatible with statistical analyses, publication or database deposition.

Then, by creating a database containing the information form members of selected families that want to be monitored from our synthetic Ac-D-Arg library. Skyline proteomics software was a useful platform to generate all data required to inject samples to QqQ.

4.4.4.2. Sequence confirmations through transitions co-elution

Q-trap spectra were manually and individually assigned using Skyline Proteomics software. Spectra show all transitions of the screened sequence (specific family members) within a

retention time chromatogram (*x* axis). Each transition is drawn in a different colour. Co-elution of enough transitions (unequivocally describing the screened sequence) in the chromatogram ensures the presence of that particular sequence. Figure 4.21 shows, as a case example, the co-elution of all transitions defining RAIIPI assignment on peak at RT = 16.7 minutes.

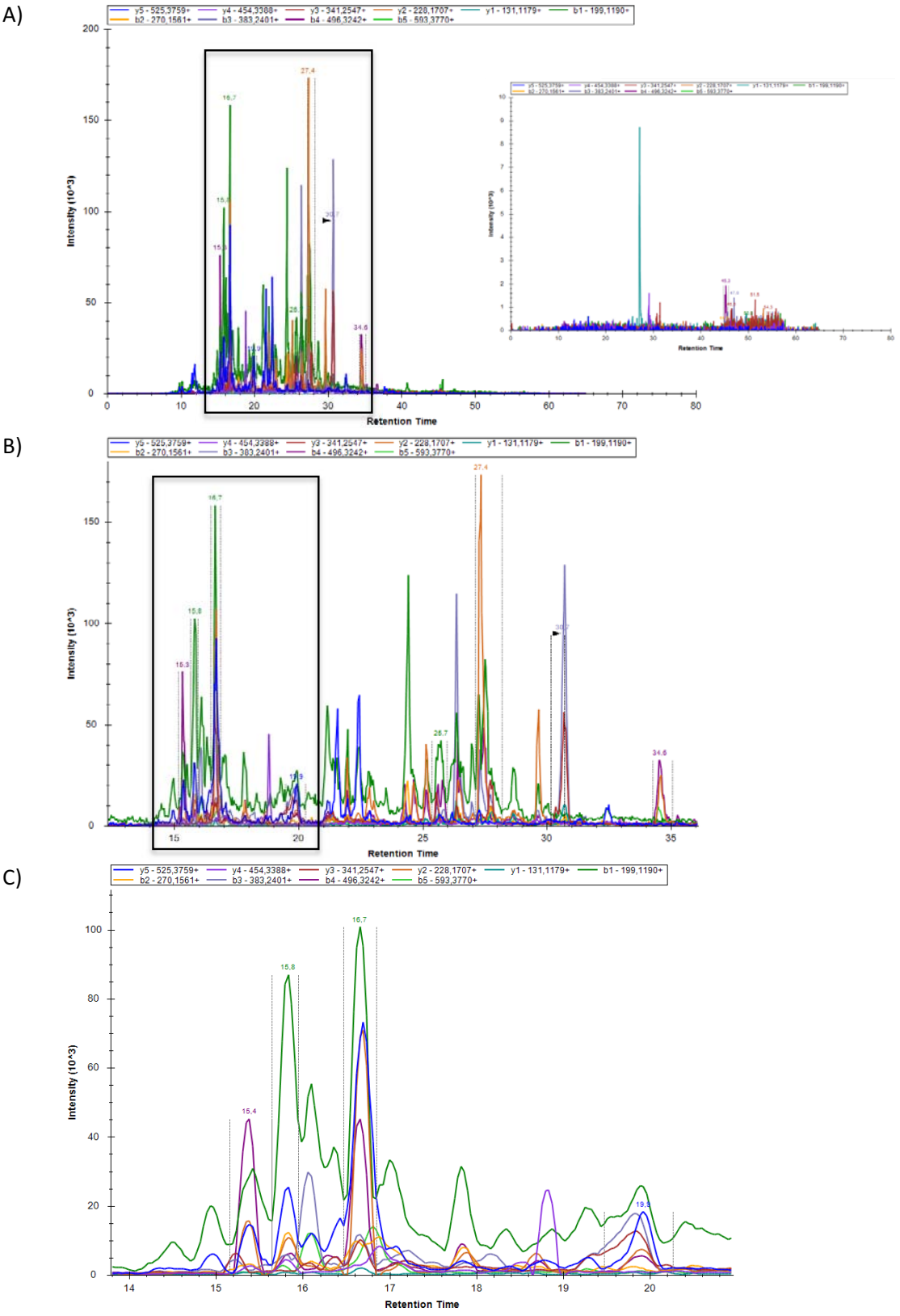


Figure 4.21. This figure show as a case of example (RAIIPI) the process followed for every single targeted sequence to search for complete sets of co-eluting transitions. A, B and C correspond to the SRM spectra and zoom in images of the same spectra. A also shows the blank spectra monitoring the same sequence.

Other mechanisms transport list		*Passive diffusion transport list*		
I3-A1-P1-R1	I4-P1-R1	I3-P2-R1	P3-S2-R1	A3-R2-P1
RAIIIP	RIIIIP	RIIIPP	RPPPSS	RAAAPR
RAIIPI	RIIIPI	RIIPIP	RPPSPS	RAAARP
RAIPII	RIIPII	RIIPPI	RPPSSP	RAAPAR
RAPIII	RIPIII	RIPIIP	RPSPPS	RAAPRA
RIAIIP	RPIIII	RIPIPI	RPSPSP	RAARAP
RIAIPI		RIPPII	RPSSPP	RAARPA
RIAPII		RPIIIP	RSPPPS	RAPAAR
RIIAIP		RPIIPI	RSPPSP	RAPARA
RIIAPI		RPIPII	RSPSPP	RAPRAA
RIIIAP		RPPIII	RSSPPP	RARAAP
RIIIPA				RARAPA
RIIPAI				RARPAA
RIIPIA				RPAAAR
RIPAII				RPAARA
RIPIAI				RPARAA
RIPIIA				RPRAAA
RPAIII				RRAAAP
RPIAII				RRAAPA
RPIIAI				RRAPAA
RPIIIA				RRPAAA

Legend
Co-eluting transitions not found in any triplicate
Co-eluting transitions found in some triplicates, but not at the same retention time
Co-eluting transitions found in some triplicates, at the same retention time
Co-eluting transitions found in all triplicates, but not at the same retention time
Co-eluting transitions found in all triplicates, at the same retention time, but not on the same precursor ion
Co-eluting transitions found in all triplicates, at the same retention time

Table 4.11. Summarized information extracted from SRM experiments performed in Q-trap. Information showed in the table refers to co-eluting transitions found in triplicate samples and in which retention time they were found. Detailed information can be found in *Appendix 7*. No significant information was found neither for E2-P2-A1-R1 nor P2-W2-I1-R1.

See Appendix 7

Detailed information of transitions found for sequences monitored by SRM.

4.4.4.3. Prospects

Those selected sequences need to be resynthesized and tested alone in *in vitro* cell-based BBB model to validate their permeation ability. Furthermore, quantification should be performed to rank those sequences in terms of *in vitro* cell-based BBB model permeability. Further discussion on these issues is in section 4.5.1.

4.4.4.4. Hypothesis, expectations and observations from QqQ analysis

HPLC family profiles usually show several peaks in spectra. Selected ion chromatograms (SIC) selecting a tight m/z range (0.01 Da window) in LTQ-Orbitrap, not only show signals corresponding to our family but also to other species matching within this m/z range. However, we assume that most peaks in those spectra are sequences (members) from the targeted family with distinct chromatographic retention times. Extra peaks appearing in SIC spectra due to isotopic peaks in MS1 spectra are not an issue on QqQ MS-identification step since they most likely will not fully match the selected transitions profile (co-eluting).

Figure 4.22 shows as a case example, an m/z signal matching a library family (A3-R2-E1) where peaks displayed on SIC LTQ-Orbitrap spectra (0.01 Da window) correspond both to molecular peak (most likely from A3-R2-E1) and to isotopic peaks (corresponding to other species).

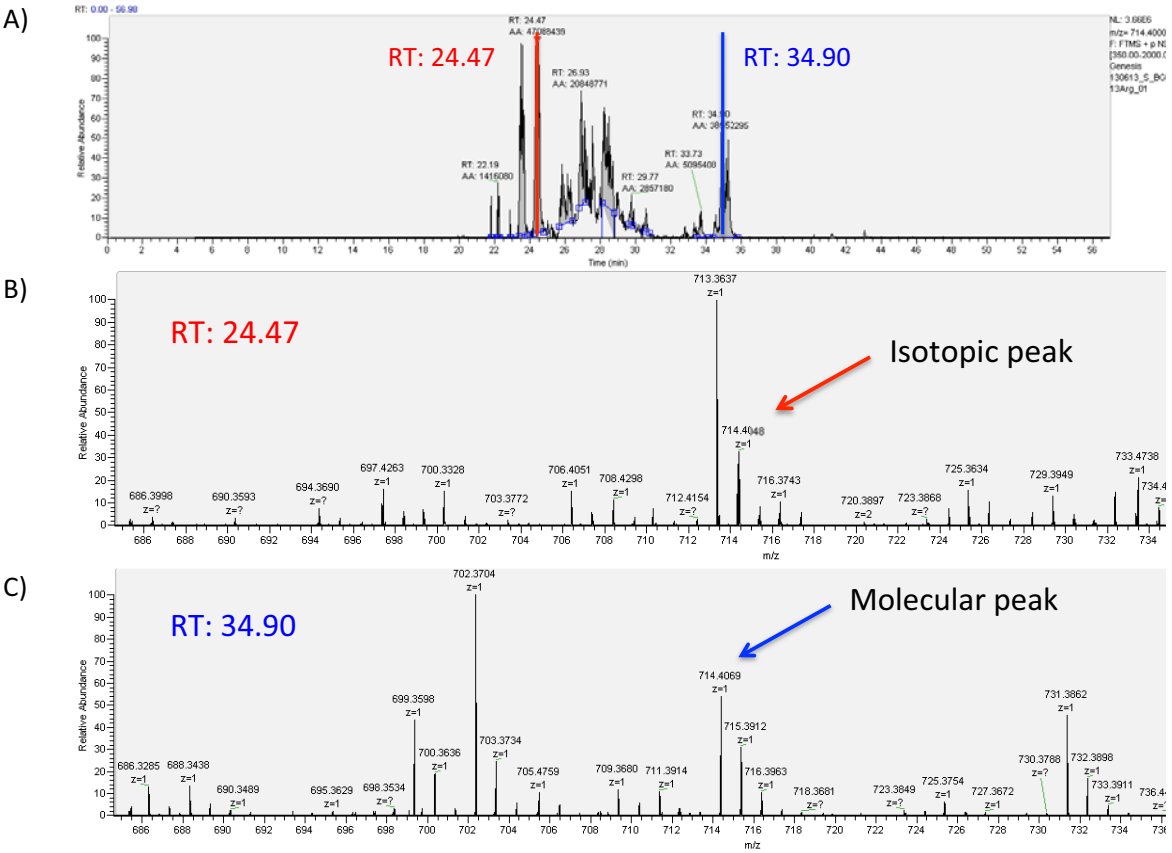
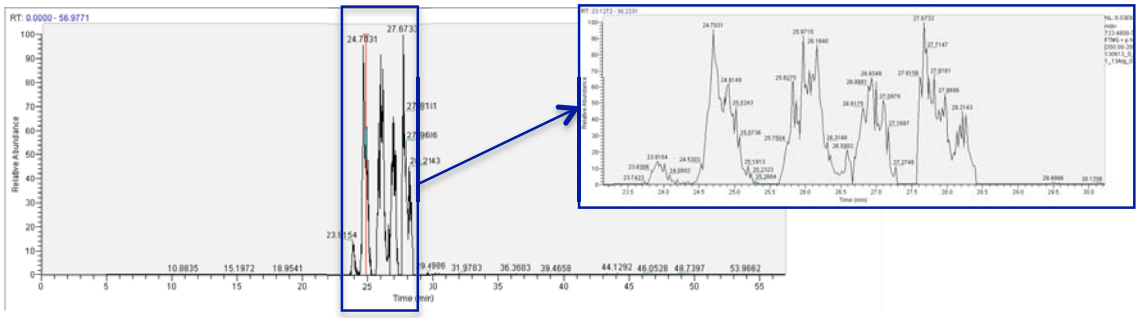


Figure 4.22. Mass list extracted from MS1 spectra of LTQ-Orbitrap contains all peaks detected. However, several peaks correspond to isotopic peaks. Since *Bibliopepfinder* is not able to discriminate between molecular or isotopic peaks on the list, it is required a manual checking of those compositions that are selected. m/z signals corresponding to each composition must appear as a molecular peak with its typical isotopic envelope (fewer intense peaks in the mass range of Ac-D-Arg-library). In the figure, it is shown this common phenomena (stated as *isotopic* in Tables 4.4 and 4.6). Those peptide compositions are usually discarded for further study. In the present case, SIC, selected ion chromatogram, for species comprised on m/z signals between 714.40-714.41 is shown. Those species appear on distinct retention times. For example, MS1 signal corresponding to retention time of 24.47 minutes is due to an isotopic peak. However, in this specific case, a specie with a molecular peak matching targeted mass appears on MS1 spectra at 34.90 minutes. So that, even though it could lead to misleading results in some retention times, this specie could be selected for further study. Nonetheless, this family was discarded since other species were more appealing for targeted studies.

Once checked for every QqQ analysed family which peaks displayed on its SIC LTQ-Orbitrap spectra (0.01 Da window) correspond to molecular peaks, we assumed all of them represent the same family. Then, we reasoned that all these confirmed peaks correspond to one or more sequences of the same family with different retention times.

In QqQ spectra, we expected to see profiles similar to those observed in LTQ-Orbitrap SIC chromatogram since b_1 fragment is present in all family members. Then, highly populated spectra were expected as the one shown below as a case example for family I3-A1-P1-R1 targeting RAIIPI.

A)



B)

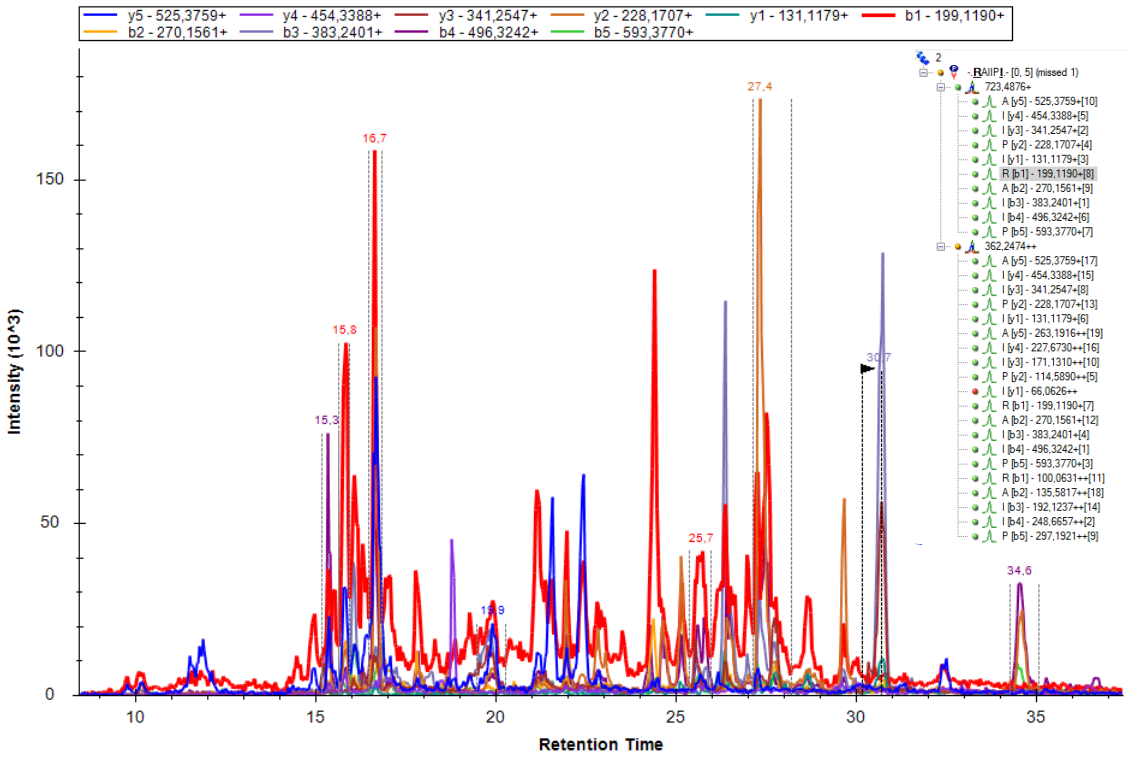


Figure 4.23. A) Selected ion chromatogram on LTQ-Orbitrap of m/z signals comprised between 723.48 and 723.49, where I3-A1-P1-R1 family is present. B) QqQ spectra monitoring transitions of RAIIPI. Red trace corresponds to transition monitoring b_1 ion. b_1 ion is shared by all family members (Ac-D-Arg- moiety). This could explain that much crowded spectra when monitoring a single family member. However, co-elution of transition signals should be found in few peaks.

Family members share most transitions. For example, RAIIPI and RAIIIP share all transitions but y_1 and b_5. So that, to discriminate between RAIIPI and RAIIIP either y_1 or b_5 are strictly required. Then, even though we are looking on the spectra targeting RAIIPI transitions, several peaks (in case sequences appear in distinct retention time) with incomplete co-elution profile will be indeed present in the spectra corresponding to RAIIPI and other sequences within the family. Ideally, in the spectrum targeting RAIIPI transitions only one peak with retention time RT_{RAIIPI} will display all co-eluting transitions defining that specific specie. We hypothesize and expect that in

the spectra targeting RAIIPI transitions, any peak with an incomplete set of co-eluting transitions corresponds to any other member of the family. Hence, we reasoned these peaks (appearing at specific RT) would display the complete profile of co-eluting transitions in the spectra targeting that specific sequence. This approach should allow mapping sequences appearing on each LTQ-Orbitrap SIC spectra. Altogether, this dissertation assumes all transitions could be seen when monitoring each sequence, which it is presumed unlikely.

Moreover, singly and doubly charged precursor ions fragmentation spectra from the same specie were not expected to be similar. Fragmentation and ionization suggest that mainly one precursor ion (either singly or doubly charged) would be preferred and then detected from its spectrum. Still, this cannot be stated as a general rule.

However, in acquired experimental data, transitions describing specific sequences were found both in singly or doubly charged precursor ion fragmentation spectra as shown in Figure 4.24. Unexpectedly, in some cases, transitions describing specific sequences can be found in not only one but in some peaks from the same spectra.

It was hypothesized that presence of proline within monitored families might partially explain such phenomena. Distinct peptide conformations could lead to substantial shifts in retention times. Then, each conformation of proline-containing peptides could have different retention times and thus appearing in numerous peaks in the same spectrum. However, it was odd to observe large retention time differences between peaks that were assigned to the same sequence. Hence, a complete set of transitions profiling a sequence in distinct retention times among triplicates cannot be completely ruled out, although rare. Triplicates, showing transitions co-eluting in the same retention time, clearly show higher consistency than others appearing in different retention times.

Moreover, QqQ precursor ion filtration window (in the first quadrupole, Q_1) is 0.7 Da. Although this value is so tiny compared with those achievable for equipment from years ago, it is far from being ideal. This 0.7 Da window is still huge for analysing complex mixtures, when stringent ion filtration conditions are desired. Thus, all compositions from peptides within the library sharing similar retention time and m/z values would be filtered together when targeting a specific m/z value. We found that 40 families within the library have ambiguous compositions on QqQ, since their masses are closer than 0.7 Da.

See Appendix 8
Complete list of family compositions with masses closer than 0.7 Da.

This is the worst-case scenario, since distinct retention time of coincident species would avoid this drawback. Presumably, stringent assays such as *in vitro* cell-based BBB assay should enormously simplify or at least reduce the amount of peptides within the sample.

Altogether, it makes spectra assignment much more complicated than presumed. Although using state-of-the-art equipment, these extremely redundant peptide libraries seem to be still too complex for direct identification with no other strategies assisting on simplifying the complexity such as use of sub-libraries or other deconvolution or encoding strategies.

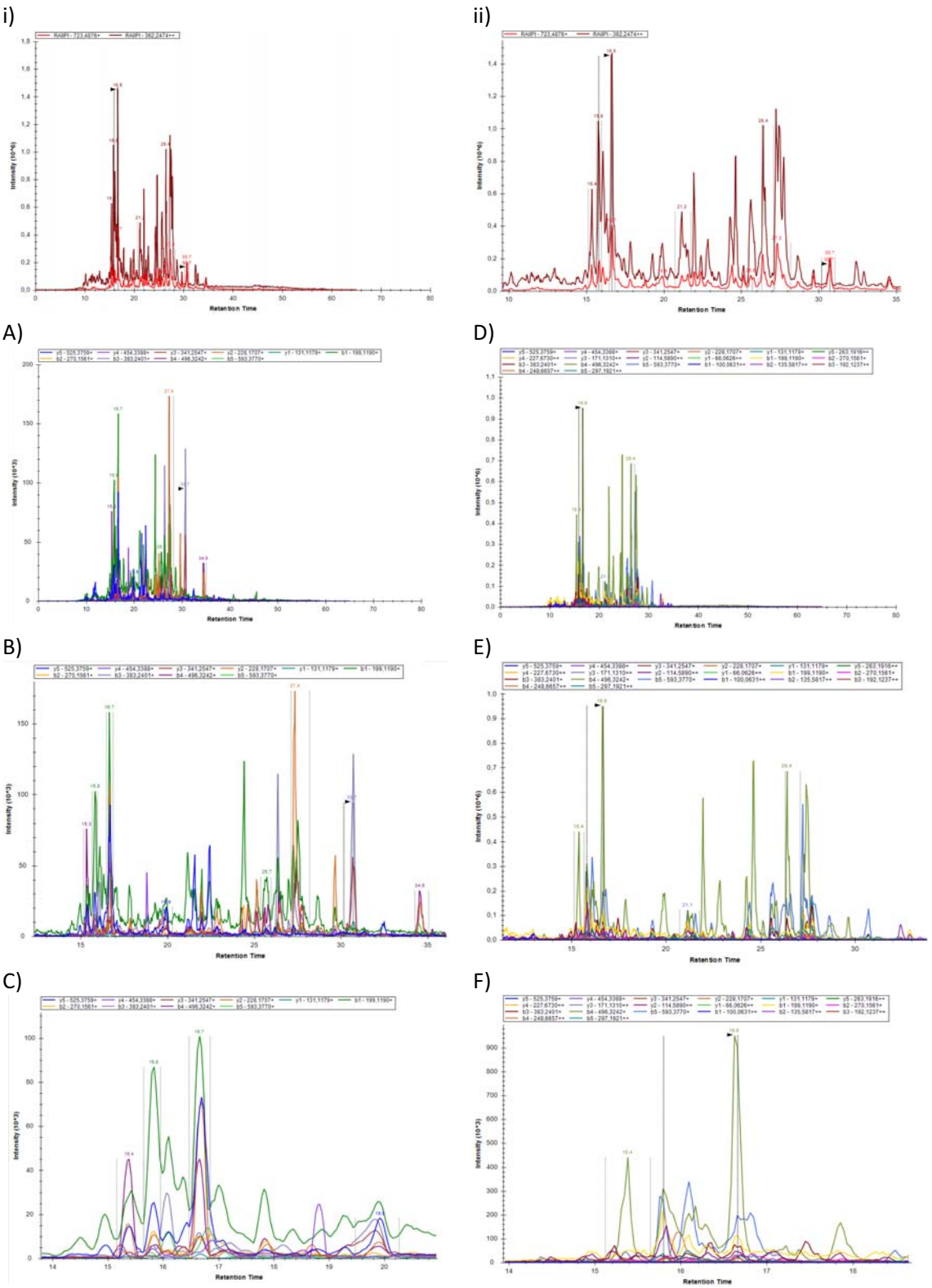


Figure 4.24. Figure showing QqQ spectra of precursor ions singly and doubly charged of RAIIPI, as a case example. i) shows the trace of precursor ions along RT while ii) is a zoom of the spectra. Singly and doubly charged precursor ions traces are surprisingly similar. A) shows the spectra of singly charged precursor ion (B and C are zoom images of A spectra), while D shows the spectra of doubly charged precursor ion (E and F are zoom images of D spectra).

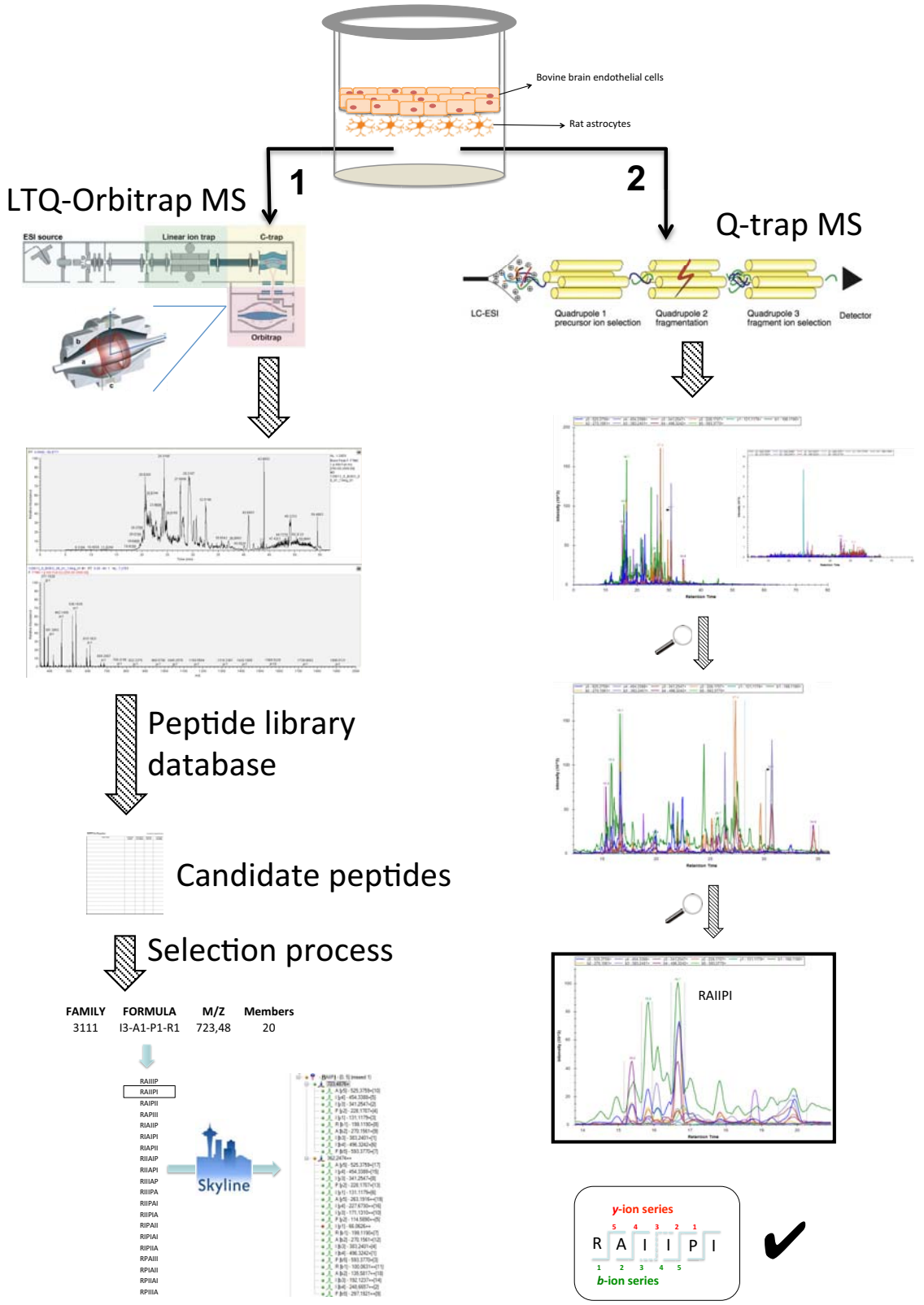


Figure 4.25. General scheme of proposed novel high-throughput screening identification methodology based on mass spectrometry. A workflow based on two MS-identification levels is set up. 1) First MS-identification step consist on choosing most promising peptide species corresponding to specific compositions (peptide families) with LTQ-Orbitrap MS, only operating on MS1 level. 2) Subsequently, targeted mass spectrometry (SRM) was applied as the second MS-identification level by monitoring transitions defining specific amino acidic sequences. Hence, peptide sequences (family members) comprised within the selected composition can unequivocally be validated. As case example RAIIPI is shown.

4.4.5. *In vivo* approach

In 1996 Pasqualini *et al*. applied phage display technology *in vivo* by injecting phage libraries into mice.[44] Since then, this *in vivo* approach has successfully evolved and lead to the discovery of several peptide sequences targeting specific healthy or carcinoma organs and tissues and it is been applied to human patients.

One of the most interesting features of *in vivo* assays is that allows performing biodistribution studies. By cross-correlating peptides identified in distinct organs, biodistribution information can be obtained. Then, peptides targeting specific organs or tissues can be found.

Since our one-bead one-compund libraries approach has a strong parallelism to phage display approach, we reasoned to perform biodistribution experiments using Ac-D-Arg-library described above. This approach is regarded as a potential future application of this methodology, but is not deeply explored in the present thesis since library was initially designed for *in vitro* assays and *in vivo* approach was considered during my stay in Dr. Renata Pasqualini and Dr. Wadih Arap lab at MD Anderson Cancer Center in Houston, TX (USA).

Hence, some crucial steps are discussed and first trials to tackle this approach are described. Experience gained on this *in vivo* scenario, lead us to some considerations that will be required to better design and set up an *in vivo* identification workflow for peptide high-throughput screening. Circulation time is the first parameter that has to be optimized. Short circulation times may not show a real picture of the whole situation while too long circulation times can lead to peptide elimination and thus giving misleading information. During my stay in Dr. Renata Pasqualini and Dr. Wadih Arap lab at MD Anderson Cancer Center in Houston, TX (USA), we injected the library via intra-tail vein injection; 3 circulation times were considered: 30 minutes, 6 hours and 24 hours. When injecting a peptide mixture via intra-tail vein injection, all peptides circulating through blood are exposed to tissue, organs, cells and eventually receptors of the whole body.

4.4.5.1. Focusing the brain

Focusing on the brain, peptides circulate through the blood throughout brain capillaries. Hence, all peptides are in close contact with BBB, which is surrounding all brain capillaries.

Generally speaking, these peptides can undergo four pathways. First, they might cross the BBB and enter the brain parenchyma. Second, they might be internalized by endothelial cells thus not crossing the BBB. Third, they might interact with a receptor or any part of the cell membrane being retained there: neither internalizing the cell nor crossing the BBB. Finally, they might not undergo any BBB effect and they might pass through to continue with blood circulation flow.

BBB-shuttle definition only fits with the first option. A BBB-shuttle peptide must be able to cross to the brain parenchyma. Theoretically speaking these differences are clear, but when moving to real experiments to get rid out of all peptides which are not BBB-shuttles is a difficult task.

As was mentioned before (*see* 4.4.1.3), mice were perfused with PBS. Well-performed perfusion cleans thoroughly all capillaries. This perfusion step removes those peptides which are not interacting with the BBB. One can trust, that also those peptides interacting with any part of the cell membrane in the brain capillaries will be detached if perfusion is abundant, then not interfering on BBB-shuttles discovery. However, peptides internalized into endothelial cells will be taken together with those which might be in brain parenchyma being real BBB-shuttles candidates when harvesting the whole brain.

Then, after peptide identification, performance of further experiments such as fluorescenct imaging experiments with peptides carrying a fluorescence tag and immunohistochemical studies will be needed to assess whether peptides cross the BBB or get stuck into the endothelial cells.[45]

Capillary depletion protocols had been described for a long time.[46-48] However, this protocol is a quite rudimentary dextran gradient and do not provide reliable, consistent and quantitative results (only a percentage of all capillaries are depleted). We discarded to use this capillary depletion protocol.

Once the organ is collected, it is chopped and stored. To analyse organ samples a regular protocol was performed to break up all cells and tissue to release all peptides entrapped. Cell membrane disrupting solution with 4% SDS in PBS was prepared and sonication was performed.

As was reported in section 4.4.2, three sample cleaning methods were performed. Most bibliographic references recall on organic solvent precipitation to get rid out of proteins. Usually, methanol or acetonitrile are used as a solvent to precipitate proteins overnight at -20ºC. However, we observed that while removing a considerable amount of precipitated protein, samples were still not enough clean and they block HPLC capillary columns.

On the other hand, microfiltration with 10 KDa spin filters showed better cleaning evidences. At a glance, it was possible to determine that samples were much cleaner. Since samples treated with this procedure did not block HPLC columns, this was the protocol of choice.

Generally speaking, this cleaning protocol must get rid out of proteins while not throwing off tested peptides. Those two objectives might be difficult to completely achieve together. Both spin filters microfiltration and organic solvent precipitation, have drawbacks that may trap peptides within removed proteins.[49, 50]

■ **Assumption 5.** Peptide-protein interactions are not under our control and peptides might be entrapped within protein-precipitated clusters (organic solvent precipitation) or peptides might be retained in the upper side of the filter laying in a protein bed that eventually stuck the sample passage through the spin filter. However, all this drawbacks are inherent to our library design, which initially did not consider *in vivo* scenario, and must be assumed as possible peptide-loosing steps. ■

Microfiltration process might last for several hours (4-6 hours) and proteases may break down endogenous proteins in some extent. While these are good news in terms of protein cleaning process, small size and middle size protease-generated peptides will pass through the spin filter and will be collected together with Ac-D-Arg-library peptides. Ac-D-Arg-library peptides are all D-peptides so; in this case we are not afraid for peptide cleavage protease activity on Ac-D-Arg-library peptides. However, endogenous and peptides directly coming from protease action may mask our peptides which probably are present in fewer quantities. Masking will depend both on relative abundance of endogenous peptides and Ac-D-Arg-library peptides, and on dynamic range of mass spectrometry equipment used for identification.

Two protocols were run with *in vivo* organ and tissue samples. On one hand, sequences found to be present *in vitro* were tested *in vivo* to try to find and validate *in vivo* those targeted sequences.

On the other hand, another identification protocol was intended for *in vivo* samples. The very same high-throughput screening workflow applied to *in vitro* samples was also performed *in vivo*. However, results obtained were really poor and suggest a full revision of library design to better set up conditions for *in vivo* high-throughput screening.

4.4.5.2. *In vivo* follow up of *in vitro* sequences

Since Ac-D-Arg-library was injected *in vivo*, we thought it was worth trying to look for those species found in *in vitro* assays into brain samples. Brain tissue samples were analysed monitoring transitions for both *passive diffusion transport list* selected families and *other mechanisms transport list* selected families.

No relevant data was found for *passive diffusion transport list* sequences since spectra had no co-eluting transitions and high signal/noise ratio. This was somehow expected since in the *in vivo* scenario, peptides circulating the blood are exposed in the whole animal body to passive diffusion mechanisms, thus highly diluting the final concentration of those peptide potentially being in the brain due to this low specificity. However, peptides crossing by active transport ideally being much more specific to receptors expressed in the endothelial cells of brain capillaries have more chances to be found.

In *other mechanism transport list*, a full profile of co-eluting transitions was found for all members of I4-P1-R1 family. However, retention time was 45 minutes. While this elution time was still fitting the gradient (it is not on acetonitrile cleaning stage) all signals found for those sequences *in vitro* were in the retention time range of 15-30 minutes. This retention time shift to 45 minutes cannot be attributed to an equipment malfunction performance. So that, this suggest that it is another specie previously not reported in *in vitro* samples (maybe masked or hidden) but present *in vivo*.

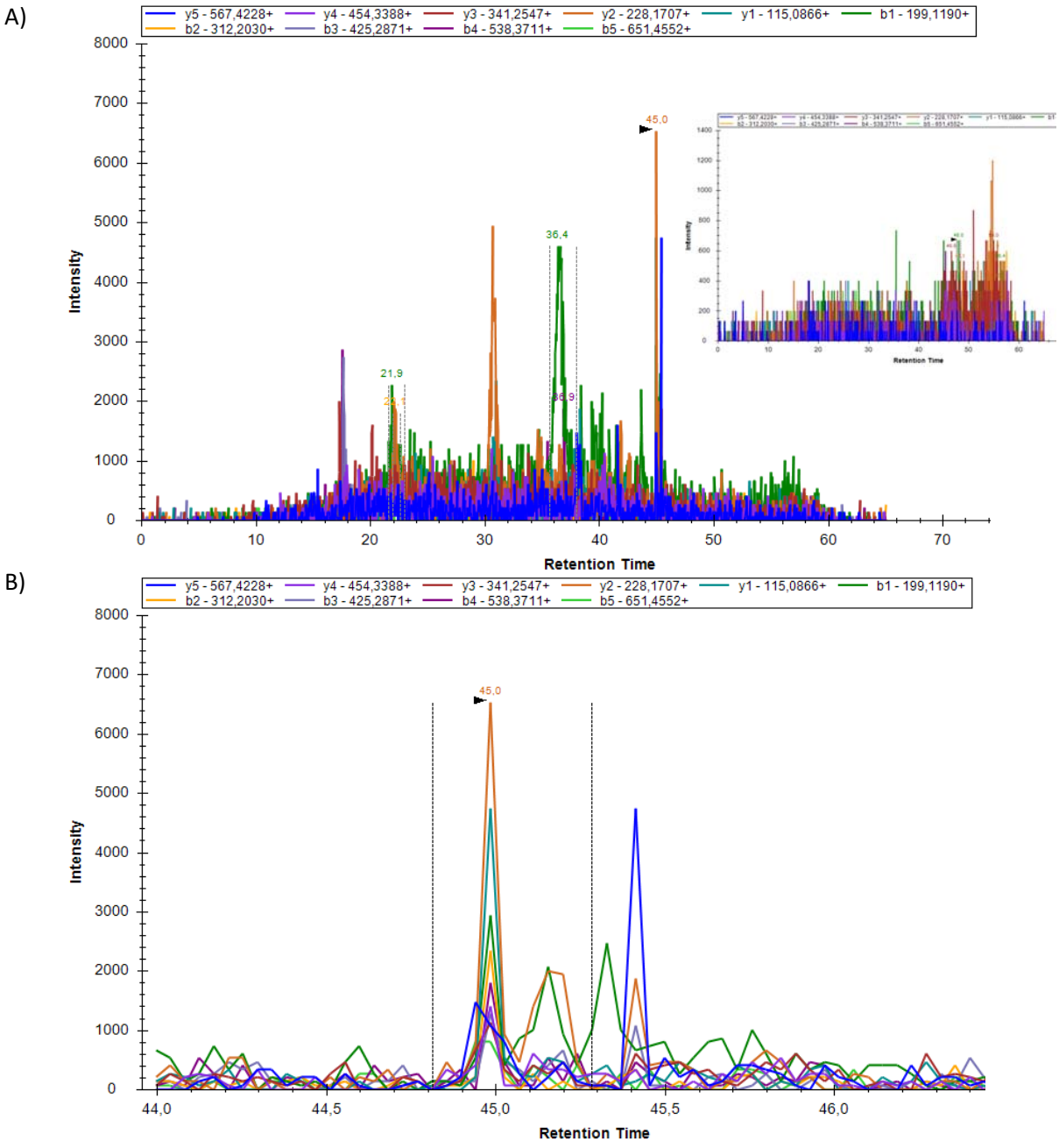
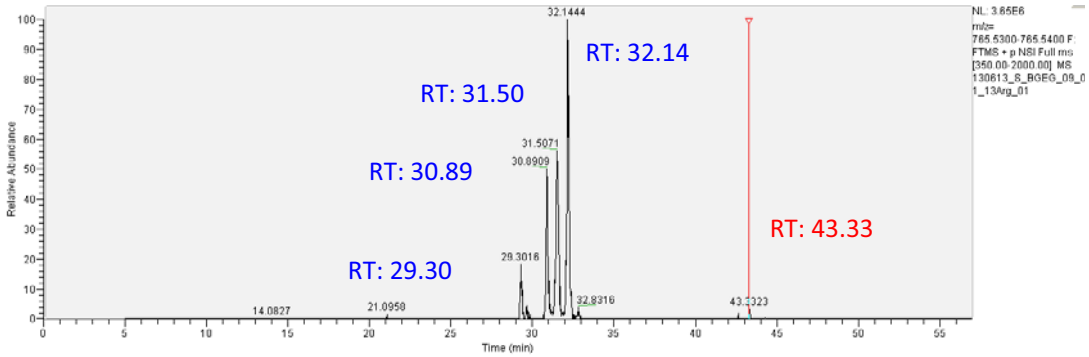


Figure 4.26. QqQ spectra monitoring transitions of RIIIIP on *in vivo* brain sample (A) and on blank sample (top-right). A complete co-eluting set of transitions defining the studied sequence appears on RT = 45 min (zoom in on B).

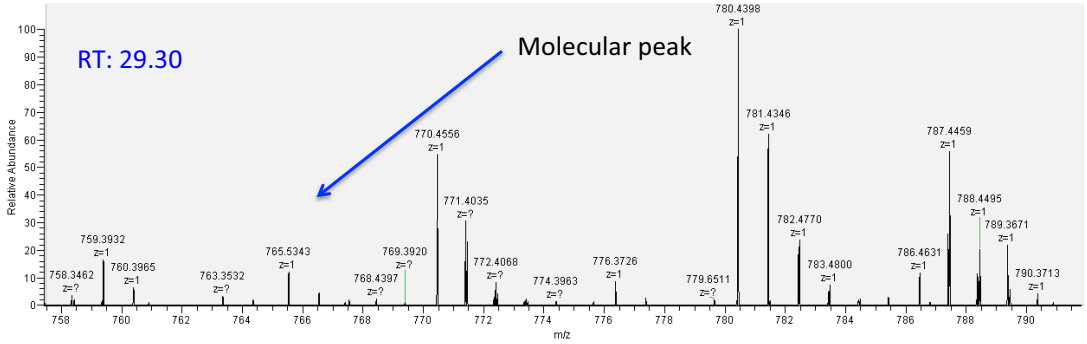
Complete profiles of co-elution transitions strongly suggest that this specie is present in the brain. To confirm there is indeed a specie matching m/z signal we look into MS1 LTQ-Orbitrap spectra both *in vitro* and *in vivo*. *In vitro* spectra showed a tiny peak around 45 minutes (even using same HPLC columns model, retention time of both equipment may be slightly different), but targeted m/z corresponding to that tiny peak seems to be an isotopic peak of a specie with (m-1)/z, as shown in Figure 4.27.

Inconsistent data makes difficult to find a coherent conclusion on the fate of that molecules. Further validation experiments should be performed to validate those results (4.5).

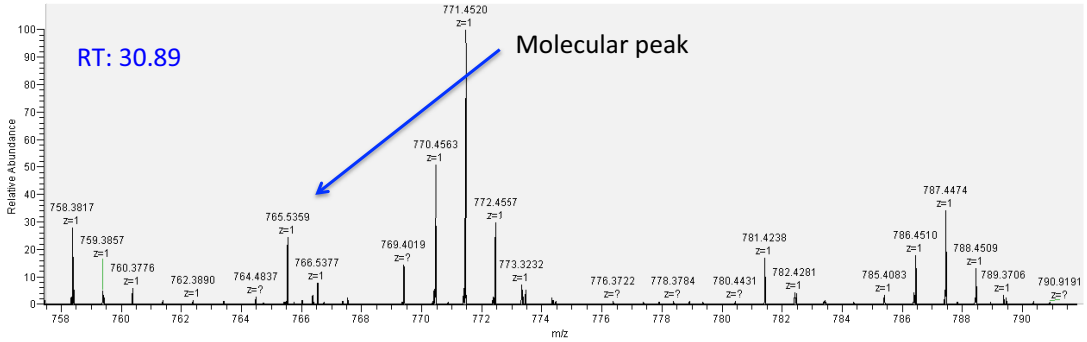
A)



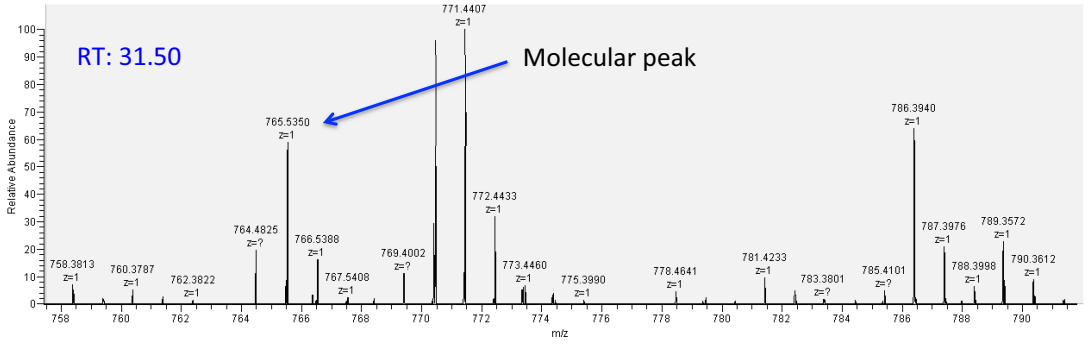
B)



C)



D)



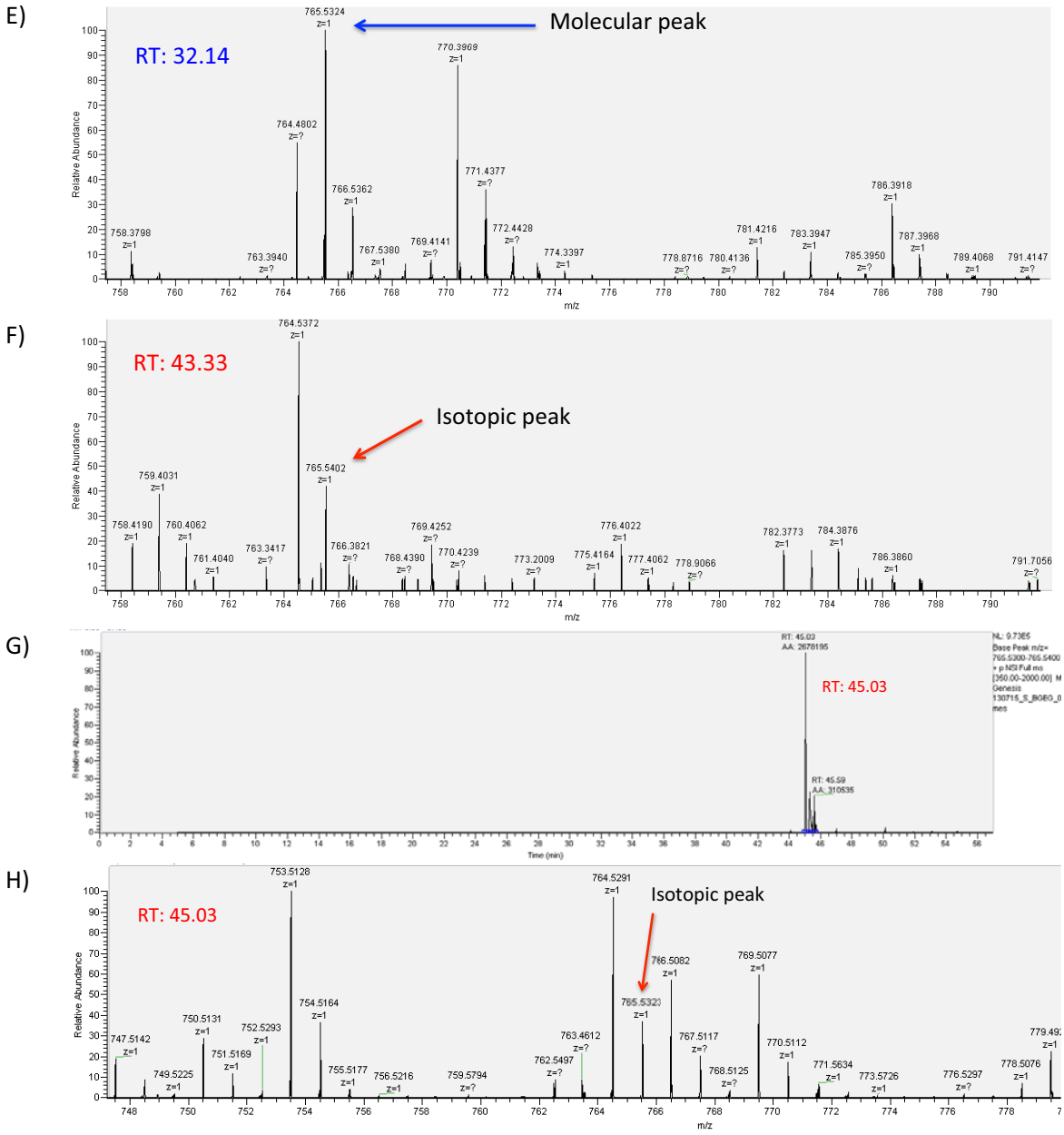


Figure 4.27. A-F) Selected ion chromatogram (SIC) from *in vitro* sample on LTQ-Orbitrap of m/z signals comprised between 765.53 and 765.54, where I4-A1-R1 family might be present. Main peaks appear at RT: 29.30; 30.89; 31.50; 32.14 minutes. In all these peaks 765.53 m/z signal is due to a molecular peak. Another peak (much tinier) appears at RT: 43.33 which should correspond to the one appearing at RT: 45.00 minutes on QqQ spectra (*in vivo*). However, 765.53 m/z peak correspond to an isotopic peak, meaning this peak is contribution of another specie in the spectra. G,H) Similarly, selected ion chromatogram from *in vivo* sample on LTQ-Orbitrap of m/z signals comprised between 765.53 and 765.54 was regarded and only peak with RT: 45.03 minutes was found. Again 765.53 m/z signal correspond to an isotopic peak. No signals between RT: 29 and 33 minutes were found, meaning I4-A1-R1 family was not present in brain homogenate injection.

4.4.5.3. Steps to HTSM *in vivo* identification

On the attempt to directly high-throughput screen Ac-D-Arg-library *in vivo* in mice we focused on identifying BBB-shuttle peptides. By studying in parallel peptide identification in brain, as well as in liver and kidney, we explored both the most restrictive peptide-accepting organ (brain) and the most peptide-retaining organs (liver and kidney) to picture up our technique frame.

In the next section, we discuss what we found with the present library and some hints on new library designs that might better suit technique performance for peptide library identification. Brain, liver and kidney spin micro-filtered cleaned samples (6 hours circulation time) were

injected to LTQ-Orbitrap and mass list were analysed with *Bibliopepfinder*. We began analysing kidney sample as we expected to find much more library peptides than in brain.

Kidney

Kidney blank signals matching possible families from the library were deleted from LTQ-Orbitrap analysed *in vivo* kidney sample (22%). *Bibliopepfinder* assigned 121 families to 109 m/z signals in orbitrap spectrum, 97 m/z signals matched one single family each while 12 m/z signals could be assigned to two families each.

See Appendix 9
Ac-D-Arg-library families found in Kidney.

Since we did not have information from triplicates, applied family selection criteria was simpler that for *in vitro* samples. Then, only two criteria were regarded: first, family group must fit compositions such as 5-1, 4-2, 4-1-1, 3-3, 3-2-1 or 2-2-2. Second, *m/z* peak must appear as molecular peak in LTQ-Orbitrap spectra.

Following those criteria. 5 families were selected for further Q-trap analysis [family members in brackets]: P4-E1-R1[5], A4-R2[5], E3-W2-R1[10], E3-A2-R1[10] and E4-I1-R1[5]. *b*-ion and *y*-ion series transitions were generated for both singly and doubly charged precursor ions.

Unfortunately no sequence confirmation was achieved since not enough transitions co-eluting were found in Q-trap spectra to unequivocally outline any specific sequence.

Liver

No blank signals matching possible families from the library were found from LTQ-Orbitrap analysed *in vivo* liver sample. *Bibliopepfinder* assigned 71 families to 65 m/z signals in LTQ-Orbitrap spectrum, 59 m/z signals matched one single family each while 6 m/z signals could be assigned to two families each. Surprisingly very few peptide signals were found in liver sample. Contrarily, liver is an organ where most peptides are described to be retained previous body elimination. Thus suggesting cleaning sample protocol should be improved to do not miss peptides. Signals found were really tiny (Intensity < 5000) and no further analysis was performed.

See Appendix 10
Ac-D-Arg-library families found in Liver.

Those few peaks with higher intensities showed peak profiles most likely corresponding to endogenous species (high relative intensity). We postulated this hypothesis even though blank sample did not show any or those peaks to be present because we could not envision any other reason.

Brain

Only two blank signals matching possible families from the Ac-D-Arg-library were found from LTQ-Orbitrap analysed *in vivo* brain sample. *Bibliopepfinder* assigned 68 families to 63 m/z signals in LTQ-Orbitrap spectrum, 58 m/z signals matched one single family each while 5 m/z signals could be assigned to two families each.

See Appendix 11
Ac-D-Arg-library families found in Brain.

Although intense enough peptides do not appear in brain blank sample, relative intensity within LTQ-Orbitrap *in vivo* brain spectra suggests they must be endogenous peptides. So that, no further analysis was considered until cleaning protocol improvement.

To sum up, we were not able neither to confirm any sequence from selected families in kidney sample nor find trustable family candidates for brain and liver samples. Then, we concluded even working with state-of-the-art mass spectrometry instruments such as QqQ with high sensitivity; this *in vivo* methodology needs slightly modified library design to facilitate cleaning sample step. Cleaning sample step is a milestone in this *in vivo* approach and library design is one of the options may help to overcome this issue. In section 4.5 we discuss some hints as proposed steps to a next generation library design.

4.5. Prospects

As a future road book, here we propose some points that should be addressed to further validate results presented in this thesis. Moreover, proposal of a new library design is discussed.

4.5.1. Identification validation and quantification

Despite of the high selectivity of QqQ SRM analysis thanks to the two consecutive mass filtering stages performed in Q_1 and Q_3, a particular precursor/fragment combination may not be specific for a peptide targeted in a complex and redundant sample such as Ac-D-Arg-library. Unspecific signals may correspond to other peptides with transitions (precursor/fragment ion pairs) of similar masses. These peptides might have closely related sequences, since members of the same family share part of the transitions. Moreover, completely unrelated sequences might, by chance, have transitions with mass too close to be sufficiently filtered out in the consecutive quadrupoles (0.7 Da filtering window). This might be specially pronounced when analysing *in vitro* or *in vivo* samples with a huge amount of endogenous peptides.

Furthermore, peaks due to non-canonical fragments or natural isotope distribution increase the chances of unspecific signals. These peaks are usually depreciable in proteomic experiments unless targeted peptides are several orders of magnitude less abundant than the most abundant peptides. However, in the present case those peaks cannot be completely ruled out since transitions of target peptides appear in low intensities (10^3-10^5).

Therefore, it is important to validate the results to ensure all transition set appearing co-eluting in a given retention time correspond to the target sequence.

At least eight data points per peak are required for precise quantification. So that, dwell time and cycle time should be adjusted to that purpose, (*see* section 4.3.2.3).

Again, most validation strategies applied for proteomic approach usually involving acquisition and deeply study of specific MS/MS spectra are not fully implementable in our synthetic library.[37]

However, heavy isotope-labelled peptides matching the target peptide sequence spiking into the sample will exactly co-elute with the light version peptide. Moreover, the signal intensity ratios of transitions corresponding to each version are the same. We believe this strategy, also used in proteomics, is the most suitable for validation of Ac-D-Arg-library results.

Interestingly, this strategy can be used for both validation and quantification of targeted signals. Similarly, here we suggest two other approaches that will address only one of those issues: validation or quantification.

4.5.1.1. Validation and quantification: Heavy version synthesis of targeted sequences

This approach involves testing Ac-D-Arg-library or any other synthesized library. Synthesis of heavy-versions of candidate sequences would be performed. Spiking of the tested sample (Ac-D-Arg-library) with heavy peptide with known concentration would both allow validation and quantification. On one hand, validation would be confirmed if heavy-peptide monitored transitions co-elute and have the same intensity ratio profile with those of the identified peptide. On the other hand, quantification would be achieved by calculating the relative ratios between the same selected transitions of pair-peptides (light-heavy). This quantification method is usually referred as SID (stable-isotope dilution) and share the same basis of AQUA approach proposed for proteomic quantification.[51]

This approach would confirm which co-elution signal sets would really correspond to the target sequence. Then it would be possible to further understand crowded spectra such as those found for I3-A1-P1-R1 family.

4.5.1.2. Validation: Resynthesis of light-version selected peptides

In case in which only validation but not quantification is desired synthesis of light version (the peptide version comprised in the library) is only required. Independent injection of resynthesized peptide with the very same conditions would validate the RT and transition ratio profile of targeted peptides. Being enough for validation, this strategy significantly decreases the cost of resynthetic peptides. It could be argued that label-free quantification strategies could be considered with this approach. However, we discourage this option since differences of assayed and resynthesis samples are significant. It has been described that sample matrix and background can dramatically affect ionization and transition intensities, thus hampering the comparison of target analytes.[37] Furthermore, single peptide resynthesized would not suffer from peptide ionization competition as the same peptide comprised in the complex mixture of analysed library.

4.5.1.3. Quantification: Synthesis of a isotopically-labelled library

This strategy consist on the synthesis of a whole heavy library by splitting the resin on the last step. An equivalent library containing all peptides pairs is achieved. Then, a strategy inspired in SILAC (stable isotopic labelling by amino acids in cell culture) proteomic approach could be carried out.[52] By spiking with the same/proportional amount of heavy library in both t_0 and acceptor compartments of the *in vitro* assay or organ lysate in case of *in vivo* assays, relative quantification could be achieved in a similar approach as the one used in chapters 1 and 2. However, this approach is unable to validate peaks present on QqQ spectra since shares the same complexity of its light-version counterpart.

4.5.1.4. Quantification at first MS level: LTQ-Orbitrap

The previous discussion was proposed in terms of QqQ equipment, which offers higher sensitivity than LTQ-Orbitrap. However, since injection in LTQ-Orbitrap is performed in the first step of identification it could be considered a faster quantification method based on that technique in the MS1 level.

High-throughput quantification of specific sequences cannot be achieved with this approach because MS1 is unable to distinguish among distinct sequences with the same composition. Quantification approach could be carried out in terms of family transport rate using the whole isotopically-labeled library. An additional evident drawback of this approach is the significant increase of complexity in the library. While a stringent assay limited the complexity of the light-library the addition of the heavy-library spikes all peptides. Then, LTQ-Orbitrap undergoes less sensitive measurements and low abundance peptides would not be distinguished.

Alternatively, resynthesis of specific identified sequences could be performed in their light and heavy versions. Then, a new assay should be performed with the light peptide version alone. Standard SID approach, as described in section 1.5, could be performed by spiking a known amount of heavy-peptide. MS1 spectra would be enough in terms of sensitivity, because only synthesized sequence is present in the sample and no other members of the family could influence the measurement.

Acetyl tag used has a further drawback that could affect on SRM quantification since carry 3 deuterium atoms on each acetyl moiety. Deuterium containing isotopomers are reported to have slight differences on chromatographic retention time in reverse phase HPLC, while ^{13}C, ^{15}N, ^{18}O isotope-containing tags are far less pronounced and can generally be omitted.[53, 54] Thus, deuterium containing isotopic tags are discouraged on LC-MS or LC-MS/MS quantification strategies.

4.5.2. Designing next generation library: OBOC 2.0

Two main improvements should be included in second-generation peptide library to better face challenges on that new approach. On one hand, peptide recovery after *in vivo* injection must be far more efficient. On the other hand, since these total mix and split peptide libraries are extraordinarily complex for MS/MS identification (even with the new methodology described above) the new library should contain peptides with broader physicochemical properties among them to facilitate peptide separation and differentiation. With peptides ranging a wider physicochemical space there will be more chances for accurate identification. Even though those libraries will screen a broader physicochemical space, certainly it will be covered less exhaustively.

To address the first improvement, we suggest to implement a technology to increase peptide recovery within complex *in vivo* mixtures such as biotin decorated peptides. Biotin decorated peptides can be captured and separated from tissue lysate by streptavidin-coated magnetic beads. This method developed to quantify transport of cell-penetrating peptides by Burlina *et al*. could be here implemented for both identification and quantification purpose.[55]

To address the second stated improvement, we suggest the following peptide scaffold design:

$$\text{Biotin-K(Ac/Ac*)-K(Ac/Ac*)-K-G-G-}X_1\text{-G-G-}X_2\text{-G-G-}X_3\text{-G-G-K-NH}_2$$

where Ac* represents isotopically labelled acetyl tags and all scaffold amino acids being D-amino acids. Mix and split methodology would only be implemented in X positions. Then, three distinct groups of amino acids would be selected to fill up each X position.

Amino acids are usually grouped depending on their physicochemical properties. Those amino acids with similar properties will be placed in separate groups. Then, each set of amino acids must contain a single amino acid representative of each physicochemical group.

The following selection is proposed for a first trial library:

X_1: D-Arg, D-Glu, D-Ser, D-Phe, D-Asn and D-Ile.

X_2: D-Lys, D-Asp, D-Ala, D-Tyr, D-Gln and D-Leu.

X_3: D-His, D-Thr, D-Trp, D-Pro, D-Val and D-Gly.

Thus, while restricting synthetic amino acid positions we are screening a broader physicochemical space.

Peptide terminus, *C*-terminus and *N*-terminus, are planned to be amidated and biotinylated, respectively. Thus, no charge will be able to be stabilized on those terminus positions on ionization. To assist peptide ionization, two free lysine residues will be located in the *N*-terminus

and *C*-terminus sides. On one hand, they improve precursor ion chances to be analysed by MS detector. On the other hand, they improve both *b*-ion and *y*-ion series detection on MS/MS facilitating sequence identification.

For quantification purpose, two lysine residues will be located next to biotin to introduce the acetyl tag in their side chains. Isotopically label acetyl tag can be synthetically easily introduced in those *N*-terminus peptide positions. Two lysines increase heavy versions up to 8 atomic mass units, which are required since precursor ion will be most likely doubly charged. To introduce acetyl tag within the synthesis, those lysines must be protected with any orthogonal Fmoc/*t*Bu protecting group such as *Alloc*. These acetyl groups can be also considered as the most basic and tiniest cargo for such shuttles.

Of course, this design does not solve all problems chemically synthesized peptide libraries may have but we strongly believe that will increase both peptide library performance and peptide identification with mass spectrometry techniques.

Bibliography

1. Dolle, R.E., Le Bourdonnec, B., Goodman, A.J., Morales, G.A., Thomas, C.J. & Zhang, W. Comprehensive survey of chemical libraries for drug discovery and chemical biology: 2008. *J. Comb. Chem.* **2009**, 11, 739-790.
2. Dolle, R.E., Bourdonnec, B.L., Worm, K., Morales, G.A., Thomas, C.J. & Zhang, W. Comprehensive survey of chemical libraries for drug discovery and chemical biology: 2009. *J. Comb. Chem.* **2010**, 12, 765-806.
3. Kennedy, J.P., Williams, L., Bridges, T.M., Daniels, R.N., Weaver, D. & Lindsley, C.W. Application of combinatorial chemistry science on modern drug discovery. *J. Comb. Chem.* **2008**, 10, 345-354.
4. Kodadek, T. The rise, fall and reinvention of combinatorial chemistry. *Chem. Commun.* **2011**, 47, 9757-9763.
5. Lopez-Vallejo, F., Giulianotti, M.A., Houghten, R.A. & Medina-Franco, J.L. Expanding the medicinally relevant chemical space with compound libraries. *Drug Discov. Today* **2012**, 17, 718-726.
6. Hou, J., Liu, X., Shen, J., Zhao, G. & Wang, P.G. The impact of click chemistry in medicinal chemistry. *Expert Opin. Drug Discov.* **2012**, 7, 489-501.
7. Furka, A., Sebestyen, F., Asgedom, M. & Dibo, G. General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Pept. Protein Res.* **1991**, 37, 487-493.
8. Lam, K.S., Salmon, S.E., Hersh, E.M., Hruby, V.J., Kazmierski, W.M. & Knapp, R.J. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **1991**, 354, 82-84.
9. Houghten, R.A., Pinilla, C., Blondelle, S.E., Appel, J.R., Dooley, C.T. & Cuervo, J.H. Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature* **1991**, 354, 84-86.
10. Geysen, H.M., Rodda, S.J. & Mason, T.J. A priori delineation of a peptide which mimics a discontinuous antigenic determinant. *Mol. Immunol.* **1986**, 23, 709-715.
11. Zhao, P.L., Nachbar, R.B., Bolognese, J.A. & Chapman, K. Two new criteria for choosing sample size in combinatorial chemistry. *J. Med. Chem.* **1996**, 39, 350-352.
12. Zhao, P.L., Zambias, R., Bolognese, J.A., Boulton, D. & Chapman, K. Sample size determination in combinatorial chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, 92, 10212-10216.
13. Smith, G.P. & Petrenko, V.A. Phage Display. *Chem. Rev.* **1997**, 97, 391-410.
14. Web site: http://mathworld.wolfram.com/SpherePacking.html, 25th October 2013.
15. Jaeger, H.M. & Nagel, S.R. Physics of the granular state. *Science* **1992**, 255, 1523-1531.
16. Belda, I. Final degree project. Química combinatoria virtual: disseny de pèptids que travessen la barrera hematoencefàlica. **2001**, Universitat Ramon Llull.
17. Chou, P.Y. & Fasman, G.D. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1978**, 47, 45-148.
18. Kyte, J. & Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, 157, 105-132.
19. Glaser, F., Steinberg, D.M., Vakser, I.A. & Ben-Tal, N. Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins* **2001**, 43, 89-102.
20. Palmblad, M., Drijfhout, J.W. & Deelder, A.M. High resolution mass spectrometry for rapid characterization of combinatorial peptide libraries. *J. Comb. Chem.* **2010**, 12, 65-68.
21. Demirev, P.A. & Zubarev, R.A. Probing combinatorial library diversity by mass spectrometry. *Anal. Chem.* **1997**, 69, 2893-2900.
22. Nawrocki, J.P., Wigger, M., Watson, C.H., Hayes, T.W., Senko, M.W., Benner, S.A. & Eyler, J.R. Analysis of combinatorial libraries using electrospray Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* **1996**, 10, 1860-1864.
23. Marshall, A.G., Hendrickson, C.L. & Jackson, G.S. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **1998**, 17, 1-35.
24. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, 422, 198-207.
25. Eng, J.K., McCormack, A.L. & Yates III, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, 5, 976-989.
26. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, 20, 3551-3567.
27. Xu, H. & Freitas, M.A. MassMatrix: a database search program for rapid characterization of proteins and peptides from tandem mass spectrometry data. *Proteomics* **2009**, 9, 1548-1555.

28. Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L., Xu, M., Maynard, D.M., Yang, X., Shi, W. & Bryant, S.H. Open mass spectrometry search algorithm. *J. Proteome Res.* **2004**, 3, 958-964.
29. Craig, R., Cortens, J.P. & Beavis, R.C. Open source system for analyzing, validating, and storing protein identification data. *J. Proteome Res.* **2004**, 3, 1234-1242.
30. Wu, R., Haas, W., Dephoure, N., Huttlin, E.L., Zhai, B., Sowa, M.E. & Gygi, S.P. A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat. Methods* **2011**, 8, 677-683.
31. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **2012**, 9, 555-566.
32. Laskin, J. & Futrell, J.H. Collisional activation of peptide ions in FT-ICR mass spectrometry. *Mass Spectrom. Rev.* **2003**, 22, 158-181.
33. Wells, J.M. & McLuckey, S.A. Collision-induced dissociation (CID) of peptides and proteins. *Meth. Enzymol.* **2005**, 402, 148-185.
34. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass. Spectrom.* **1984**, 11, 601.
35. Biemann, K. Mass spectrometry of peptides and proteins. *Annu. Rev. Biochem.* **1992**, 61, 977-1010.
36. Zhang, Z. Prediction of low-energy collision-induced dissociation spectra of peptides. *Anal. Chem.* **2004**, 76, 3908-3922.
37. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **2008**, 4, 222.
38. Stahl-Zeng, J., Lange, V., Ossola, R., Eckhardt, K., Krek, W., Aebersold, R. & Domon, B. High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol. Cell. Proteomics* **2007**, 6, 1809-1817.
39. Kansy, M., Senner, F. & Gubernator, K. Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. *J. Med. Chem.* **1998**, 41, 1007-1010.
40. Lu, H., Demny, S., Zuo, Y., Rea, W., Wang, L., Chefer, S.I., Vaupel, D.B., Yang, Y. & Stein, E.A. Temporary disruption of the rat blood-brain barrier with a monoclonal antibody: a novel method for dynamic manganese-enhanced MRI. *NeuroImage* **2010**, 50, 7-14.
41. Joshi, S., Ergin, A., Wang, M., Reif, R., Zhang, J., Bruce, J.N. & Bigio, I.J. Inconsistent blood brain barrier disruption by intraarterial mannitol in rabbits: implications for chemotherapy. *J. Neuro-Oncol.* **2011**, 104, 11-19.
42. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C. & MacCoss, M.J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, 26, 966-968.
43. Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **2008**, 24, 2534-2536.
44. Pasqualini, R. & Ruoslahti, E. Organ targeting in vivo using phage display peptide libraries. *Nature* **1996**, 380, 364-366.
45. Bertrand, Y., Currie, J.C., Demeule, M., Regina, A., Che, C., Abulrob, A., Fatehi, D., Sartelet, H., Gabathuler, R., Castaigne, J.P., Stanimirovic, D. & Beliveau, R. Transport characteristics of a novel peptide platform for CNS therapeutics. *J. Cell. Mol. Med.* **2010**, 14, 2827-2839.
46. Triguero, D., Buciak, J. & Pardridge, W.M. Capillary Depletion Method for Quantification of Blood-Brain Barrier Transport of Circulating Peptides and Plasma Proteins. *J. Neurochem.* **1990**, 54, 1880-1888.
47. Boado, R.J. & Pardridge, W.M. A One-Step Procedure for Isolation of Poly(A)+ mRNA from Isolated Brain Capillaries and Endothelial Cells in Culture. *J. Neurochem.* **1991**, 57, 2136-2139.
48. Staquicini, F.I., Ozawa, M.G., Moya, C.A., Driessen, W.H., Barbu, E.M., Nishimori, H., Soghomonyan, S., Flores, L.G., 2nd, Liang, X., Paolillo, V., Alauddin, M.M., Basilion, J.P., Furnari, F.B., Bogler, O., Lang, F.F., Aldape, K.D., Fuller, G.N., Hook, M., Gelovani, J.G., Sidman, R.L., Cavenee, W.K., Pasqualini, R. & Arap, W. Systemic combinatorial peptide selection yields a non-canonical iron-mimicry mechanism for targeting tumors in a mouse model of human glioblastoma. *J. Clin. Invest.* **2011**, 121, 161-173.
49. Burden, D.W. Guide to the Disruption of Biological Samples – 2012 *Random Primers* **2012**.
50. Web site: http://www.pressurebiosciences.com/downloads/publications/Not_the_same_old_grind.pdf, 3rd Desember 2013.
51. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. & Gygi, S.P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 6940-6945.

52. Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. & Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, 1, 376-386.
53. Zhang, R., Sioma, C.S., Wang, S. & Regnier, F.E. Fractionation of isotopically labeled peptides in quantitative proteomics. *Anal. Chem.* **2001**, 73, 5142-5149.
54. Zhang, R. & Regnier, F.E. Minimizing resolution of isotopically coded peptides in comparative proteomics. *J. Proteome Res.* **2002**, 1, 139-147.
55. Burlina, F., Sagan, S., Bolbach, G. & Chassaing, G. A direct approach to quantification of the cellular uptake of cell-penetrating peptides using MALDI-TOF mass spectrometry. *Nat. Protoc.* **2006**, 1, 200-205.

CONCLUSIONS

Conclusions

1- Regarding the stereoisomeric peptide library, we found that chirality plays a significant role in molecule transport across the PAMPA, an assay that exclusively mimics passive diffusion conditions of the BBB. The homochiral compounds showed greater permeability than the heterochiral ones. Enantiomeric discrimination was present in some cases, but we were unable to detect any global trend. The all-D version of Ac-(*N*-MePhe)$_4$-NH$_2$ showed better permeability than the reported all-L version. Moreover, there was an inverse correlation between membrane retention and permeability. The lower the peptide retention by the membrane, the better the permeability. Retention time on HPLC was used as a measure of molecule polarity since peptides of the same family were compared. Heterochiral versions were observed to be more hydrophobic than their homochiral counterparts.

2- Peptides in the HAIYPRH analogue library did not show significant permeability differences compared to the original peptide. However, in general, slightly higher permeability values were achieved for D-library than for L-library molecules. This finding thus suggests improved stability of D-library peptides in the *in vitro* cell-based BBB model.

3- MALDI-TOF mass spectrometry was successfully applied to measure peptide transport across *in vitro* assays modelling the BBB (PAMPA and cell-based BBB model) by using isotopically labelled pair-peptides (light-heavy). Acetyl tag, and its 4 a.m.u. isotopically labelled heavy version ($CD_3{}^{13}CO$-) were chosen for isotopic labelling of library peptides for individual transport assessment.

4- A simplified MALDI-TOF transport quantification protocol was set up in such a way as to avoid the need for a calibration curve and thus facilitate high-throughput transport assessment of relatively large libraries with minimal precision loss. Selected compounds showing the most promising transport values require an exhaustive MALDI-TOF transport quantification protocol including calibration curves for faithful transport assessment.

5- A CX_3C phage display peptide library was successfully synthesized and quality control-validated. A phage library displaying tripeptides constrained by cysteines inserted on the pIII protein of fd-phage were tested *in vivo* in BALB/c mouse models. We assessed biodistribution and brain targeting through three biopanning rounds and an extra post-panning round. CFLFC-phage, CRWEC-phage, and CNSQC-phage were selected as promising sequences targeting the brain.

6- Biodistribution experiments by q-PCR sequencing of phage-DNA, by testing each selected phage individually in BALB/c mice, revealed more than a one-fold higher abundance than non-displaying insert phage. On the other hand, accumulation of each individual phage in the brain was three-fold lower than in liver.

7- Immunostaining of fixed brain slides and confocal imaging of individually injected phages to BALB/c mice revealed their promising features as BBB-shuttles. A colocalization experiment with CD31 (endothelial cell marker) did not work as expected. However, the shapes and location of signals corresponding to the injected phage with respect to the cell nucleus (DAPI staining) reveal encouraging scenarios in which these phages target specific sites inside the brain parenchyma, thus crossing the BBB.

8- A mix and split peptide library, Ac-D-Arg-library, was successfully synthesized and validated. This library was assayed on PAMPA and in an *in vitro* cell-based BBB model.

9- A novel high-throughput screening identification method based on mass spectrometry was set up to identify the most promising peptides within the library crossing the *in vitro* assay. A workflow based on two MS-identification levels was established. The first MS-step consisted of choosing the peptide species corresponding to specific compositions (peptide families) by means of LTQ-Orbitrap MS, operating only on MS1 level. Subsequently, targeted mass spectrometry (SRM) was applied as the second MS-identification level by monitoring transitions defining specific amino acid sequences. Hence, peptide sequences (family members) within the selected composition were unequivocally validated.

10- On the basis of data obtained from the PAMPA and *in vitro* cell-based BBB model, we propose that some peptide families cross the BBB by *passive diffusion transport* while others use *other mechanisms of transport* (including active transport and paracellular flux). Most sequences in the I3-A1-P1-R1 and I4-P1-R1 families corresponding to *other mechanisms of transport* were validated in SRM experiments. Similarly, most sequences in the I3-P2-R1, P3-S2-R1 and A3-R2-P1 families corresponding to *passive diffusion transport* were also validated in SRM experiments. Further validation and quantification experiments must be performed.

11- *In vivo* evaluation of the Ac-D-Arg-library was performed in BALB/c mice, and a high-throughput screening method was applied to brain, liver, and kidney samples. Specific compositions matching peptides in the library were found only in kidney in the first MS-identification level, although they could not be validated by SRM at the second identification level. No significant results were achieved for brain and liver samples. A sample cleaning step seems to be the cornerstone for the success of the project. For this reason, we postulate a new library design to facilitate organ and tissue sample cleaning in order to achieve a sufficient amount of peptide for successful MS identification and quantification.

MATERIALS AND METHODS

1. Solid-phase peptide synthesis (SPPS) materials and methods

1.1. Solvent and reagents

All the reagents employed in the present section have been purchased to following suppliers: Anaspec, Bachem AG, Calbiochem-Novabiochem AG, Iris Biotech, KaliChemie, Luxemburg Industries, Neosystems and Sigma-Aldrich. All solvents employed on the present section have been purchased to following suppliers: Panreac, Scharlau, SDS and Sigma-Aldrich.

1.2. Instrumentation

1.2.1. Liquid chromatography

1.2.1.1. HPLC-UV

HPLC chromatograms were recorded on a Waters Alliance 2695 separation module coupled with a Waters 2998 photodiode array detector and Empower software (Waters) using a Sunfire C_{18} column (100 mm x 4.6 mm, 3.5 µm; Waters). Column flow was set at 1 mL/min, and solvents were H_2O (0.045% TFA) and MeCN (0.036% TFA). Samples sere previously filtered by a Millipore system with a filter of 0.45 µm.

1.2.1.2. Semi-preparative HPLC

Peptides were purified by semi-preparative HPLC on a Waters 2700 Sample manager equipped with a Waters 2487 dual λ absorbance detector, a Waters 600 controller, a Waters fraction collection II and Millenium chromatography manager software (Waters) using a Symmetry C_{18} column (100 mm x 19 mm, 5 µm; Waters). Column flow was set at 15 mL/min, and solvents were H_2O (0.1% formic acid) and MeCN (0.07% formic acid).

1.2.1.3. HPLC-MS

HPLC-MS were recorded using a Waters Alliance 2695 equipped with a Waters 2998 photodiode array detector, ESI-MS micromass ZQ (Waters) and Masslynx software (Waters) using a Sunfire C_{18} column (100 mm x 2.1 mm, 3.5 µm, Waters). Column flow was set at 0.3 mL/min, and solvents were H_2O (0.1% formic acid) and MeCN (0.07% formic acid).

1.2.2. MALDI-TOF

MALDI-TOF mass spectra were obtained with two instruments. If not stated in the text, the instrument used was a MALDI-TOF/TOF Applied Biosystem 4700 with a N_2 laser of 337 nm. In the text might be referred as MALDI-4700. On the other hand, an MALDI-TOF/TOF Applied Biosystem 4800 with a laser of 355 nm, 200Hz and a poulse of 3-7 ns. In the text, this instrument is referred as MALDI-4800.

Sample was prepared by mixing a solution of peptide to be analysed (1 µL) and MALDI matrix (1 µL) and plating the mixture in a MALDI plate and dried by air. MALDI matrix was prepared by dissolving 10 mg of α-cyano-4-hydroxycinnamic acid (ACH) in H_2O/MeCN (1:1) containing 0.1% of TFA.

1.2.3. High-resolution mass spectroscopy (FT-ICR)

Exact masses were determined by high-resolution mass spectroscopy recorded on a LTQ-FT Ultra (Thermo Scientific). The Triversa NanoMate (Advion BioSciences, Ithaca, NY, USA) aspirated the samples from a 384-well plate with disposable, conductive pipette tips, and infused the samples through the ESI Chip, wich consists on 400 nozzles in a 20x20 array. Spray voltage ws 1.75 kV and delivery pressure 0.5 psi. Capillary voltage was 3 KV; sampling cone, 45 V; source temperature, 80ºC; Trap collision energy, 6; and transfer collision energy, 4. Screened m/z range was comprised between 200 and 2000 a.m.u, calibrated with GluFib (m/z range 200 to 2000); RF: Source 350, IMS 350, Trap, 350 and Transfer 350. Lack Spray was GluFib 8 pmols/µL; reference scan frequency, 15 seconds; and reference cone voltage, 35 V. Data was acquired with Xcalibur software, vs.2.0SR2 (ThermoScientific).

1.3. Solid-phase peptide synthesis (SPPS)

1.3.1. General considerations

All peptides were synthesized by SPPS using the 9-fluorenylmethoxycarbonyl/*tert*-butyl (Fmoc/*t*Bu) strategy.[1] In all cases peptide elongation and other manipulations were done manually in polypropylene syringes provided with a polyethylene porous disk. Solvents and soluble reagents were removed by suction. Washings between synthetic steps were carried out with DMF (5 x 30 s) and DCM (5 x 30 s) using approximately 5 mL of solvent/g of resin. During coupling and deprotection reactions the mixture was allowed to react with intermittent manual stirring (Teflon stick).

1.3.2. Colorimetric tests

1.3.2.1. Ninhydrin or Kaiser test[2]

The Kaiser test, also known as ninhydrin test, is a method used for free primary amines detection widely used in SPPS to qualitatively determine the extent of coupling and deprotection reactions on primary amines. Peptide-resin is washed with DCM to avoid primary amines traces on DMF and few resin beads are transferred to a small glass tube where the reaction is performed. Six drops of reagent solution A and 2 drops of reagent solution B are added to the glass tube and mixture was allowed to react with resin beads at 110ºC for 3 minutes. Formation of a blue colour on beads or on the supernatant indicates the presence of free primary amines, while yellowish colour (no colour change) indicates negative test. The method is highly sensitive and negative test ensures a good coupling rate.

Reagent solution A: 400 g of phenol are dissolved in 100 mL of absolute EtOH and the mixture is heated until complete dissolution of the phenol. 20 mL of 10 mM KCN (65 mg in 100 mL of H_2O) are added to 1 L of freshly distilled pyridine over ninhydrin. Both solutions are stirred for 45 minutes with 40 g of Amberlite MB-3 ion exchange resin, filtered and mixed.

Reagent solution B: 2.5 g of ninhydrin are dissolved in 50 mL of absolute EtOH and resulting solution is stored in a flask protected from light.

1.3.2.2. Chloranil test[3]

The chloranil test is used for detection of secondary amines and is widely used on SPPS for monitoring coupling and deprotection reactions involving secondary amines. Peptide-resin is washed with DCM and solvent filtered by suction. Few resin beads are transferred to a small glass tube where 20 µL of saturated chloranil solution (750 µg of 2,3,5,6-tetrachloro-1,4-benzoquinone in 25 mL of toluene) and 200 µL of acetone are added. Mixture is thoroughly stirred for 5 minutes at room temperature. Formation of a blue-greenish colour indicates the presence of free secondary amines, while yellow, amber or brown colours are considered negative test.

1.3.2.3. De Clercq test or *p*-nitrophenyl ester test[4]

De Clercq test is used for detection of secondary amines with higher sensitivity than the chloranil test.

The peptide-resin is washed with MeOH and solvent filtered by suction. Few resin beads are transferred to a small glass tube with 5 drops of De Clercq reagent. The suspension is heated at 70ºC for 8 minutes and resin beads are washed with MeOH (3 times), DMF (3 times) and DCM (once). The presence of free secondary amines is indicated by red-stained resin beads whereas no colour change on beads is considered a negative test.

Reagent synthesis: A three-step synthesis leads to *p*-nitrophenyl ester from commercially available disperse red 1. A solution of ethyl diazoacetate (8.4 mL, 80 mmol) in toluene (40 mL) is added to a solution of disperse red 1 (6.28 mg, 20 mmol) and $Rh_2(OAc)_4$ (150 mg, 0.34 mmol) in DCM/toluene (1:1, 200 mL) at 40 ºC over 1 hour. The reaction is performed overnight with continued stirring and purified by flash chromatography. Obtained product (5 g, 12.5 mmol) and KOH (4.062 g 62.5 mmol) are dissolved in MeOH/toluene (4.3:1, 370 mL) and refluxed until under N_2 atmosphere for 90 minutes. The product is purified by a series of extractions. Obtained product (2.322 g, 6 mmol) and *p*-nitrophenol (0.834 g, 6 mmol) are dissolved in pyridine/DCM (1:1.2, 220 mL) at -15ºC and a solution of $POCl_3$ in DCM (10% v/v) slowly added over 1 hour. The *p*-nitrophenyl ester obtained is purified by a series of extractions and it is used at a concentration of 0.02 M in MeCN.

1.3.3. Initial conditioning of the resin

Conditioning of the resin was performed as follows both for Sieber resin and Fmoc-Rink amide AM resin.

Step	Treatment	Reagents / solvents	Time
1	Wash	DCM	5 x 1 min
2	Wash	DMF	5 x 1 min
3	Pre-activation	20% piperidine in DMF	1 x 1 min, 2 x 10 min
4	Wash	DMF	5 x 1 min

1.3.4. Fmoc group removal

The Fmoc group was removed by treating the resin with 20% piperidine in DMF (3-4 mL/g resin, 1 x 1 min and 2 x 10 min). To remove the Fmoc group from Fmoc-*N*-Me-Aa-OH or Fmoc-Pro-OH and additional treatment with DBU, toluene, piperidine and DMF (5:5:20:70) (2 x 5 min) was performed.

1.3.5. Fmoc group quantification/resin loading capacity

Piperidine washes were collected and measured by UV spectroscopy (λ = 290 nm) to determine the loading capacity of the resin after first amino acid coupling. Loading capacity was determined according the equation:

$$Z = \frac{A \cdot X}{\varepsilon \cdot Y \cdot l}$$

A	Absorbance
X	Volume of solvent (mL)
ε	Molar absorbance coefficient (5800 L · mol^{-1} · cm^{-1})
Y	Resin weight (g)
l	Length of the cell (cm)
Z	Loading of the resin

1.3.6. Amino acid coupling and chain elongation

Protocol A was used for coupling onto primary amines, while Protocol B was used for coupling onto secondary amines.

Protocol A:

Step	Treatment	Reagents / solvents	Time
1	Wash	DMF	5 x 1 min
2	Coupling	Fmoc-Aa-OH (4 eq), Oxima pure® (4 eq), COMU (4 eq), DIEA (8 eq) in DMF	1 x 90 min
3	Wash	DMF	5 x 1 min
4	Fmoc group removal	20% piperidine in DMF	1 x 1 min, 2 x 10 min
5	Wash	DMF	5 x 1 min

Protocol B:

Step	Treatment	Reagents / solvents	Time
1	Wash	DMF	5 x 1 min
2	Fmoc group removal	20% piperidine in DMF	1 x 1 min, 2 x 10 min
3	Fmoc group removal	DBU, toluene, piperidine in DMF (5:5:20:70)	2 x 5 min
4	Wash	DMF	5 x 1 min
5	Coupling	Fmoc-Aa-OH (4 eq), Oxima pure® (4 eq), COMU (4 eq), DIEA (8 eq) in DMF	2 x 90 min
6	Wash	DMF	5 x 1 min

Couplings were performed at room temperature with intermittent manual stirring. The extent of coupling reactions was checked by the ninhydrin, *p*-nitrophenylester or chloranil test (depending on amine *N*-terminal nature, primary or secondary, of the last amino acid on the peptide chain). Fmoc deprotection of Proline or *N*-alkyl-Aa residues was carried out using an extra treatment of

DBU, toluene, piperidine and DMF and next amino acid residue introduced in the peptide chain was systematically recoupled.

1.3.7. Peptide *N*-terminus capping

Peptides were capped on resin once peptide chain was fully elongated. Only acetyl or benzoyl tags were used. Acetylation was performed with acetic acid and DIPCDI and benzylation with benzoic acid and DIPCDI. Similarly, isotopically labelled tags were introduced by using CD_3COOD, $CD_3{}^{13}COOD$ and C_6D_5COOH.

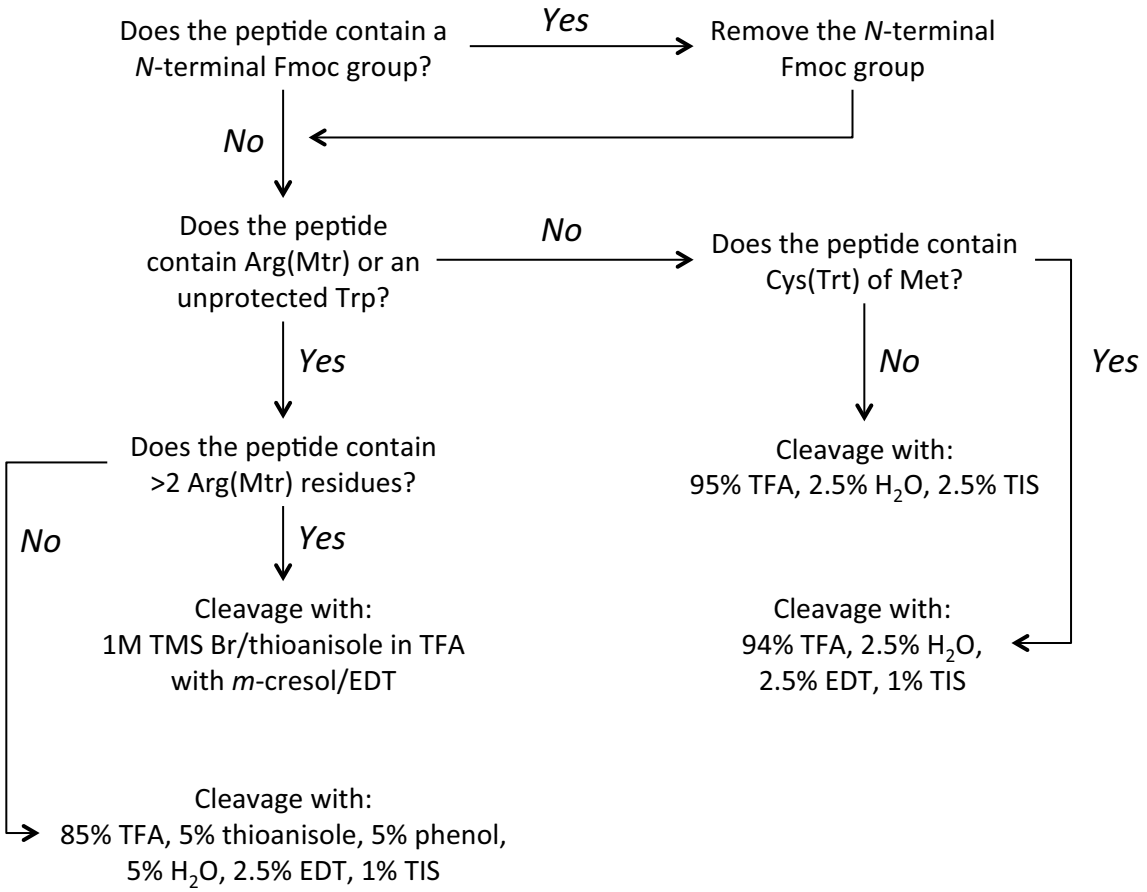
Step	Treatment	Reagents / solvents	Time
1	Wash	DCM	5 x 1 min
2	Coupling	R-COOH (6 eq), DIPCDI (3 eq) in DCM	2 x 15 min
3	Wash	DCM	5 x 1 min

1.3.8. Disulphide bridge formation

After Fmoc group removal, the disulphide bridge was formed under solid-phase conditions by double treatment with I_2 (5 eq) in DMF for 10 min.

1.3.9. Cleavage of peptides from the resin

Peptide cleavage from the resin and deprotection of side chains was performed by acidolytic treatment with TFA solution, since SPPS was carried out by Fmoc/*t*Bu chemistry. Selection of appropriate scavengers on the cleavage cocktail are required for minimize undesired side reactions. The following flowchart was used as a guide to select such cleavage cocktails by means of peptide sequence.



	Peptide	Purification required?	HPLC gradient	HPLC-RT/min	Mw	MALDI	Yield (%)
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-LLLL-NH$_2$	No	AG50100t8	4.0	703.86	726.9 (+Na$^+$)	86.3
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-DDDD-NH$_2$	No	AG50100t8	3.8	703.86	726.9 (+Na$^+$)	78.6
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-LDDD-NH$_2$	No	AG50100t8	4.3	703.86	726.9 (+Na$^+$)	67.3
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-DLLL-NH$_2$	No	AG50100t8	4.3	703.86	726.9 (+Na$^+$)	70.4
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-LDLL-NH$_2$	No	AG50100t8	5.2	703.86	726.9 (+Na$^+$)	65.2
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-DLDD-NH$_2$	No	AG50100t8	5.2	703.86	726.9 (+Na$^+$)	79.5
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-DDLL-NH$_2$	No	AG50100t8	4.7	703.86	726.9 (+Na$^+$)	81.4
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-LLDD-NH$_2$	No	AG50100t8	4.8	703.86	726.9 (+Na$^+$)	59.4
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-LLDL-NH$_2$	No	AG50100t8	4.9	703.86	726.9 (+Na$^+$)	70.5
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-DDLD-NH$_2$	No	AG50100t8	5.0	703.86	726.9 (+Na$^+$)	69.9
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-LDLD-NH$_2$	No	AG50100t8	5.3	703.86	726.9 (+Na$^+$)	75.8
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-DLDL-NH$_2$	No	AG50100t8	5.3	703.86	726.9 (+Na$^+$)	61.2
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-LDDL-NH$_2$	No	AG50100t8	4.5	703.86	726.9 (+Na$^+$)	73.8
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-DLLD-NH$_2$	No	AG50100t8	4.5	703.86	726.9 (+Na$^+$)	84.1
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-DDDL-NH$_2$	No	AG50100t8	4.3	703.86	726.9 (+Na$^+$)	59.3
Ac-(*N*-MePhe)$_4$-NH$_2$ library	Ac-LLLD-NH$_2$	No	AG50100t8	4.2	703.86	726.9 (+Na$^+$)	68.3
Ac-(*N*-MePhe)$_4$-NH$_2$ tags	CD_3COO-LLLL-NH$_2$	No	AG20100t8	6.6	706.88	729.8 (+Na$^+$)	80.1
Ac-(*N*-MePhe)$_4$-NH$_2$ tags	$CD_3{}^{13}COO$-LLLL-NH$_2$	No	AG20100t8	6.7	707.88	730.9 (+Na$^+$)	73.5
Ac-(*N*-MePhe)$_4$-NH$_2$ tags	C_6H_5-LLLL-NH$_2$	No	AG50100t8	5.5	765.94	789.0 (+Na$^+$)	73.7
Ac-(*N*-MePhe)$_4$-NH$_2$ tags	C_6D_5-LLLL-NH$_2$	No	AG50100t8	5.5	770.97	794.0 (+Na$^+$)	78.4
Ac-HAIYPRH-NH$_2$ library	1L	Yes	AG0100t8	3.6	934.1	935.0 (+H$^+$); 957.1 (+Na$^+$); 973.1 (+K$^+$)	11.1
Ac-HAIYPRH-NH$_2$ library	1L*	Yes	AG0100t8	3.7	938.1	939.0 (+H$^+$); 961.1 (+Na$^+$); 977.1 (+K$^+$)	14.9
Ac-HAIYPRH-NH$_2$ library	2L	Yes	AG0100t8	3.8	988.2	989.1 (+H$^+$); 1011.2 (+Na$^+$); 1027.2 (+K$^+$)	15.2
Ac-HAIYPRH-NH$_2$ library	2L*	Yes	AG0100t8	3.8	992.2	993.1 (+H$^+$); 1015.2 (+Na$^+$); 1031.2 (+K$^+$)	16.1
Ac-HAIYPRH-NH$_2$ library	3L	Yes	AG0100t8	3.9	946.1	947.0 (+H$^+$); 969.1 (+Na$^+$); 985.1 (+K$^+$)	12.7
Ac-HAIYPRH-NH$_2$ library	3L*	Yes	AG0100t8	3.9	950.1	951.0 (+H$^+$); 973.1 (+Na$^+$); 989.1 (+K$^+$)	8.4
Ac-HAIYPRH-NH$_2$ library	4L	No	AG0100t8	3.4	933.1	934.0 (+H$^+$); 956.1 (+Na$^+$); 972.1 (+K$^+$)	68.6
Ac-HAIYPRH-NH$_2$ library	4L*	No	AG0100t8	3.4	937.1	938.0 (+H$^+$); 960.1 (+Na$^+$); 977.1 (+K$^+$)	54.4
Ac-HAIYPRH-NH$_2$ library	5L	No	AG0100t8	4.0	936.1	937.0 (+H$^+$); 959.1 (+Na$^+$); 975.1 (+K$^+$)	58.8
Ac-HAIYPRH-NH$_2$ library	5L*	No	AG0100t8	4.0	940.1	941.0 (+H$^+$); 963.1 (+Na$^+$); 979.1 (+K$^+$)	51.1
Ac-HAIYPRH-NH$_2$ library	6L	Yes	AG0100t8	4.2	986.1	987.1 (+H$^+$); 1009.2 (+Na$^+$); 1025.2 (+K$^+$)	10.1
Ac-HAIYPRH-NH$_2$ library	6L*	Yes	AG0100t8	4.2	990.1	991.1 (+H$^+$); 1013.2 (+Na$^+$); 1029.2 (+K$^+$)	47.5

	Peptide	Purification required?	HPLC gradient	HPLC-RT/min	Mw	MALDI	Yield (%)
Ac-HAIYPRH-NH_2 library	7L	No	AG0100t8	4.1	968.2	969.0 ($+H^+$); 991.1 ($+Na^+$); 1007.1 ($+K^+$)	56.8
Ac-HAIYPRH-NH_2 library	7L*	No	AG0100t8	4.1	972.2	963.0 ($+H^+$); 995.1 ($+Na^+$); 1011.1 ($+K^+$)	66.9
Ac-HAIYPRH-NH_2 library	8L	No	AG0100t8	3.2	906.0	907.0 ($+H^+$); 929.1 ($+Na^+$); 945.1 ($+K^+$)	43.0
Ac-HAIYPRH-NH_2 library	8L*	No	AG0100t8	3.3	910.0	911.0 ($+H^+$); 933.1 ($+Na^+$); 949.1 ($+K^+$)	57.1
Ac-HAIYPRH-NH_2 library	9L	Yes	AG0100t8	3.7	934.1	935.0 ($+H^+$); 957.1 ($+Na^+$); 973.1 ($+K^+$)	13.9
Ac-HAIYPRH-NH_2 library	9L*	Yes	AG0100t8	3.7	938.1	939.0 ($+H^+$); 961.1 ($+Na^+$); 977.1 ($+K^+$)	5.3
Ac-HAIYPRH-NH_2 library	11L	Yes	AG0100t8	3.2	934.1	935.0 ($+H^+$); 957.1 ($+Na^+$); 973.1 ($+K^+$)	21.4
Ac-HAIYPRH-NH_2 library	11L*	Yes	AG0100t8	3.2	938.1	939.0 ($+H^+$); 961.1 ($+Na^+$); 977.1 ($+K^+$)	26.7
Ac-HAIYPRH-NH_2 library	1D	No	AG0100t8	3.3	934.1	935.0 ($+H^+$); 957.1 ($+Na^+$); 973.1 ($+K^+$)	62.4
Ac-HAIYPRH-NH_2 library	1D*	No	AG0100t8	3.3	938.1	939.0 ($+H^+$); 961.1 ($+Na^+$); 977.1 ($+K^+$)	67.2
Ac-HAIYPRH-NH_2 library	2+D	No	AG0100t8	3.8	1010.1	1011.1 ($+H^+$); 1023.2 ($+Na^+$); 1039.2 ($+K^+$)	52.5
Ac-HAIYPRH-NH_2 library	2+D*	No	AG0100t8	3.8	1014.1	1015.1 ($+H^+$); 1027.2 ($+Na^+$); 1043.2 ($+K^+$)	75.9
Ac-HAIYPRH-NH_2 library	2-D	No	AG0100t8	3.7	1010.1	1011.1 ($+H^+$); 1023.2 ($+Na^+$); 1039.2 ($+K^+$)	69.3
Ac-HAIYPRH-NH_2 library	2-D*	No	AG0100t8	3.7	1014.1	1015.1 ($+H^+$); 1027.2 ($+Na^+$); 1043.2 ($+K^+$)	65.1
Ac-HAIYPRH-NH_2 library	3D	Yes	AG0100t8	3.4	948.1	949.0 ($+H^+$); 971.1 ($+Na^+$); 987.1 ($+K^+$)	10.5
Ac-HAIYPRH-NH_2 library	3D*	Yes	AG0100t8	3.4	952.1	953.0 ($+H^+$); 975.1 ($+Na^+$); 991.1 ($+K^+$)	5.3
Ac-HAIYPRH-NH_2 library	4D	Yes	AG0100t8	3.6	951.4	952.0 ($+H^+$); 974.1 ($+Na^+$); 990.1 ($+K^+$)	30.5
Ac-HAIYPRH-NH_2 library	4D*	Yes	AG0100t8	3.6	955.4	956.0 ($+H^+$); 978.1 ($+Na^+$); 994.1 ($+K^+$)	12.6
Ac-HAIYPRH-NH_2 library	5D	Yes	AG0100t8	3.8	946.1	947.0 ($+H^+$); 969.1 ($+Na^+$); 985.1 ($+K^+$)	13.6
Ac-HAIYPRH-NH_2 library	5D*	Yes	AG0100t8	3.8	950.1	951.0 ($+H^+$); 973.1 ($+Na^+$); 989.1 ($+K^+$)	14.6
Ac-HAIYPRH-NH_2 library	6D	Yes	AG0100t8	4.2	986.1	987.1 ($+H^+$); 1009.2 ($+Na^+$); 1025.2 ($+K^+$)	15.6
Ac-HAIYPRH-NH_2 library	6D*	Yes	AG0100t8	4.2	990.1	991.1 ($+H^+$); 1013.2 ($+Na^+$); 1029.2 ($+K^+$)	16.6
Ac-HAIYPRH-NH_2 library	7D	Yes	AG0100t8	4.2	968.2	969.0 ($+H^+$); 991.1 ($+Na^+$); 1007.1 ($+K^+$)	17.6
Ac-HAIYPRH-NH_2 library	7D*	Yes	AG0100t8	4.2	972.2	963.0 ($+H^+$); 995.1 ($+Na^+$); 1011.1 ($+K^+$)	18.6
Ac-HAIYPRH-NH_2 library	8D	Yes	AG0100t8	4.1	951.1	952.0 ($+H^+$); 974.1 ($+Na^+$); 990.1 ($+K^+$)	19.6
Ac-HAIYPRH-NH_2 library	8D*	Yes	AG0100t8	4.1	955.1	956.0 ($+H^+$); 978.1 ($+Na^+$); 994.1 ($+K^+$)	20.6
Ac-HAIYPRH-NH_2 library	9D	Yes	AG0100t8	3.9	934.1	935.0 ($+H^+$); 957.1 ($+Na^+$); 973.1 ($+K^+$)	21.6
Ac-HAIYPRH-NH_2 library	9D*	Yes	AG0100t8	3.9	938.1	939.0 ($+H^+$); 961.1 ($+Na^+$); 977.1 ($+K^+$)	22.6
Ac-HAIYPRH-NH_2 library	11D	Yes	AG0100t8	3.3	934.1	935.0 ($+H^+$); 957.1 ($+Na^+$); 973.1 ($+K^+$)	22.5
Ac-HAIYPRH-NH_2 library	11D*	Yes	AG0100t8	3.3	938.1	939.0 ($+H^+$); 961.1 ($+Na^+$); 977.1 ($+K^+$)	37.3
Phage display peptides	cyclic Ac-CFLFC-NH_2	Yes	AG0100t8	6.6	592.6	593.1 ($+H^+$)	45.3
Phage display peptides	cyclic Ac-CRWEC-NH_2	Yes	AG0100t8	4.1	734.8	735.2 ($+H^+$)	39.6
Phage display peptides	cyclic Ac-CNSQC-NH_2	Yes	AG030t8	5.2	770.8	771.3 ($+H^+$)	51.3

* stands for $CD_3{}^{13}CO$- acetylated peptides.

Once chain elongation is finished, resin was dried by suction and transferred to a 50 mL centrifuge tube. Freshly prepared cleavage cocktail solution (3-5 mL) was added to the resin and reaction performed at room temperature for 1 hour with gentle agitation. Cleavage reactions of peptide sequences containing Arg(Pbf) residues were allowed to react for an extra hour for each Arg residue. Cleaved peptide was precipitated through addition of cold *tert*-butyl methyl ether and centrifuged at 4ºC and 4000 rpm for 10 minutes. The ether fraction was discarded and precipitation repeated up to 3 times to remove all scavengers and by-products from cleavage reaction. Finally, cleaved peptide was dried with a N_2 flow, resuspended in H_2O/MeCN (1:1), separated from the resin through filtration and lyophilized.

1.3.10. Peptide purification

For peptide crudes with purities lower than 95%, peptides were purified by means of semi-preparative HPLC. Samples were dissolved (preferably in low-organic fraction mixtures), filtered through 0.45 µm nylon filter and injected into the semi-preparative HPLC (10 mL loading loop). 30 minutes linear gradients of MeCN in H_2O and 15 mL/min flow rates with λ = 220 nm for detection was the typical set up. Peaks of interest were analysed by analytical HPLC (and MALDI-TOF if required), combined and lyophilized.

1.3.11. Peptide characterization

Identity of peptides synthesized was assessed by MALDI-TOF MS and HPLC-MS. Purity was checked by analytical HPLC using a Sunfire C_{18} column.

1.3.12. Library synthesis

Library synthesis was performed by mix and split methodology in SPPS. Seven small reactors consisting on 20 mL polypropylene syringes provided with a polyethylene porous disk were used for coupling reactions and Fmoc group removal. A 60 mL polypropylene syringe provided with a polyethylene porous disk was used as a container for mixing all resin for further split. Coupling reactions were performed on small reactors with distinct amino acid on each. For D-Ala, D-Arg, D-Ile, D-Glu, D-Ser and D-Trp reactors, protocol A was followed, whereas in D-Pro reactor protocol B was applied. A schematic workflow and reaction steps are detailed in table at page 181.

Steps 5-11 were repeated up to 5 times. On the fifth synthetic cycle step 11 (mix) is skipped and coupling of Fmoc-D-Arg(Pbf)-OH on all reactors and Fmoc group removal is performed. Subsequently, mix of all resin is performed on the 60 mL reactor. Resin is split in two 20 mL reactors to perform acetylation with CH_3COOH and $CD_3{}^{13}COOD$ respectively. Finally cleavage of peptide library from the resin is performed with a cocktail cleavage of TFA, TIS, H_20 (95:2.5:2.5). Cleavage was performed in a 20 mL polypropylene syringe provided with a polyethylene porous disk. Three treatments of 1 hour were performed and cleaved peptide and cocktail cleavage were filtered by suction. Cleaved peptide library was precipitated through addition of cold *tert*-butyl methyl ether and centrifuged at 4ºC and 4000 rpm for 10 minutes. The ether fraction was discarded and precipitation repeated up to 3 times to remove all scavengers and by-products from cleavage reaction. Finally, cleaved peptide was dried with a N_2 flow, resuspended in H_2O/MeCN (1:1), separated from the resin through filtration and lyophilized.

Step	Step	Reagent/Solvent				Time
1	Resin conditioning	DCM				5 x 1 min
2	Resin conditioning	DMF				5 x 1 min
3	Preactivation	20% piperidine in DMF				1 x 1 min, 2 x 10 min
4	Wash	DMF				5 x 1 min
5		Split (x7)				
		Protocol A		**Protocol B**		
		Reagent/Solvent	**Time**	**Reagent/Solvent**	**Time**	
6	Coupling	Fmoc-D-Aa-OH (4 eq), Oxima pure® (4 eq), COMU (4 eq), DIEA (8 eq) in DMF	2 x 90 min	Fmoc-D-Pro-OH (4 eq), Oxima pure® (4 eq), COMU (4 eq), DIEA (8 eq) in DMF	2 x 90 min	
7	Wash	DMF				
8	Fmoc group removal	20% piperidine in DMF	1 x 1 min, 2 x 10 min	20% piperidine in DMF	1 x 1 min, 2 x 10 min	
9	Fmoc group removal	-	-	DBU, toluene, piperidine in DMF (5:5:20:70)	1 x 5 min	
10	Wash	DMF				5 x 1 min
12	*Coupling*	*Fmoc-D-Arg(Pbf)-OH (4 eq), Oxima pure® (4 eq), COMU (4 eq), DIEA (8 eq) in DMF*	*1 x 90 min*	*Fmoc-D-Arg(Pbf)-OH (4 eq), Oxima pure® (4 eq), COMU (4 eq), DIEA (8 eq) in DMF*	*2 x 90 min*	
13	*Fmoc group removal*	*20% piperidine in DMF*	*1 x 1 min, 2 x 10 min*	*20% piperidine in DMF*	*1 x 1 min, 2 x 10 min*	
11		Mix				
14		Split (x2)				
15	Wash	DMF				5 x 1 min
16	Wash	DCM				5 x 1 min
17	Acetylation	CH_3COOH (6 eq), DIPCDI (3 eq) in DCM	2 x 15 min	$CD_3{}^{13}COOD$ (6 eq), DIPCDI (3 eq) in DCM	2 x 15 min	
18	Cleavage	TFA, TIS, H_20 (95:2.5:2.5)	3 x 60 min			

1.3.13. Library characterization

Library characterization was performed by 3 methods.

1.3.13.1. Cleavage of single beads

Appropriate library synthesis should lead to one-bead one-compound library. Then, previous to the cleavage step some resin beads were picked up and cleaved in individual vials with a cocktail cleavage composed by TFA, TIS, H_20 (95:2.5:2.5). Cleavage cocktail was dried with N_2 flux and dissolved in H_2O/MeCN (1:1). 1 µL of peptide solution was mixed with 1 µL of ACH matrix in a MALDI-plate and MALDI-TOF spectra performed. Each peptide was analysed by MALDI-TOF MS in tandem mode. Analysis was performed to fifteen beads and single peptides sequences were purely characterized on each bead.

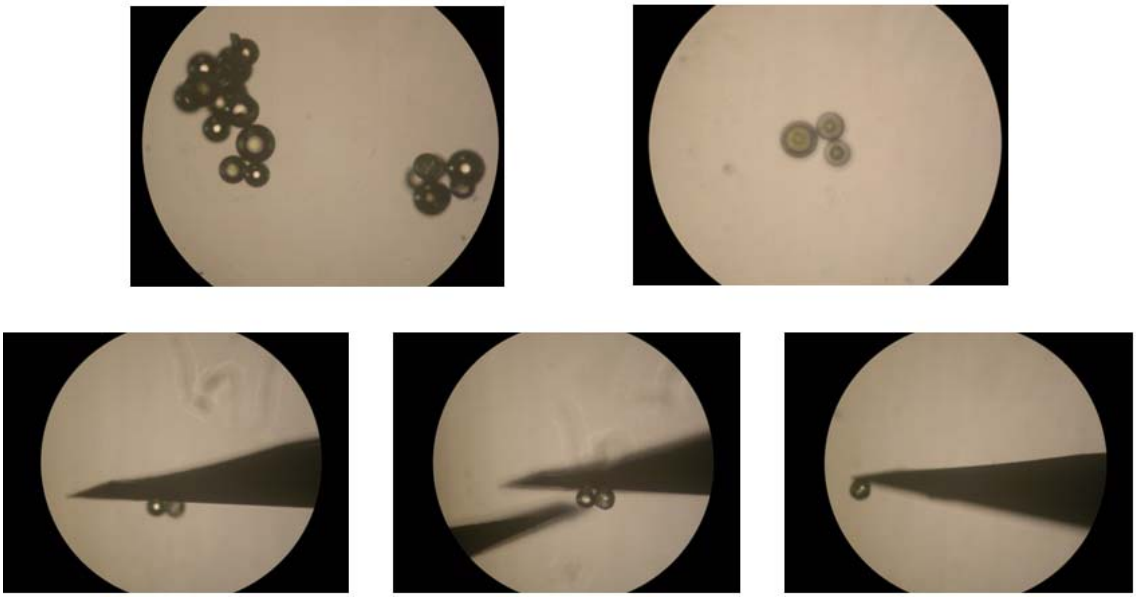


Figure M.1. Picture sequence showing how single resin beads were picked up from library reactor using an optic microscope handling two needles.

1.3.13.2. FT-ICR

500 µg of library sample was reconstituted in 50 µL of H_2O/MeCN (1:1) with 0.1% formic acid. Sample direct infusion (automated nanoelectrospray) was performed on LTQ-FT Ultra (Thermo Scientific). The NanoMate (Advion BioSciences, Ithaca, NY, USA) aspirated the samples from a 384-well plate (protein Lobind) with disposable, conductive pipette tips, and infused the sample through the nanoESI Chip (which consists of 400 nozzles in a 20x20 array) towards the mass spectrometer set to positive ionization. Spray voltage was 1.90 kV and delivery pressure was 0.8 psi. Capillary temperature was 200ºC, capillary voltage 44 V and tube lens 120 V. Screened m/z range was comprised between 200 and 1500 a.m.u. Data was acquired with Xcalibur software, vs.2.0SR2 (ThermoScientific).

1.3.13.3. Analysis of amino acids

A known amount of peptide or peptide library was transferred to a pyrex tube for analysis of amino acids. Samples was hydrolysed with HCL 6 N at 110ºC for 8 hours. The resulting solution was evaporated under vacuum and derivatized according to the derivatization protocol provided within the derivatization kit (AccQ-Tag™, Waters). Amino acid concentration was determined by HPLC compared with an internal standard.

1.3.14. Library cleavage

Peptide library cleavage from the resin and deprotection of side chains was performed by acidolytic treatment with TFA solution, since SPPS was carried out by Fmoc/*t*Bu chemistry. Once chain elongation was finished, resin was dried by suction and freshly prepared cleavage cocktail solution (3-5 mL; TFA/TIS/H_2O, 95:2.5:2.5) was added to the resin. Reaction was performed at room temperature for 2 hour with gentle agitation. Since peptide sequences contained one or more Arg(Pbf) residues, the reaction was performed twice. Cleaved peptide was precipitated through addition of cold *tert*-butyl methyl ether and centrifuged at 4ºC and 4000 rpm for 10 minutes. The ether fraction was discarded and precipitation repeated up to 3 times to remove all scavengers and by-products from cleavage reaction. Finally, cleaved peptide was dried with a N_2 flow, resuspended in H_2O/MeCN (1:1), separated from the resin through filtration and lyophilized. Lyophilised library was redissolved with 1M HCl H_2O/MeCN (1:1) and lyophilized three times to replace trifluoroacetate counterion for a chloride ion.

2. Cellular biology

2.1. Reagents

All the reagents employed in the present section have been purchased to following suppliers: Biological industries, Calbiochem, Gibco, Lonza, Panreac, Sigma-Aldrich and Worthington. Bovine brain microvascular endothelial cells were purchased at Cell Applications and Wistar rats at Charles River.

2.2. Instrumentation

2.2.1. Transendothelial electrical resistance measurement

Transendothelial electrical resistance measurement (TEER) was performed by using an ohmmeter Millicell ERS system (MERS 000 01, Millipore). TEER ($\Omega \cdot cm^2$) was calculated from obtained electrical resistance by substraction of the electrical resistance of collagen-fibronectin coated filter with no cells and correction for filter surface area.

$$TEER = \left(Resistance_{filter\ with\ cells} + Resistance_{filter\ without\ cells}\right) \cdot A$$

where *A* is the area of the filter in cm^2.

2.2.2. Fluorescence measurement (LY)

Fluorescence of Lucifer Yellow (LY) was determined by using an FL600 Microplate Fluroescence Reader, BIO-TEK.

2.3. Primary cell lines: General considerations

Coating of culture flasks surfaces is required for proper cell seeding: poly-D-lysine coating for astrocytes and collagen and fibronectin coatings for bovine brain endothelial cells.

Poly-D-lysine: cell culture flasks were treated with 6 mL of poly-D-lysine (10 µg/mL in H_2O) for 2 hours at 37ºC. Poly-D-lysine solution was eliminated by suction and flasks were left on the hood until completely dry (30 minutes).

Collagen and fibronectin coating: cell culture flasks were coated by adding 6 mL of collagen (10 µg/mL in H_2O, 0.1% v/v acetic acid) for 2 hours at room temperature. Collagen coating solution was eliminated by suction and flasks washed three times with PBS. Subsequently, 5 ml of fibronectin coating solution was added (10 µg/mL in H_2O) for 30 minutes. Fibronectin coating solution was aspirated and used for cell culturing straightaway.

DMEM (+/+) was used as the standard media for cells maintenance, both rat astrocytes and bovine brain endothelial cells. Culture medium was changed every other day. DMEM (4500 mg/L glucose) was enriched depending on the final usage.

DMEM (-)	2 mM glutamine 50 U/mL of penicillin 0.05 g/mL streptomycin	Rat dissection
DMEM (+)	10% heat inactivated calf serum 2 mM glutamine 50 U/mL of penicillin 0.05 g/mL streptomycin	Astrocytes conditioning
DMEM (+/+)	10% heat inactivated calf serum 2 mM glutamine 25 mM HEPES 5 mL of MEM non-essencial amino acids 50 U/mL of penicillin 0.05 g/mL streptomycin	Cell culture and maintenance

Trypsinization of cells was performed with 0.25% trypsin-EDTA for 1 minute at 37ºC. Cells were washed down with DMEM (+/+) and centrifuged at 1000 rpm for 1 minute to remove all trypsin. If cells were not detached with trypsin treatment a scraper was used. Cell pellet was resuspended in DMEM (+/+) and cell density was determined in a Neubauer counting plate.

2.4. Astrocytes isolation

All process was performed with strictly sterilized instruments. Solutions 1 to 5 were freshly prepared the same day of dissection.

Solution 1	50 mL Krebs-Ringer buffer 1X 0.15 g BSA 0.4 mL $MgSO_4$ at 3.8% (w/v)
Solution 2	10 mL solution 1 2.5 mg trypsin
Solution 3	10 mL solution 1 0.8 mg DNAse 5.2 mg STBI 0.1 mL $MgSO_4$ at 3.8% (w/v)
Solution 4	8.4 mL solution 1 1.6 mL solution 3
Solution 5	5 mL solution 1 40 µL $MgSO_4$ at 3.8% (w/v) 6 µL $CaCl_2$ at 1.2% (w/v)

Six newborn Wistar rats (2 pups/flask) at the age of 2 to 3 days were decapitated with scissors and each head was dipped in 70% ethanol. The brains were extracted from the skulls and placed on a petri dish with ice-cold DMEM (-). Cortex of each brain was carefully isolated and meninges were peeled out. Cortex were pulled, chopped with a blade and transferred with a pipette to a

50 mL centrifugation tube together with solution 1. Mixture was centrifuged at 4ºC and 1000 rpm for 5 seconds. Supernatant was discarded and solution 2 was added. Mixture was carefully shaken at room temperature for strictly 5 minutes, when solution 4 was added and mixture centrifuged at 4ºC and 1000 rpm for 5 seconds. Supernatant was discarded and two Pasteur pipette volumes of solution 3 were added. The pellet was resuspended and thoroughly homogenized using a Pasteur pipette with a narrowed end. Solution 5 was added to the homogenized mixture. Mixture was centrifuged at 4ºC and 1000 rpm for 5 minutes. Supernatant was discarded and pellet was carefully resuspended in DMEM (+) and seeded in three p75 cell culture flasks. Cell cultures were maintained at 37ºC and 5% CO_2 for 3 days (it is highly recommended not to manipulate flasks during the first three days), when culture media had to be replaced. At day 6, culture media was renewed and cells culture were shaken at 90 rpm under anaerobic conditions overnight. Culture media was removed and cells still attached to the flasks were carefully rinsed with PBS and cultured again with DMEM (+) until day 9 at 37ºC and 5% CO_2. Astrocytes were subcultured in p75 culture flasks previously coated with poly-D-lysine.

All experiments using rats were approved by the appropriate institutional review committee and performed in strict compliance with the European Community Guide for the Care and Use of Laboratory Animals.

2.5. Cell-based BBB *in vitro* model

The established model consists on a co-culture of bovine brain endothelial cells and rat astrocytes. Model was build on 24-well plates containing inserts or transwells containing a polycarbonate membrane with a surface are of 0.33 cm^2 and pore-size of 0.4 µm. Upper surface of transwells was coated with collagen type IV and fibronectin. First, filter membranes were incubated with 100 µL of collagen type IV solution (10 µg/mL) in H_2O (0.1% v/v acetic acid) for 2 hours at room temperature. Transwells were rinsed three times with PBS to completely remove acetic acid traces and 100 µL of fibronectin solution (100 µg/mL in PBS) were added for transwell membrane coating for 30 minutes at room temperature, after which solution was removed by suction. Transwells were located upside down in plastic boxes containing 10 mL of PBS to keep high relative humidity inside the box. Immediately, 90 µL dropplets approximately containing 45000 astrocytes were seeded on the bottom part of each filter. Astrocytes were incubated in upside down position for 75 minutes. Fresh DMEM was added on each filter every 15 minutes. Inserts were carefully transferred to the 24 well plate containing 800 µL of DMEM media. Upper well was filled with 200 µL of DMEM. The transwell system was incubated at 37ºC and 5% CO_2 for 3 days. It is highly recommended not to manipulate transwell system during the first three days. Media was changed by enriched differentiation medium (clonetics EMVB SingleQuots, Lonza) with heparin (125 µg/mL in PBS, freshly prepared), HEPES (25 mM in PBS) and MEM (10 µL/mL) two hours prior bovine brain endothelial cells seeding. Bovine brain endothelial cells were seeded in the upper part of the filter by adding 20 µL containing approximately 45000 cells. Cells are homogeneously distributed all across the membrane by carefully pipetting up and down the upper trasnwell suspension. The transwell system was incubated at 37ºC and 5% CO_2 for three days. At day 3 of co-culture, medium was replaced by differentiation medium (clonetics EMVB SingleQuots, Lonza) supplemented with 8-(4-chlorophenylthio)-cAMP (154 mg/L in H_2O) and 4-(3-butoxy-4-methoxybenzyl)-2-imidazolidnone (4.9 mg/L in DMSO) and incubated at 37ºC and 5% of CO_2. Usually at day 8 of co-culture the system was ready to perform transport studies. To assess maturity of the cell layer transendothelial electrical resistance measurements (TEER) were performed.

2.6. Transport assays

2.6.1. Blood-brain barrier *in vitro* model

Inserts were washed with Ringer/HEPES solution. Acceptor (lower or baso-lateral) compartment were filled with 0.8 mL of Ringer/HEPES solution, while donor (upper or apical) were filled with 0.2 mL of peptides to be tested dissolved in Ringer/HEPES. Assay was performed for 2 hours incubating the plate at 37ºC and 5% CO_2. Assayed peptides were co-incubated with Lucifer Yellow (LY) at 20 µM. After the experiment, LY permeability is determined by transferring an aliquot of the samples in a fluorescence plate reader (excitation at 428 nm and emission at 536 nm). Higher than $1.7 \cdot 10^{-5}$ cm/s LY permeability indicates cell membrane disruption or breakage during the assay, and such wells discarded.

P_{app} was determined by:

$$P_{app} = \frac{\partial Q}{\partial t} \cdot \frac{1}{A \cdot C_0}$$

where (d*Q*/d*t*) is the amount of sample present in the acceptor compartment in function of time (nmol/s), *A* is the area of the insert (cm^2) and C_0 is the initial concentration of sample applied to the donor compartment (nmol/mL).

The percentage of transport is calculated following equation:

$$\%T = \frac{Q_A(t)}{Q_D(t_0)} \cdot 100$$

were $Q_A(t)$ is the compound concentration at the acceptor well at time *t* and $Q_D(t_0)$ is the compound concentration in the donor well at *t = 0*.

All samples were tested in triplicate. Peptides from the HAIYPRH library were tested at 200 µM. Ac-*D*-Arg-library was tested at 10% of saturation concentration.

2.6.2. Parallel artificial membrane permeability assay

Parallel artificial membrane permeability assay (PAMPA) was used to assess the ability of tested peptides to cross the BBB by passive diffusion by means of effective permeability (P_e). Initial concentration of compounds was set at 200 µM in a buffer solution prepared by dilution from a commercial buffer solution by pION following manufacturer's instructions. PAMPA sandwich consist on two 200 µM wells separated by a polycarbonate filter membrane. 4 µL of phospholipids in dodecane (20 mg/mL) coated polycarbonate membrane. Peptide sample (195 µL) was placed in donor compartment (lower) containing a magnetic seed, while acceptor compartment (upper) was filled with 200 µL of plain buffer. The plate was covered and incubated at room temperature in a saturated humidity atmosphere for 4 hours in Gut-Box™ chamber at 25 µm of unstirred water layer (UWL). Content of acceptor and donor compartments were evaluated by HPLC-UV or by MALDI-TOF, and P_e, percentage of transport (*%T*) and membrane retention (% membrane retention) were calculated.

$$P_e = \frac{-218.3}{t} \cdot log\left[1 - \frac{2 \cdot Q_A(t)}{Q_D(t_0)}\right] 10^{-6} \ cm/s$$

$$\%T = \frac{Q_A(t)}{Q_D(t_0)} \cdot 100$$

$$\%\ membrane\ retention = [Q_D(t_0) - Q_A(t) - Q_D(t)] \cdot 100$$

2.6.3. *In vivo* experiments

To perform *in vivo* experiments, BALB/c mice at the age of 6 weeks were used. 200 µL of Ac-D-Arg-library saturated solution in PBS at pH = 7 was intra-tail vein injected to mice with 2% (w/v) of blue evans. Three distinct circulation times were tested: 30 minutes, 6 hours and 24 hours. Each condition was tested by duplicate. A control mice was injected with 200 µL of PBS. Mice responded well and did not show any adverse effect in none of tested conditions. Blood collections were performed over time: t0, 15 minutes, 30 minutes, 1 hour, 3 hours, 6 hours, 12 hours and 24 hours. Mice were deeply anesthesized with avertin and were perfused with 10 mL PBS by heart injection. Organs and tissues of interest were harvested, chopped and stored at -80ºC. Organs and tissues collected were: brain, heart, lung, pancreas, spleen, liver, kidney, muscle and fat. In any case blue evans stained brain ensuring BBB integrity during the assay.

3. Mass Spectrometry

3.1. Off-line reverse phase purification protocol

Selected columns (Ultra MicroSpin C_{18}, 300Å silica (5-60 µg) commercial column) were conditioned with 200 µL of conditioning buffer (MeOH) two times, and spin in centrifuge (1000 x g for 1 min) to push the flow through. Column was equilibrated with 200 µL equilibration buffer three times with spinning. Peptide samples were dissolved in H_2O (with 5% formic acid) and loaded to the column. Washing buffer was applied to the column three times (200 µL) and finally peptides were eluted with three treatments of 200 µL of elution buffer. Solvent form eluted peptides was evaporated using a SpeedVac system. Samples were dissolved in 40 µL of 0.1% formic acid for injection.

Buffers:

Conditioning buffer	100% methanol (MeOH)
Equilibration/washing buffer	5% formic acid in H_2O (5% FA)
Elution buffer	5% formic acid in H_2O /MeCN (1:1, v/v), 50%MeCN

3.2. Samples treatment previous HPLC-MS injection

3.2.1. *In vitro* samples

3.2.1.1. PAMPA

Acceptor samples and t_0 samples from PAMPA assay were evaporated using a SpeedVac system. Sample was redissolved in 200 µL of H_2O (with 5% formic acid) and desalted with *off-line reverse phase purification protocol* (*see* 3.1).

3.2.1.2. Cell-based BBB *in vitro* model

Acceptor samples and t_0 samples from cell-based BBB *in vitro* assay were evaporated using a SpeedVac system. Sample was redissolved in 200 µL of H_2O (with 5% formic acid) and desalted with *off-line reverse phase purification protocol* (*see* 3.1).

3.2.2. *In vivo* samples

Brain sample and liver and kidney (taken as controls) samples corresponding to 6 hours circulation time experiments were sonicated in 1 mL of 4% SDS 0.1 M HEPES buffer (pH = 8.5) for 5 minutes with 15 seconds pulse on and 15 seconds pulse off cycle in a probe sonicator to lysate all cells and release all peptides entrapped into tissue and cells.

Samples were cleaned using three different methods to get rid out of as many proteins as possible:

a) Protein precipitation with acetone was performed by adding 5 volumes of acetone to the lysate sample and let proteins precipitate overnight at -20ºC. Samples were centrifuged at 4ºC and 10000 rpm for 30 minutes and supernatant volume reduced to dryness with SpeedVac. Extract was redissolved in 200 µL of H_2O (with 5% formic acid) and desalted in C_{18} columns (*see* 3.1).

b) Protein precipitation with methanol was performed by adding 5 volumes of methanol to the lysate sample and let proteins precipitate overnight at -20ºC. Samples were centrifuged at 4ºC and 10000 rpm for 30 minutes and supernatant volume reduced to dryness with SpeedVac. Extract was redissolved in 200 µL of H_2O (with 5% formic acid) and desalted in C_{18} columns (*see* 3.1).

c) Microfiltration with 10 KDa spinfilters (150 µL of lysated sample per filter) was performed in Microcon-10 centrifugal filters (MRCPRT010, Millipore). Collected filtreted sample was reduced to dryness with SpeedVac. Extract was redissolved in 200 µL of H_2O (with 5% formic acid) and desalted in C_{18} columns (*see* 3.1).

3.3. LTQ-Orbitrap-XL

Library peptides were analysed using a LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) coupled to an Agilent Technologies 1200 Series (CA, USA). Peptides were loaded onto C_{18} Zorbax precolumn (Agilent Technologies, cat #5065-9913) and were separated by reversed-phase chromatography using a 12-cm column with an inner diameter of 75 µm, packed with 5 µm C_{18} particles (Nikkyo Technos Co., Ltd. Japan). Chromatographic gradients started at 97% buffer A and 3% buffer B with a flow rate of 300 nL/min, and gradually increased to 90% buffer A and 10% buffer B in 1 min, and to 65% buffer A / 35% buffer B in 30 min. After each analysis, precolumn and column were washed for 10 min with 10% buffer A / 90% buffer B. *Buffer A: 0.1% formic acid in H_2O. Buffer B: 0.1% formic acid in MeCN.*

The mass spectrometer was operated in positive ionization mode with nanospray voltage set at 2.5 kV and source temperature at 200 °C. Ultramark 1621 for the FT mass analyser was used for external calibration prior the analyses. Moreover, an internal calibration was also performed using the background polysiloxane ion signal at m/z 445.1200. The instrument was operated in DDA mode and full MS scans with 1 micro scans at resolution of 60.000 were used over a mass range of m/z 350-2000 with detection in the Orbitrap. Auto gain control (AGC) was set to $1 \cdot 10^6$, dynamic exclusion (60 seconds). All data were acquired with Xcalibur software v2.2.

3.4. Q-trap

The peptide mixtures were analysed using a 5500 Q-trap mass spectrometer (AB Sciex, Framingham, MA, USA) coupled to a nanoLC Ultra-1DPlus (AB Sciex (Eksigent)).

Peptides were loaded onto C_{18} Acclaim PepMap precolumn (Thermo Scientific, cat # 164564) and were separated by reverse-phase chromatography using a 12-cm column with an inner diameter of 75 µm, packed with 5 µm C_{18} particles (Nikkyo Technos Co., Ltd. Japan). Chromatographic gradients started at 98% buffer A and 2% buffer B with a flow rate of 300 nL/min for 5 min and

gradually increased to 60% buffer A and 40% buffer B in 35 min. After each analysis, precolumn and column were washed for 10 min with 2% buffer A / 98% buffer B. *Buffer A: 0.1% formic acid in H_2O. Buffer B: 0.1% formic acid in MeCN.*

SRM data was processed using the Skyline software (v 2.1).

4. Phage display

4.1. Reagents

All reagents and material used on that section were purchased at Affymetrix, Applied Biosystem, Beckman, Bertin Technologies, Bio-Rad, Eppendorf, Fisher, Invitrogen, Jackson ImmunoResearch, New England Biolabs, Qiagen, Sigma-Aldrich, Thermo Scientific and Zymo Research Corporation.

Bacterial culture media:

LB	Purchased at Fisher
TB	For 5L: 48 g tryptone, 96 g yeast extract, 16 mL glycerol + supplements: 23.1 g KH_2PO_4/L, 125.48 g K_2HPO_4/L
SOC	Purchased at Invitrogen

4.2. Instrumentation

Ultracentrifuge	Sorvall WX Ultra 80
Centrifuge	Sorvall Lynx4000
Electroporator	BTX-ECM 630
Thermocycler	Eppendorf Mastercycler proS
Fast Real Time q-PCR	7500 Fast Real-Time PCR System, Applied Biosystem
Incubator	Gyromax 747R, Amerex Instruments
Confocal microscope	Zeiss LSM510

4.3 Phage display library synthesis

4.3.1. Phage vector preparation

4.3.1.1. Propagation of fUSE55

fUSE55 (purified via CsCl gradient) was first electroporated on Invitrogen-Electromax DH5α (cat: 11319-019) bacteria. Two 1mm electroporation cuvettes and 1.5 mL micro centrifuge tubs were chilled on ice for 15 minutes. fUSE55 and DH5α-E were thaw on ice. 20 µL of thawed DH5α and 50 ng of fUSE55 were mixed in chilled tube and mixture transferred to a 1 mm electroporation cuvette (BioRad Gene Pulser 0.1cm #1652089). A control with 1 µL of H_2O instead of plasmid was also prepared. Both cuvettes were electroporated (BTX-ECM630: 2kV, 200 Ω, 25 µF). 500 µL of SOC media were quickly added to electroporated mixture and let recuperate for 5 minutes. Then, it was incubated at 37ºC with vigorous shaking at 250 rpm for 1 hour. Finally, 100 µL and 400 µL were plated in tetracycline agar plates and let colonies grow overnight. A glycerol stock was made with a single well-isolated colony for further usage.

To culture DH5α for propagation of fUSE55, DH5α from glycerol stock was streak on a tetracycline agar plate and incubated at 37ºC overnight. A single, well-isolated colony from the freshly streaked plate is picked and inoculated in 10 mL of LB, 40 µg/µL tetracycline media and incubated for 8 hours at 37ºC and 250 rpm. This starter culture was diluted 1/500 in LB/Tet (40 µg/µL) medium. 10 mL of starter culture were inoculated to 5 L of medium (10 flasks, 2 L capacity, with 500 mL culture on each). It was grown at 37°C overnight (16 hours) with vigorous shaking 250 rpm. The culture should reach a cell density of approximately $3–4 \cdot 10^9$ cells per milliliter, which typically corresponds to a pellet wet weight of approximately 3 g/L medium. Bacterial cells were harvested by centrifugation at 6000 x g for 15 minutes at 4ºC. The bacterial pellet was thoroughly resuspended in 250 mL of lysis buffer from *Inside Qiagen Giga Prep Kit* (Buffer P1) and supplier instructions were followed to lysate cells and collect the plasmid DNA (Inside Qiagen Giga Prep kit, cat#12191).

Two subsequent purification steps were performed with Qiagen Giga Prep kit to purify plasmid DNA. Supplier instructions were followed.

Finally, another purification step, CsCl/EtBr gradient, was performed to further purify DNA. Final volume of DNA was brought up to 4 mL with TE buffer (10 mM Tris, 1mM EDTA, pH = 8; included on *Inside Qiagen Giga Prep Kit)*. Then, DNA in TE was mixed with 4.4 gr of CsCl and 400 µL EtBr (10 mg/mL), covered with aluminium foil and incubated at room temperature for 30 minutes. Simultaneously, a filler mixture was prepared with 24 mL of TE, 25.4 g of CsCl and 2.4 mL of EtBr (10 mg/mL) and also incubated. Using a 20-gauge syringe, carefully and slowly 1/4 of DNA was dispensed into 4 quick-seal tubes *(Beckman cat# 342412)*. Tubes were filled with filler mixture and balanced (± 0.1 g) and sealed (heat seal with metal caps). Balanced tubes were placed in the rotor (VTi65.2) of ultracentrifuge (Sorvall WX Ultra 80). Ultracentrifugation was performed at 58000 rpm for 48 hours at 20 ºC and <250 psi. Tubes were carefully extracted from the rotor and placed in a rack. Two red bands are observed. The top one was chromosomal DNA and the bottom one corresponded to plasmid DNA. A UV source was required to distinguish DNA bands. With a syringe and a 18-gauge needle, desired plasmid DNA band was extracted. Ultracentrifugation was performed twice. To remove CsCl/EtBr, 15 mL of isopropanol was added per mL of DNA solution to precipitate DNA. Solution was vortexed and centrifuged at 3000 rpm for 1 minute. Supernatant was removed and volume brought to the initial volume with TE. Isopropanol precipitation was performed three additional times until the solution was colourless. Finally, 2.5 volumes of TE and 2 volumes of ethanol were mixed with vortex with clean DNA. It was incubated at -20ºC overnight and centrifuged at 15000 xg for 20 minutes, ethanol decanted and DNA pellet resuspended in 500 µL of TE. After 25 L batch, 1.5 mg of clean fUSE55 was recovered.

4.3.1.2. fUSE55 digestion

Digestion of fUSE55 DNA vector was performed with BGLI leading to sticky ends.

10 µg fUSE55 plasmid
5 µL New England Biolabs BGLI Buffer 3 (10x)
1 µL New England Biolabs BGLI 10,000 U/mL
x µL H_2O (up to 50 µL reaction volume)

Reaction mixture was incubated at 37ºC overnight. Three subsequent overnight reactions were performed by addition of fresh BGLI. Enzyme was finally heat inactivated at 65ºC for 20 minutes. After vector digestion a cleaning step was performed with a Zymo column (DNA Clean & Concentrator #D4032, Zymo research corporation).

4.3.2. Olignucleotide library preparation

In annealing step, the single stranded oligonucleotide library CX_3C-oligonucleotide purchased from Sigma interacts to primer library antisense (TTCGGCCCCAAGCGGC).

2 µL CX_3C-oligo (1µg/µL)
4 µL Primer (library antisense: TTCGGCCCCAGCGGC)
2 µL 5x Buffer
2 µL H_2O

To anneal 20 µg of oligonucleotide library, 10 reactions were simultaneously performed in a thermocycler:

Annealing temperature set	
93 ºC	3 min
80 ºC	20 min
75 ºC	20 min
70 ºC	20 min
65 ºC	20 min
40 ºC	60 min

On elongation, the annealed primer became the starting point for the enzyme to create the antiparallel strand of the oligonucleotide library. On each reaction tube we added:

2 µl 10mM dNTP (25% C, G, A &T)
5 µL 0.1M stock DTT
31 µL enzyme dilution buffer
2 µL Sequenace v2.0 DNA polymerase

Final volume of each tube was 50 µL and it was incubated at 37ºC for 1 hour. After elongation, a purification step was performed with *QIAquick Nucleotide Removal Kit* (*Qiagen* #28306). Instructions from the supplier were followed: 8 columns were required for 10 performed reactions. 50 µL of EB (elution buffer) was used to elute purified oligonucleotide sample from each column. The final volume was 400 µL.

Digestion (leading to sticky ends) of double stranded oligonucleotide library was performed with BGLI:

200 µL purified elongated oligonucleotide library
25 µL New England Biolabs BGLI Buffer 3 (10x)
5 µL H_2O
20 µL BGLI (NEB, 10 U/µL)

Two reactions (250 mL each) were performed overnight at 37ºC in an orbital shaker. Three subsequent overnight reactions were performed by addition of fresh BGLI. After digestion a purification step was performed with QIAquick Nucleotide Removal Kit using 6 columns. Elution volume was reduced with SpeedVac to get a concentration solution.

4.3.3. MC-1061 Electrocompetent cells preparation

Stock MC-1061 were streak on a non-antibiotic plate and let grow overnight. A single, well-isolated colony from the freshly streaked plate is picked and inoculated in 50 mL of SOC + 40 µg/µL streptomycin media in a 500 mL flask and incubated at 37ºC and 250 rpm overnight. 1 mL of this culture was diluted into 1 L of LB + 40 µg/µL streptomycin media in 2.5 L flask and let grow at 37ºC and 250 rpm for 2.5-3.5 hours until OD was 0.8 at 550 nm. Once the target OD was achieved the culture was transferred to an ice bath and was centrifuged in centrifugation flasks at 4ºC, 5000 rpm for 15 minutes. Supernatant was discarded and bacteria cells were resuspended in 150 mL ice-cold 1 mM sterile HEPES. Centrifugation was carried out at 4ºC, 5000 rpm for 15 minutes and supernatant removed. Cell pellet were again resuspened in 150 mL of HEPES solution and centrifuged again. Then, clean cells pellet was resuspended in 24 mL of ice-cold 10% (v/v) glycerol wash buffer. Resuspended cells were centrifuged at 4ºC, 5000 rpm for 15 minutes and supernatant removed. Finally, cells pellet was resuspended in 10% (v/v) glycerol solution up to a final volume of 2 mL. Aliquots of this solution were distributed to chilled criovials quickly frozen in dry ice ethanol bath. Criovials were stored at -80ºC.

Cell competency

Dilution of pUC19 plasmid was prepared at 10 pg/µL. A 100 µL aliquot of MC-1061 was thawed and mixed with plasmid 1:10 (2.5 µL of plasmid with 22.5 µL of MC-1061 bacteria). A control was set up by mixing H_2O and bacteria (1:10). Mixture was electroporated (BTX-ECM630: 2kV, 200 Ω, 25 µF) with 1 mm electroporation cuvettes (BioRad Gene Pulser 0.1cm #1652089). Electroporated mixture was quickly resuspended with 1 mL of SOC and incubated at 37ºC and 250 rpm for 30 minutes. Bacteria were plated to no antibiotic plates (duplicates of 10 µL, 100 µL and 1000 µL), and plates incubated overnight at 37ºC.

Efficiency (colonies/µg plasmid) was calculated as follows:

$$\frac{\#\ colonies}{plating\ volume\ (\mu L)} \cdot \frac{1 \cdot 10^4 \mu L}{10\ pg} \cdot \frac{1 \cdot 10^6 pg}{1\ \mu g} =$$

4.3.4. Setting up ligation conditions

Ligation of fUSE55 and oligonucleotide library inserts requires setting up the best ratio vector:insert conditions to perform ligation reaction. Molar ratios 1:1, 1:3, 1:5, 1:10, 1:30 were tested to determine ideal ligation efficiency. Oligonucleotide library was set up at 3.7 µg/mL and fUSE55 and library insert are formed by 9206 bp and 30 bp, respectively. Then, reaction conditions were calculated in 20 µL reaction volumes:

Ratio fUSE55:insert	1:1	1:3	1:5	1:10	1:30	C-1	C-2
fUSE55 (500 ng/µL)/µL	1	1	1	1	1	1	1
Oligonucleotide library/µL	0.43	1.3	2.16	4.34	13	0	0
T4 enzyme (1 U/mL)/µL	2	2	2	2	2	2	0
T4 buffer (5x)/µL	4	4	4	4	4	4	4
H_2O/µL	12.57	11.7	10.84	8.66	0	13	15

Ligation reaction was performed at 16ºC during 12 hours. Enzyme heat inactivation was finally performed at 65ºC for 20 minutes. C-1 and C-2 were control reactions.

Each ligation reaction material was electroporated to MC-1061 electrocompetent bacteria cells. 20 µL of electrocompetent MC-1061 cells was mixed with 1 µL of ligation reaction, including C-1 and C-2. In addition, an additional control (C-3) was performed with 1 µL of H_2O instead of ligation reaction mixture. Samples were incubated on ice for 30 minutes and loaded into chilled 1 mm

electroporation cuvettes (BioRad Gene Pulser 0.1cm #1652089). Cuvettes were electroporated (BTX-ECM630: 2kV, 200 Ω, 25 µF), its content quickly resuspended with 1 mL of SOC media and incubated at 37ºC and 250 rpm for 1 hour.

Incubated bacteria were plated (by duplicate: 10 µL, 100 µL) on tetracycline agar plates. C-3 was platted in both tetracycline plate (1x 100 µL) and non-antibiotic plate (1x 100 µL). Plates were incubated at 37ºC overnight.

Colonies were counted and ligation efficiency was calculated. In addition, 96 colonies were picked for sequence plasmid DNA and determine whether contain or not the ligated insert. Single and well-isolated colonies were picked and each one resuspended in 50 µL of LB media. We performed a PCR (poly-chain reaction) of phage insert in a total volume reaction of 20 µL:

2 µL of bacteria suspension
0.5 µL of 10 mM dNTPs
2 µL of 10x Taq-pol buffer (with 2.5 mM $MgCl_2$)
1 µL of fUSE5 forward primer (8 pmol/µL)
1 µL of fUSE5 reverse primer (8 pmol/µL)
12.7 µL of H_2O
0.4 µL of DMSO
0.4 µL of Taq-pol (5 U/µL)

Then, PCR incubation cycle was performed:

PCR cycle		
1	94 ºC	3 min
2	94 ºC	10 sec
3	60 ºC	30 sec
4	72 ºC	1 min
5	72 ºC	3 min
Repeat steps 2-4, 35 times. Finally hold temperature at 4ºC.		

After PCR cycle a 4% e-gel was run on 10 random samples per 96-plate to check the quality and quantity of the amplified DNA. Amplified DNA was diluted with milli-Q H_2O: 10 µL of H_2O to 1 µL of PCR sample. Samples were sequenced on DNA sequencing core facility at MD Anderson Cancer Center.

4.3.5. Library ligation conditions for complete library synthesis

Ligation of digested vector and oligonucleotide library was done according to results obtained on the ratio tests (*see* 4.3.4) where efficiency was optimized. Ligation was performed in PCR tubes with 500 ng of fUSE55 vector.

Ratio fUSE55:insert	1:5	C-1	C-2
fUSE55 (500 ng/µL)/µL	1	1	1
Oligonucleotide library (3.71 µg/µL)/µL	2.17	0	0
T4 enzyme (1 U/mL)/µL	2	2	0
T4 buffer (5x)/µL	4	4	4
H_2O /µL	10.83	13	15

Ligation reaction was performed at 16ºC during 12 hours. Enzyme heat inactivation was finally performed at 65ºC for 20 minutes. C-1 and C-2 were control reactions.

In contrast to electroporation for setting up ligation conditions, all ligation DNA material was electroporated to synthesize the phage library. DNA resulting from ligation step was electroporated to MC-1061 electrocompetent bacteria cells. 200 µL of electrocompetent MC1061 cells was mixed with 10 µL of ligation reaction. Similarly, 10 µL of ligation reaction controls C-1 and C-2 were electroporated with 200 µL of electrocompetent MC1061 cells. C-3 was also performed by electroporating 200 µL of electrocompetent MC-1061 cells with 10 µL of H_2O. Samples were incubated on ice for 30 minutes and load in chilled 2 mm cuvettes. Cuvettes were electroporated (2kV, 200 Ω, 25 µF), its content quickly resuspended with 20 mL of SOC media and incubated at 37ºC and 250 rpm for 1 hour.

On one hand, incubated bacteria were plated (by duplicate: 10 µL, 100 µL) on tetracycline agar plates. C-3 was platted in both tetracycline plate (1x 100 µL) and non-antibiotic plate (1x 100 µL). Plates were incubated at 37ºC overnight and colonies counted to determine efficiency of the process that resulted as $4.6 \cdot 10^6$ transformats/µL DNA.

On the other hand, 830 mL of LB media was supplied to 20 mL SOC solution containing electroporated bacteria. Altogether were transferred to a 2.5 L flasks and library grown at 37ºC, and 250 rpm for 20 hours in a previously disinfected incubator.

The volume of each flask was equally divided on 3 centrifugation flasks, which were centrifuged at 4ºC, 6000 rpm for 7 minutes. Supernatant containing the first generation of CX_3C phage library was decanted in 1L sterile bottles. 25.5 g of NaCl (30 g/L) and 34 g of PEG (40 g/L) were added and thoroughly shake for phage precipitation overnight in an ice bath (first precipitation).

Simultaneously, bacteria pellet was saved for future growing of a second generation of CX_3C phage library. To that purpose, bacteria pellet was resuspended in 80 mL of 10% glycerol for washing and centrifuged at 4ºC, 5000 rpm for 7 minutes. Finally, obtained pellet was resuspended in 1 mL of 50% glycerol, frozen on criovials in dry ice and stored at -80ºC.

Precipitated phage was centrifuged at 4ºC, 10000 rpm for 30 minutes and supernatant was carefully removed. Phage-containing pellet, was resuspended in 40 mL PBS by smooth shaking at 37ºC for 1 hour and at 4ºC overnight. Resuspended phage were centrifuged for debris removal at 4ºC, 10000 rpms for 30 minutes.

Second phage precipitation was performed by adding 15% volume (6 mL) of PEG/NaCl solution (NaCl, 30 g/L and PEG, 40 g/L) for 45 minutes in an ice bath. Precipitated phage was centrifuged for 4ºC, 10000 rpm for 30 minutes and supernatant carefully removed. Phage pelled was redissolved in 400 µL of PBS overnight at 4ºC. Finally, it was transferred to a 1.5 mL centrifugation tube and centrifuged for debris removal at 4ºC, 10000 rpm for 30 minutes. Supernatant containing phage library was stored at 4ºC.

4.3.6. Library titration

Few colonies of *E. coli* K91/kan bacteria were inoculated to 10 mL of TB medium containing kanamycin (0.1 mg/mL) and grown at 37ºC, 250 rpm until OD density is between 0.190 and 0.200 at 600 nm when diluted 1:10 in TB.

Dilutions of phage library down to 10^{-7}, 10^{-8} and 10^{-9} are required for feasible counting.

200 µL of log phase *E. coli* K91/kan bacteria was added to tubs containing dilutions 10^{-7}, 10^{-8} and 10^{-9} of phage to be tittered in PBS. Incubation was allowed for 30 minutes at room temperature with no shaking. 100 µL of each phage-infected bacteria solution were plated by duplicate on Kan/Tet agar plates and let bacteria colonies grow overnight at 37ºC.

Colonies were counted in those plates containing between 20 and 600 colonies (out of these limits results are not reliable) and titration determined depending on dilution of counted plate. Library resulted on $2.4 \cdot 10^9$ TU/µL.

4.4. Panning

4.4.1. Panning protocol

4.4.1.1. First panning round

Three BALB/c mice were injected with $5 \cdot 10^9$ TU of CX_3C phage library in 200 µL of PBS. Circulation time was set for 6 hours. Perfusion of mice was performed with 10 mL of PBS. Brain, pancreas, kidney and liver were harvested and kept on ice. Organs or pieces of organ were weighted and chopped. Weighted organs were grinded with hard tissue grinding mix, Precellys 24 (Bertin Technologies) with 1 mL of cold 1% BSA, Sigma PIC (protease inhibitors cocktail tablets from Sigma #S8830) in DMEM. 150 µL was frozen and stored at -80ºC. Remaining mixture was centrifuged at 7000 rpm for 4 minutes and supernatant was removed. Pellet was washed with 0.5 mL of 1% BSA, Sigma PIC in DMEM and centrifugation repeated, except for brain samples. 700 µL of log phase *E. coli* K91/kan bacteria were added to homogenized tissue and let 30 minutes at room temperature for infection. Mixtures were diluted with 5 mL of LB (kanamycine, Kan, 50 µg/mL; tetracycline, Tet, 40 µg/mL) and 100 µL, 10 µL, 1 µL and 0.1 µL were plated in Kan/Tet agar plates and incubated overnight. 5 additional mL of LB media were added to the diluted mixture and was incubated for phage amplification at 37ºC, 250 rpm overnight. Bacteria was precipitated by centrifugation (4ºC, 8000 rpm for 10 minutes) and phage in the supernatant was precipitated with 15% PEG/NaCl (NaCl, 30 g/L and PEG, 40 g/L) for 3 hours in an ice bath. Precipitated phage were centrifuged at 4ºC, 10000 rpm for 30 minutes. Supernatant was discarded, phage pellet resuspended in 200 µL of PBS and centrifuged at 4ºC, 10000 rpm for 10 minutes in order to get rid out of debris. Phage was tittered: $7.4 \cdot 10^7$ TU/µL.

Colonies from plated homogenate solution were counted and 96 colonies were picked and sequenced after PCR DNA amplification, (*see* 4.3.4).

4.4.1.2. Second panning round

The same protocol than for first panning round was performed. Instead of injecting CX_3C library, in this case, amplified phage from first round of brain affinity selection was injected.

4.4.1.3. Third panning round

The same protocol than for first and second panning rounds was performed. Instead of injecting CX_3C library, amplified phage from second round of brain affinity selection was injected.

	Titration (TU/µL)	Phage volume/µL	Mouse injection/µL
Initial CX_3C library	$2.40 \cdot 10^9$	200	2.1
After 1st selection round	$7.40 \cdot 10^7$	200	67.5
After 2nd selection round	$2.12 \cdot 10^8$	200	23.6

4.4.2. Post-panning round

Each phage from sequence selection in panning rounds was grown separately. Sequences selected were: FLF, GGG, GRV, GSK, GWR, KGK, NSQ, NSK, RGG, RWE, VFF, WGG. To amplify each phage the same protocol was run in parallel for each phage sequence.

4.4.2.1. Individual phage amplification

300 µL of log phase *E. coli* K91/kan bacteria was injected by 4 µL of phage (picked colony stored a 4ºC in 96 well plate with LB) for 30 minutes. Infected bacteria were diluted with 36 mL of LB (Kan, 50 µg/mL; Tet, 40 µg/mL) and grown in three centrifugation tubes (50 mL), 12 mL on each at 37 ºC, 250 rpm overnight. Bacteria culture was centrifuged at 4ºC, 8000 rpm for 10 minutes and supernatant transferred to a clean tube. Precipitation of phage contained in the supernatant was performed by addition of 15% of PEG/NaCl (5.4 mL; NaCl, 30 g/L and PEG, 40 g/L) for 3 hours in an ice-bath. Precipitated phage were centrifuged at 4ºC, 10000 rpm for 30 minutes and pellet resuspended in 200 µL of PBS. To remove debris, resuspended phage were centrifuged at 4ºC, 10000 rpm for 5 minutes and supernatant stored at 4ºC.

4.4.2.2. Individual phage titration

Few colonies *E. coli* K91/kan bacteria were inoculated to 10 mL of TB medium containing kanamycin (0.1 mg/mL) and grown at 37ºC, 250 rpm until OD density was between 0.190 and 0.200 at 600 nm when diluted 1:10 in TB.

Dilutions of individual phage to be tittered down to 10^{-7}, 10^{-8} and 10^{-9} are required for feasible counting.

200 µL of log phase *E. coli* K91/kan bacteria was added to tubs containing dilutions 10^{-7}, 10^{-8} and 10^{-9} of phage to be tittered in PBS. Incubation was allowed for 30 minutes at room temperature with no shaking. 100 µL of each phage-infected bacteria solution were plated by duplicate on Kan/Tet agar plates and let bacteria colonies grow overnight at 37ºC.

Colonies were counted in those plates containing between 20 and 600 colonies (out of these limits results are not reliable) and titration determined depending on dilution of counted plate.

Phage inserted sequence	Titration (TU/µL)
FLF	$4.25 \cdot 10^8$
GGG	$7.90 \cdot 10^8$
GRV	$1.25 \cdot 10^9$
GSK	$1.55 \cdot 10^9$
GWR	$9.50 \cdot 10^8$
KGK	$2.37 \cdot 10^9$
NSQ	$9.70 \cdot 10^8$
NSK	$1.54 \cdot 10^9$
RGG	$1.19 \cdot 10^9$
RWE	$9.10 \cdot 10^8$
VFF	$1.26 \cdot 10^9$
WGG	$5.85 \cdot 10^8$

4.4.2.3. Post-panning injection round

Two BALB/c mice were injected with a mixture containing $5 \cdot 10^9$ TU of each selected phage from the panning rounds in 200 µL of PBS. Circulation time was set for 6 hours. Perfusion of mice was performed with 10 mL of PBS. Brain, pancreas, kidney and liver were harvested and kept on ice. Organs or pieces of organ were weighted and chopped. Weighted organs were grinded with hard tissue grinding mix, Precellys 24 (Bertin Technologies) with 1 mL of cold 1% BSA, Sigma PIC in DMEM. 150 µL was frozen and stored at -80ºC. Remaining mixture was centrifuged at 7000 rpm for 4 minutes and supernatant was removed. Pellet was washed with 0.5 mL of 1% BSA, Sigma PIC in DMEM and centrifugation repeated, except for brain samples. 700 µL of log phase *E. coli* K91/kan bacteria were added to homogenized tissue and let 30 minutes at room temperature for

infection. Mixtures were diluted with 5 mL of LB (Kan, 50 µg/mL; Tet, 40 µg/mL) and 100 µL, 10 µL, 1 µL and 0.1 µL were plated in Kan/Tet agar plates and incubated overnight.

Colonies from plated homogenate solution were counted and 96 colonies were picked and sequenced after PCR DNA amplification.

4.5. Biodistribution experiment: q-PCR

Phage from the library, and the ones selected (CFLFC, CRWEC and CNSQC), are based on fd-Tet phage construct. Since q-PCR experiments require a negative control, fd-Amp phage was used to this purpose because its DNA material do not interact with primers used to amplify fd-Tet phage, and vice versa. Thereby, biodistribution of both phage constructs can be determined simultaneously in the same individual.

q-PRC experiments require titration in terms of amount of phage DNA instead of infective ability of phage as in colony counting titration protocols.

4.5.1. q-PCR titration

For q-PCR titration a calibration curve is required with dilutions of standard fd-Tet and fd-Amp. Standard solutions and samples to be tittered were set up for q-PCR reaction in volumes of 15 µL. Fd-Tet and fd-Amp require distinct primers for reaction.

fd-Tet reaction mixture	fd-Amp reaction mixture
1 µL sample or standard	1 µL sample or standard
1 µL fw/rev primers (2.5 µM each)	1 µL fw/rev primers (2.5 µM each)
7.5 µL SYRB*	7.5 µL SYRB*
5.5 µL H_2O	5.5 µL H_2O

*SYRB: 1x Power SYBR Green PCR Master Mix (Applied Biosystems).

q-PCR cycle		
1	50 ºC	2 min
2	95 ºC	10 min
3	95 ºC	15 sec
4	60 ºC	1 min
Repeat steps 3&4, 40 times. Finally hold temperature at 4ºC.		

4.5.2. Injection of phage + fd-Amp control

Three BALB/c mice were injected with $5 \cdot 10^9$ TU of each selected phage (CFLFC, CRWEC, CNSQC) and with $5 \cdot 10^9$ TU of fd-Amp as a internal control in 200 µL of PBS. In total, 9 mice were used. Circulation time was set for 6 hours. Perfusion of mice was performed with 10 mL of PBS. Brain, lung, liver, spleen, pancreas, kidney, muscle and fat were harvested, weighted and chopped.

4.5.3. q-PCR

DNA extractions were performed with DNeasy (Qiagen) following the supplier instructions. Phage content was determined by quantitative PCR (qPCR). PCR templates consisted of 5 mL of a 1:20 dilution of DNA, 1x Power SYBR Green PCR Master Mix (Applied Biosystems) and 3.75 pmol of each oligonucleotide primer (fUSE5F1: 59-TGAGGTGGTATCGGCAATGA-39 and fUSE5R1: 59-

GGATGCTGTATTTAGGCCGTTT-39) directed to the amplification of a fragment of the Tet^R gene, in a final reaction volume of 15 mL. The temperature program consisted on:

q-PCR cycle		
1	50 ºC	2 min
2	95 ºC	10 min
3	95 ºC	15 sec
4	60 ºC	1 min
Repeat steps 3&4, 40 times. Finally hold temperature at 4ºC.		

Standard curves were generated with serial phage dilutions (from 3 to 3^8 plasmids) for each run. Each point of the curve and each sample DNA were amplified in triplicates. The standard curve was calculated by a linear regression analysis and serial dilutions. Amplification efficiency (AE) of each PCR cycle was calculated fro the slope of the standard curve by the equation: $AE = 10^{1/(-s)}$.

4.6. Immunofluorescence experiment

For immunofluorescence experiment, each selected phage sequence (FLF, RWE, NSQ) was injected to a single mouse: $5 \cdot 10^9$ TU in 200 µL of PBS. After 6 hours circulation time, mice were perfused with 10 mL of PBS and brains were harvested as well as liver, spleen, pancreas and muscle as control organs. Organs and tissues were fixed in 4% paraformaldehyde solution at 4ºC overnight.

4.6.1. Immunostaining

Fixed brains in 4% paraformaldehyde solution were sucrose embedded in increasing gradient of sucrose concentration until they sunk: 10%, 20% and 30% sucrose. Organs were OCT embedded over dry ice to be sliced in a cryostat and 10 µm slices were mounted in slides. All tissue sections were air-dried at room temperature for 30 minutes to dry OCT compound. Tissues were washed three times with PBS and once with PBS containing 0.3% triton X-100 (PBST) for 5 minutes per wash. Tissue sections were circumscribed with a hydrophobic marking pen to avoid tissue to dry out. Tissue sections were incubated in a humidity chamber in 200 µL of 5% normal serum diluted in PBST for 15 minutes at room temperature for blocking. Serum solution was removed by gentle tapping and tissue sections were incubated with the primary antibodies diluted in PBST containing 1% normal serum for one hour. Primary antibodies used were CD31 (for endothelial cells blood vessels) and anti-fd bacteriophage antibody produced in rabbit (for phage tracking). Incubated tissues were rinsed three times with PBST (5 minutes each). Sections were incubated with proper secondary antibodies in PBST for one hour at room temperature (Donkey anti-rat and donkey anti-rabbit IgG Cy3 with 1% donkey serum, Jackson ImmunoResearch). Sections were rinsed three times with PBST (5 minutes each). Tissue sections were fixed with 4% paraformaldehyde in PBS for 2 minutes. Sections were rinsed three times with PBS for 5 minutes each. Finally tissue sections were mounted with Vectashield mounting medium and edges of the coverslip were sealed with Cytoseal mounting media.

4.6.2. Confocal microscopy

Confocal images were acquired on a laser scanning confocal microscope (Zeiss LSM510) equipped with krypton-argon and helium-neon lasers. Image analysis was performed with the Zeiss LSM 3.2 software package.

5. Buffers

Lysis buffer	4% (w/v) SDS 0.1 M HEPES pH = 8.5
Ringer/HEPES	150 mM NaCl 5.2 mM KCl 2.2 nM $CaCl_2$ 0.2 mM $NaHCO_3$ 2.8 mM glucose 5 mM HEPES pH = 7.2
KRB 10X	1.2 mM NaCl 48.3 mM KCl 12.2. mM KH_2PO_4 254.7 mM $NaHCO_3$ 71.3 mM glucose

Bibliography

1. Albericio, F. & Kates, S.A. Solid-Phase Synthesis: A Practical Guide. Taylor & Francis, New York; **2000**. pp. 79-196.
2. Kaiser, E., Colescott, 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.
3. Christensen, T. Qualitative test for monitoring coupling completeness in solid-phase using chloranil. *Acta Chem. Scand. B. Org. Chem. Biochem.* **1979**, 33, 763-766.
4. Madder, A., Farcy, N., Hosten, N.G.C., De Muynck, H., De Clercq, P.J., Barry, J. & Davis, A.P. A novel sensitive colorimetric assay for visual detection of solid-phase bound amines. *Eur. J. Org. Chem.* **1999**, 1999, 2787-2791.

Resum en català

Introducció

La quantitat de malalties que afecten el cervell i el sistema nerviós central (SNC) és molt ampli. Aquestes malalties tenen un gran impacte en la societat moderna i es calcula que en el futur continuaran augmentant en importància degut a l'increment en l'expectativa de vida de la població mundial. És per això, que s'estan dedicant molts esforços a entendre les disfuncions que provoquen aquestes malalties, així com a trobar-hi solucions. En el tractament de qualsevol malaltia, la focalització en l'administració del fàrmac a la zona desitjada és imprescindible per augmentar-ne l'eficàcia i per evitar efectes secundaris. Tanmateix, en les malalties que afecten el SNC, els fàrmacs han de superar un altre escull, la barrera hematoencefàlica (BH) localitzada als capil·lars sanguinis del cervell.

La barrera hematoencefàlica

La funció principal de la BH és protegir el cervell de possibles agents neurotòxics que circulen per la sang i mantenir l'homeòstasi en el parènquima cerebral. La BH està formada per cèl·lules endotelials que difereixen de les que trobem en altres òrgans i teixits degut a la influència que reben dels astròcits, els perícits o altres cèl·lules glials. Les cèl·lules són polaritzades i formen una barrera molt estanca gràcies a la presència d'unions estretes entre les seves parets cel·lulars. A més, presenten molt poc transport vesicular i una activitat metabòlica molt alta. Aquestes cèl·lules endotelials expressen una gran quantitat de receptors que seran els responsables del transport de moltes de les molècules que requereix el funcionament del cervell. Així com s'hi troben receptors per facilitar el flux d'entrada, també trobem potents receptors que controlen el flux de sortida del cervell per tal de mantenir l'homeòstasi. Així doncs, la BH es presenta com un escull difícil de superar per la majoria dels fàrmacs dissenyats per tractar malalties del SNC o del cervell. De fet, s'ha descrit que el 98% de molècules petites no són capaces de passar la BH, mentre que pràcticament cap molècula gran és capaç de fer-ho. Els mecanismes de transport a través de la BH més habituals es divideixen en dos grans grups. Per una banda el transport passiu, representat per la difusió passiva i el flux paracel·lular. Per altra banda el transport actiu, representat per la transcitosi mediada per receptor, transcitosi mediada per transportador i transcitosi mediada per absorció.

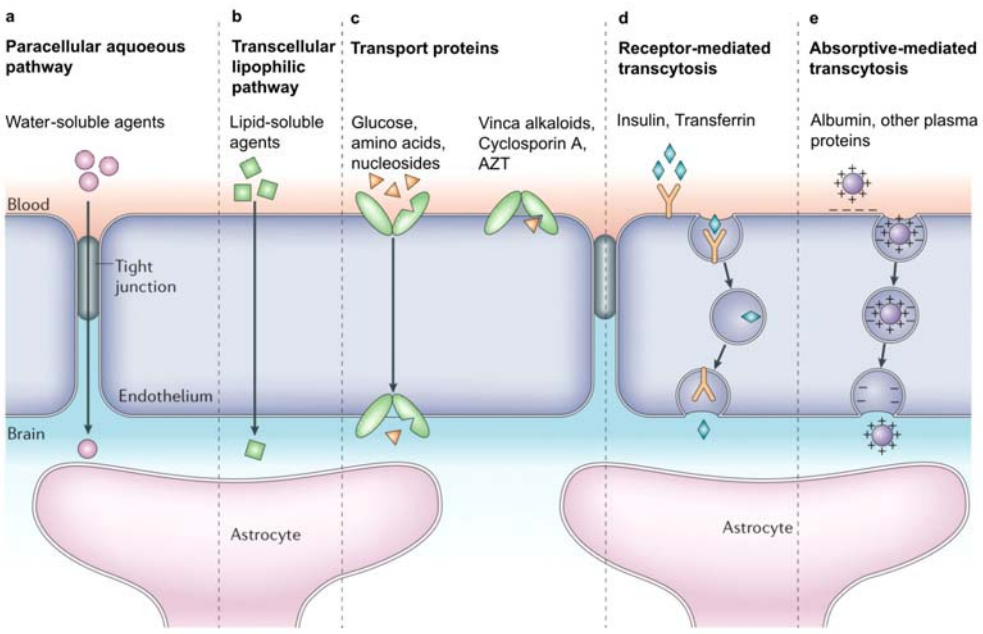


Figura 1. Esquema de la barrera hematoencefàlica i els seus principals mecanismes de transport.

S'han descrit diverses estratègies per travessar la BH, moltes d'elles però suposen o bé la punció i injecció a dintre el cervell o bé la disrupció temporal de la BH, fet que pot provocar greus efectes irreversibles i que descarten la seva aplicació. Així doncs, les estratègies explorades a dia d'avui només contemplen escenaris que no provoquin disrupció de la membrana. Una de les estratègies amb més potencial actualment és la que utilitza llançadores per travessar la barrera a mode de *Cavall de Troia*. Així, una molècula que sigui capaç de travessar la BH (idealment per un mecanisme de transport actiu) ajuda a creuar la barrera a una altra molècula incapaç de fer-ho per si mateix. D'aquesta manera, el sistema llançadora-fàrmac arribaria al lloc d'acció per poder-hi desenvolupar l'efecte desitjat. S'han descrit diverses llançadores o vectors de naturalesa diversa. En aquesta tesi però ens centrem en les llançadores peptídiques.

La modelització de la BH és de vital importància per tal de poder assajar tals llançadores peptídiques així com avaluar la capacitat de transport dels fàrmacs dissenyats. És per això que s'han dedicat i es dediquen molts esforços per mimetitzar les condicions de la BH *in vivo* en assajos *in vitro*. De totes maneres, els assajos *in vitro* encara són lluny de poder modelitzar correctament i completament aquestes propietats. Els models més avançats consten de cocultius cel·lulars de cèl·lules endotelials i astròcits, tot i que s'han descrit assajos amb altres línies cel·lulars. La combinació més estandarditzada fins a dia d'avui és el cocultiu de cèl·lules endotelials de capil·lars de cervell boví i astròcits de rata. Pel que fa als assajos *in vivo* es presenten dificultats importants per l'assaig massiu de compostos així com d'extrapolar els resultats al context humà. Hi ha diferències significatives entre els receptors expressats a la BH en diverses espècies tals com rosegadors o humans. Per això s'estan dedicant molts esforços a la construcció de models *in vitro* amb cèl·lules humanes.

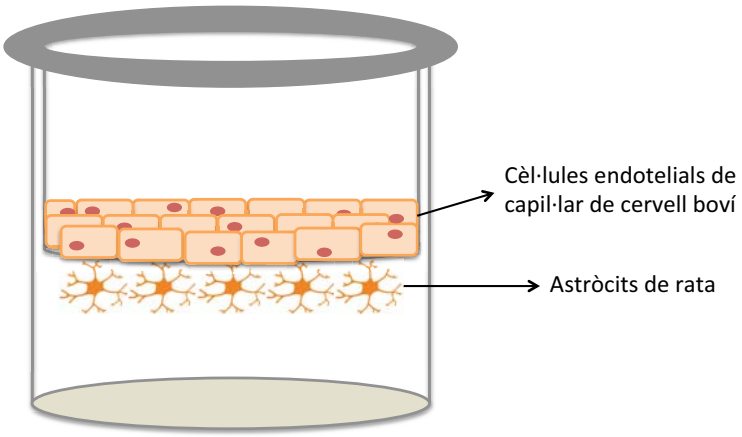


Figura 2. Esquema de model cel·lular *in vitro* de barrera hematoencefàlica. Cocultiu de cèl·lules endotelials de capil·lar de cervell boví i astròcits de rata.

Química combinatòria

El descobriment de fàrmacs és un camp complex, car i que requereix molt temps. Històricament, la major font de fàrmacs han estat els productes naturals. Tanmateix la química combinatòria va aparèixer com una nova font de fàrmacs potencials que va obrir un nou horitzó en el camp, permetent la síntesi de moltíssims compostos en poc temps i a un cost molt menor que fins llavors. Junt a la química combinatòria sovint hi trobem lligades tècniques pel cribratge d'alt rendiment, ja que el gran volum de reactius, productes i operacions han requerit que aquest camp hi evolucioni paral·lelament. Tot i que inicialment la química combinatòria prometia grans expectatives a la industria farmacèutica, al llarg dels anys s'ha comprovat que tot i ser una eina molt potent i útil, la química combinatòria té els seus factors limitants. Així, el disseny de l'estructura bàsica d'una biblioteca de molècules és de vital importància per dirigir la cerca en un espai químic rellevant biològicament. Les possibilitats de trobar molècules d'interès no augmenten de forma lineal amb la quantitat de molècules sintetitzades. D'altra banda, la síntesi de grans quantitats de molècules planteja problemes d'assaig i d'identificació.

Les primeres biblioteques foren sintetitzades en fase sòlida amb mètodes com els anomenats *pins*, bossetes de té o en suports de cel·lulosa. D'aquesta manera era possible obtenir quantitats variables de pèptid (en general quantitats petites) però permetia la possibilitat de sintetitzar-ne molts a la vegada per obtenir-los finalment per separat. Per altra banda, posteriorment es van desenvolupar biblioteques basades en barreges (sobretot també amb síntesi en fase sòlida). Ja fos en barreges de reactius (equimolars o equireactives) o amb el mètode *separa i mescla* que dóna lloc a biblioteques amb un únic tipus de molècula per bola de suport polimèric. En alguns casos, depenent de la naturalesa de l'assaig, la biblioteca s'assaja ancorada a la resina. Tanmateix un cop escindides del suport polimèric aquestes biblioteques resulten amb totes les molècules mesclades. Un altre mètode desenvolupat fou la síntesi paral·lela d'alt rendiment per obtenir petitíssimes quantitats de cada molècula pura, suficients per ser assajades. L'avantatge és que d'aquesta manera evitem tots els problemes derivats de la identificació, però necessitarem una gran capacitat per assajar tots els compostos generats.

Per tal de millorar i facilitar la identificació de compostos de les biblioteques on es troben barrejats, s'han desenvolupat una sèrie d'estratègies en funció de la naturalesa específica de cada biblioteca. Per una banda, l'anàlisi directa és el mètode més senzill i efectiu si la naturalesa de la biblioteca permet aplicar-lo. Per exemple, el cas de l'espectrometria de masses o la degradació d'Edman pel que fa als pèptids o la seqüenciació d'ADN pel que fa als oligonucleòtids. Un altre mètode força emprat és la deconvulació, iterativa o ortogonal, o l'exploració posicional. Finalment, l'estratègia de codificació i descodificació és potser la més emprada en aquests casos ja que permet la identificació d'informació estructural de forma més accessible. Per això, és necessària una estratègia de síntesi on s'hi puguin incorporar etiquetes de naturalesa diferent a la de la biblioteca i que siguin de química compatible per tal d'identificar per altres canals al de la biblioteca principal l'estructura de cada compost analitzat. De totes les modalitats de codificació la més exitosa ha resultat ser la que utilitza oligonucleòtids ja que pot ser fàcilment descodificada mitjançant la seqüenciació d'ADN. La química d'oligonucleòtids permet tractaments ortogonals sense interferència amb altres molècules, com per exemple els pèptids, i permet una gran variabilitat que no limita la variabilitat i l'extensió de la biblioteca sintetitzada.

Espectrometria de masses

En l'última dècada l'espectrometria de masses ha esdevingut finalment el millor aliat de la proteòmica, que d'alguna manera ha retroalimentat l'interès i la importància de les dues disciplines. La identificació i la quantificació de pèptids i proteïnes en mescles complexes i la comparació fiable d'una gran varietat de mostres és imprescindible per postular conclusions o establir models per processos biològics o en el camp de la biologia de sistemes.

Per l'anàlisi de mescles complexes de pèptids, la mostra no s'introdueix directament a l'espectròmetre de masses sinó que s'injecta prèviament a un aparell de cromatografia de líquids acoblat en sèrie. Així els pèptids s'elueixen de forma fraccionada en funció del gradient aplicat per simplificar la mescla que analitza en cada moment l'espectròmetre. S'han desenvolupat una gran varietat de tècniques i aparells d'espectrometria de masses. La tria en la seva utilització serà funció de la complexitat de la mostra i de l'objectiu analític.

La ionització i fragmentació, així com l'analitzador de masses, són elements importants a definir de cara a dissenyar un experiment analític i triar un equip per desenvolupar-ho. Pel que fa a la ionització hi ha dues tècniques que predominen sobre la resta en el si d'aquesta tesi. Per una banda el MALDI (*matrix-assisted laser desorption/ionization*), que consisteix en la ionització d'una mostra determinada per l'aplicació d'un feix làser a una longitud d'ona apropiada. La mostra s'ha de cocristalitzar amb una matriu de naturalesa aromàtica i àcida. En el procés les molècules d'analit s'ionitzen i es dirigeixen mitjançant una sèrie de potencials elèctrics cap a l'analitzador de masses. El MALDI és una tècnica que no permet l'acoblament en sèrie de la tècnica cromatogràfica i l'espectròmetre de masses. D'altra banda, l'ESI (*electrospray ionization*) és una

tècnica que utilitza l'aplicació d'un fort potencial elèctric per vaporitzar la mostra i ionitzar les molècules, que es separen del solvent mitjançant l'aplicació d'un flux de gas inert. La detecció de les espècies ionitzades per part de l'analitzador de masses i la seva detecció genera els espectres de pics moleculars o de primer ordre (MS1). Com que la ionització depèn en gran mesura de la natura de cada molècula l'espectre resultant no té perquè representar acuradament les proporcions de cada analit a la mostra. Per tant, la tècnica no és quantitativa *per se*. Un cop generats els ions moleculars o els ions precursors la fragmentació d'aquests genera informació estructural molt valuosa per la identificació. Hi ha una gran varietat de tècniques de fragmentació que resulten en la generació de diferents patrons de fragmentació. La més utilitzada en proteòmica és l'anomenada CID (*collision-induced dissociation*) que genera principalment fragments peptídics *b* i *y*, que representen el trencament dels enllaços amida de l'esquelet peptídic. Tanmateix, el procés de fragmentació no sempre és predictible i no sempre genera una informació completa de totes les molècules fragmentades.

Els espectròmetres de masses generalment tenen 3 parts fonamentals: la font d'ionització, l'analitzador de masses i el detector. La combinació de les diverses tècniques en cada apartat genera equips amb capacitats específiques. Menció especial mereix l'equip anomenat LTQ-Orbitrap, que al ser un equip híbrid pot operar en paral·lel registrant espectres a la vegada que fragmenta.

La seqüenciació de pèptids és un primer pas per a l'elucidació de les proteïnes d'una mostra. L'aproximació *shotgun proteomics* consisteix en la digestió enzimàtica de les proteïnes en pèptids que s'analitzen, es contrasten i correlacionen amb bases de dades de proteïnes obtingudes a partir del genoma. Una limitació evident d'aquesta aproximació és que només es pot aplicar a organismes dels quals s'hagi seqüenciat el seu genoma.

Per altra banda, una altra tècnica anomenada *targeted proteomics* (o *single reaction monitoring*, SRM) permet la identificació i quantificació precisa de mescles complexes, si bé requereix de certa informació prèvia per tal de poder abraçar grans anàlisis. Aquesta tecnologia permet monitoritzar transicions (combinació d'ió precursor i ió de fragmentació) que permeten determinar seqüències peptídiques. La seva naturalesa no és d'exploració d'alt rendiment i per això permet l'adquisició de dades de gran qualitat amb una sensibilitat fins a un o dos ordres de magnitud superior al d'altres tècniques. Tanmateix, en comparació a *shotgun proteomics*, la SRM està limitada a una selecció molt menor de molècules a analitzar.

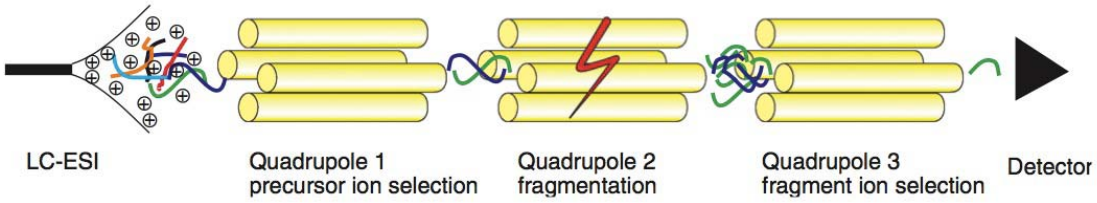


Figura 3. Esquema representatiu d'un triple quadrupol, per desenvolupar la tècnica SRM.

Per altra banda, la característica més interessant de la SRM és que permet la quantificació de molècules determinades. Amb aquesta tècnica i monitoritzant transicions intenses i diferencials entre els pèptids d'una mostra és possible la quantificació, ja sigui relativa o absoluta depenent de la tècnica utilitzada per quantificar. La tècnica més utilitzada és l'anomenada SID (*stable-isotope dilution*), que consisteix en comparar la intensitat d'una transició determinada de la molècula a quantificar amb la mateixa transició d'una molècula patró químicament idèntica a l'analitzada però marcada isotòpicament després de coinjectar les dues molècules.

Phage Display

La tècnica *phage display* és una tècnica que permet l'estudi d'interaccions proteïna-proteïna i pèptid-proteïna, que es basa en la modificació del material genètic de bacteriòfags (fags) per tal d'expressar modificacions en les seves proteïnes i així poder crear, eventualment, biblioteques per l'exploració i cribratge d'alt rendiment.

Els bacteriòfags són un grup de virus que infecten cèl·lules procariotes. Contenen un material genètic encapsulat al si d'una llarga càpside proteica format per una única cadena circular d'ADN. Tenen la particularitat de no matar la cèl·lula hosta en el procés d'infecció i replicació. El seu material genètic està format per onze gens que codifiquen onze proteïnes, algunes amb funcions estructurals i d'altres amb funcions en el procés d'infecció, replicació o auto-assemblatge. La modificació del material genètic dels fags permet el disseny de biblioteques de pèptids que s'expressen fusionats a proteïnes del fag, exposats a la superfície d'aquest. Les proteïnes més utilitzades per aquest fi són les anomenades pVIII i pIII. La pVIII forma la càpside principal que embolcalla el material genètic i està present en grans quantitats (més de 2700 còpies). La pIII només està present en 5 còpies i es localitza a un dels extrems del fag. L'expressió dels pèptids en aquesta proteïna presenta certs avantatges respecte a fer-ho a la pVIII, tals com millor especificitat i capacitat de suportar majors inserts. Tanmateix, s'ha de tenir en compte que la proteïna pIII és la responsable de la infecció i que grans inserts la desafavoreixen i per tant pot afectar considerablement la capacitat de supervivència del fag i la seva utilitat durant els experiments.

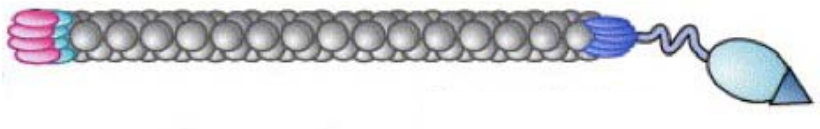


Figura 4. Esstructura esquemàtica d'un bacteriòfag. En gris, múltiples copies de la proteïna pVIII conformen la càpside principal, responsable d'encapsular l'ADN de cadena simple. En blau fosc, la proteïna pIII on s'hi expressa la biblioteca de pèptids desitjada.

Aquesta tècnica permet un procés de selecció per afinitat molt potent anomenat *panning* o *biopanning*. Aquest procés consisteix en la selecció dels fags positius en un assaig determinat i en l'enriquiment en assajos iteratius després d'un procés d'amplificació. Així, en estadis successius del procés de *panning*, la variabilitat de la biblioteca va minvant i es va enriquint amb individus que mostren especificitat per l'assaig en qüestió. Tanmateix, en certs casos pot no ser convenient abusar de les rondes de *panning*. Es poden donar casos on en l'amplificació de la biblioteca prèvia a cada ronda, la capacitat d'infecció i propagació de cada fag sigui diferent, i individus amb interacció poc específica però amb una gran capacitat de propagació dominin en relació als veritables fags d'interès amb bona capacitat d'interacció però poca capacitat de propagació. Aquest procés de selecció pot ser més o menys restrictiu segons el disseny de cada experiment. S'ha descrit tant en assajos *in vitro* com en assajos i*n vivo*.

L'aplicació *in vivo* de les tècniques de *phage display* va suposar un gran avenç en els estudis de biodistribució i la possibilitat de buscar molècules directores a diverses dianes específiques del cos. Així, s'han descrit estudis de *phage display* fins i tot en pacients humans amb mort cerebral per tal d'estudiar i poder treure informació valuosa de possibles molècules directores a diferents òrgans i teixits així com en cèl·lules cancerígenes.

Tot i el gran potencial d'aquesta tècnica, el *phage display* presenta alguns inconvenients. Per una banda, l'espai químic explorat està limitat als aminoàcids proteinogèncis. En aquest sentit s'han desenvolupat gran diversitat d'estructures on desenvolupar la variabilitat dels pèptids expressats per potenciar la variabilitat i rellevància biològica. Altres exemples permeten la selecció de pèptids D mitjançant l'assaig de fags en assajos *in vitro* contra proteïnes D. O fins i tot, s'han fet reaccions químiques amb les cadenes laterals de les cisteïnes del pèptid expressat al fag per

explorar altres espais químics. Per altra banda, un altre inconvenient lligat a l'anterior és la poca estabilitat a peptidases que presenten els pèptids seleccionats degut a la seva naturalesa L. Finalment, en alguns casos una altre maldecap és la solubilitat del pèptid quan és sintetitzat i administrat sense el fag. Tots aquests inconvenients fan necessari un treball químic d'ajust de les molècules seleccionades per *phage display* ja que al final l'efectivitat d'una molècula serà la combinació equilibrada de l'activitat, l'estabilitat i la solubilitat.

Resultats i discussió

Disseny, síntesi i avaluació de biblioteques peptídiques basades en llançadores de la BH per millorar-ne les seves propietats

Capítol 1. Estudi del rol de la quiralitat en el transport passiu a través de la BH

Les propietats fisicoquímiques de les molècules determinen la seva capacitat de transport a través de barreres biològiques per difusió passiva. En concret, la lipofilicitat, el pes molecular i les cadenes laterals dels aminoàcids que conformen un pèptid es consideren els principals descriptors que determinen la capacitat d'aquest a creuar la barrera hematoencefàlica per transport passiu.

Al laboratori, prèviament al projecte d'aquesta tesi, s'han estudiat els efectes de la *N*-metilació de l'esquelet peptídic, la influència de les cadenes laterals, l'halogenació, així com la llargada o flexibilitat dels pèptids en qüestió. En aquest capítol hem explorat la influència i el paper que juga l'estereoquímica en la difusió passiva a través de la BH, on els fosfolípids presents a la membrana són quirals.

Amb aquest propòsit es va sintetitzar una biblioteca d'estereoisòmers d'un pèptid descrit com una llançadora eficient: Ac-(*N*-MePhe)$_4$-NH$_2$. La biblioteca fou avaluada a través de l'assaig PAMPA (*parallel artificial membrane permeability assay*), que consta d'una membrana de fosfolípids equivalent a 300 bicapes lipídiques que mimetitza les condicions de transport passiu de la BH. El transport a través de l'assaig PAMPA es va determinar mitjançant les següents equacions per determinar la permeabilitat (P_e) i el percentatge de transport (*%T*):

$$P_e = \frac{-218.3}{t} \cdot log\left[1 - \frac{2 \cdot Q_A(t)}{Q_D(t_0)}\right] 10^{-6}\ cm/s; \qquad \%T = \frac{Q_A(t)}{Q_D(t_0)} \cdot 100$$

on $Q_A(t)$ és la quantitat de compost al pou acceptor a temps t i $Q_D(t_0)$ és la quantitat de compost al pou donador a l'inici de l'assaig. La quantitat de compost present a cada pou es determina mitjançant l'àrea compresa en el pic cromatogràfic després d'injectar cada mostra a l'HPLC.

Per a una anàlisi exhaustiva dels resultats, es va utilitzar el temps de retenció dels pèptids en HPLC com a mesura relativa de la lipofilicitat entre els membres de la biblioteca. Els resultats obtinguts, mostren una preferència en la permeabilitat dels compostos homoquirals a través de l'assaig PAMPA, així com també exhibeixen una menor retenció a la membrana fosfolipídica. Concretament la versió tot-D presenta una permeabilitat superior a la versió tot-L. Tot i que generalment s'accepta que a major lipofilicitat major és la difusió de molècules a través de membrana, els resultats mostren una correlació inversa entre lipofilicitat i permeabilitat, així com entre permeabilitat i retenció a membrana. Lògicament, trobem una correlació entre la lipofilicitat i la retenció a membrana. A més, hem analitzat la discriminació enantiomèrica, que indica la diferència de permeabilitat entre parelles enantiomèriques. En la majoria dels casos,

però no en tots, hem observat diferències significatives de permeabilitat entre parelles d'enantiòmers, però no hem pogut deduir una tendència que ens permeti predir els valors de discriminació enantiomèrica.

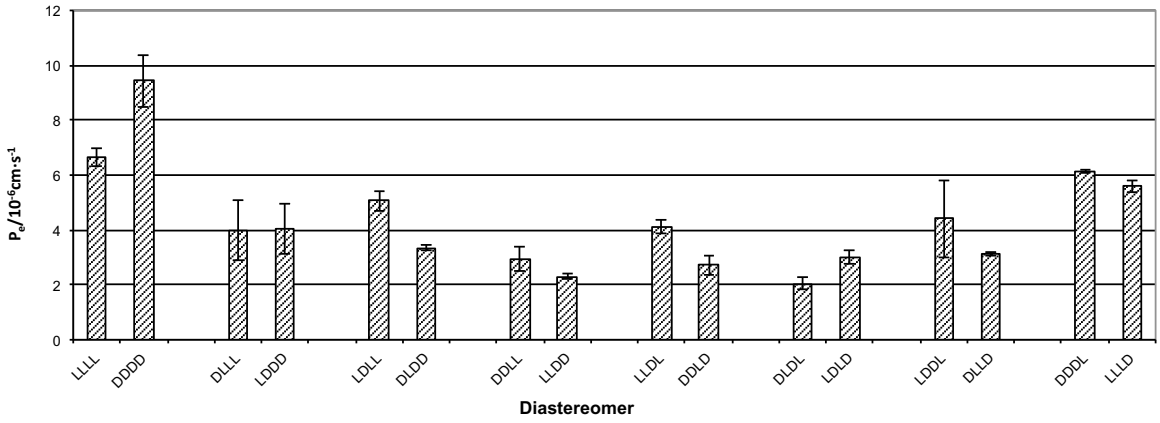


Figura 5. Resultats de transport a través de l'assaig PAMPA de les diverses parelles enantiomèriques presents a la biblioteca.

Valoració de la quantificació pel mètode d'espectrometria de masses

Tot i que la integració de l'àrea dels pics cromatogràfics és un mètode àmpliament acceptat per tal de calcular la permeabilitat a través d'assajos de transport, té algunes limitacions intrínseques que ens van portar a valorar la possibilitat de quantificar els assajos de transport amb espectrometria de masses. Entre els avantatges que presenta l'espectrometria de masses trobem una major selectivitat i sensibilitat. Tanmateix, la quantificació per espectrometria de masses requereix d'un patró intern que es comporti químicament de manera idèntica a la molècula a quantificar, ja que diferents pèptids es comporten de forma no equivalent en processos com la ionització o la detecció. Així doncs, per una correcta quantificació és necessària la presència d'un isotopòmer. Hi ha diferents maneres d'aconseguir isotòpomers, ja que poden ser generats químicament, metabòlicament o enzimàticament. En el cas que ens ocupa, s'utilitza la incorporació de reactius de síntesi que contenen àtoms marcats isotòpicament. Per a la quantificació, després de l'assaig de transport, a cada mostra a analitzar si afegeix una quantitat coneguda del patró intern i segons l'àrea o intensitat relativa dels pics corresponents a cada espècie podem conèixer la quantitat de pèptid present a cada compartiment. El MALDI-TOF va ser la tècnica d'espectrometria de masses utilitzada per a la quantificació rutinària de transport de pèptids. Per introduir els isòtops pesants al patró intern, cada pèptid es va sintetitzar en un mateix reactor fins a l'últim pas sintètic en què la resina es va dividir i en cada reactor s'hi va acoblar un bloc de síntesi amb diferent marcatge isotòpic. Es van analitzar diferents blocs de síntesi que portaven a isotopòmers amb pesos moleculars diferenciats per 3, 4 o 5 unitats de massa atòmica: CH_3-CO-, CD_3CO-; CH_3-CO-, $CD_3{}^{13}$CO-; C_6H_5CO-, C_6D_5CO-. Els pèptids es van assajar en l'assaig PAMPA i el transport es va quantificar utilitzant la parella pertinent en cada cas, tant per HPLC-UV com per MALDI-TOF. Per una banda, es va observar que la unitat C_6H_5CO- canvia les propietats de transport del pèptid dràsticament, fet que descarta la seva utilització ja que queda retingut a la membrana. Observem que amb la unitat CD_3CO-, el tercer pic isotòpic de la molècula que conté la unitat CH_3-CO- es solapa amb el pic molecular de la molècula marcada amb CD_3CO-. Així doncs, la parella CH_3-CO-, $CD_3{}^{13}$CO-; és la més adient per marcar i generar els pèptids i isotopòmers per la quantificació per espectrometria de masses.

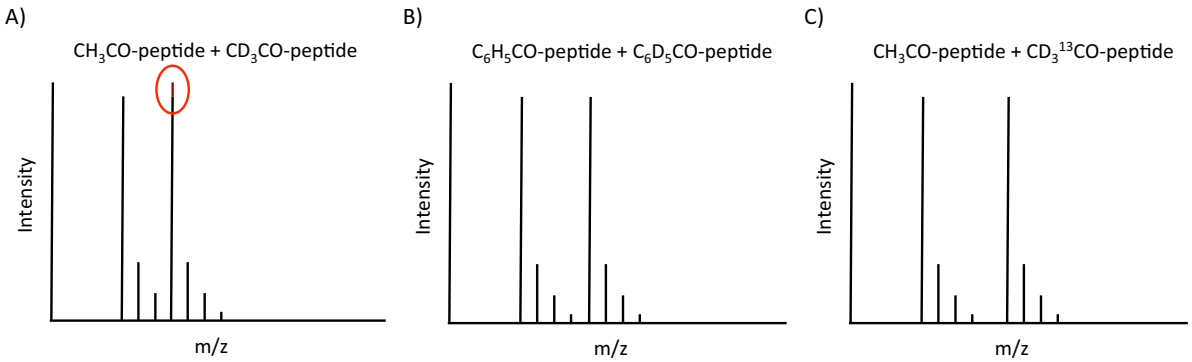


Figura 6. Perfil teòric en un espectre de masses d'una molècula i el seu patró intern amb 3, 4 o 5 u.m.a. superiors.

La quantificació de la permeabilitat i el transport per espectrometria de masses es determina mitjançant les mateixes equacions descrites anteriorment, per tant és necessari determinar $Q_A(t)$ i $Q_D(t_0)$. Aquests dos valors es determinen mitjançant una recta de calibrat de cinc punts. Cada punt consisteix en l'addició d'una quantitat major de patró intern respecte la mostra del pou acceptor a $t = t$ i la mostra del pou donador a $t = 0$. D'aquesta manera, interpolant cada recta es pot determinar $Q_A(t)$ i $Q_D(t_0)$ i així P_e i $\%T$. Així, la quantificació de cada espècie requereix de dues rectes de calibrat, que en total contenen 10 punts, i per tant 10 punxades al MALDI-TOF. És per això, que en el capítol 2 es discuteix un mètode simplificat de quantificació per tal de poder utilitzar rutinàriament la quantificació per espectrometria de masses en l'anàlisi del transport de biblioteques.

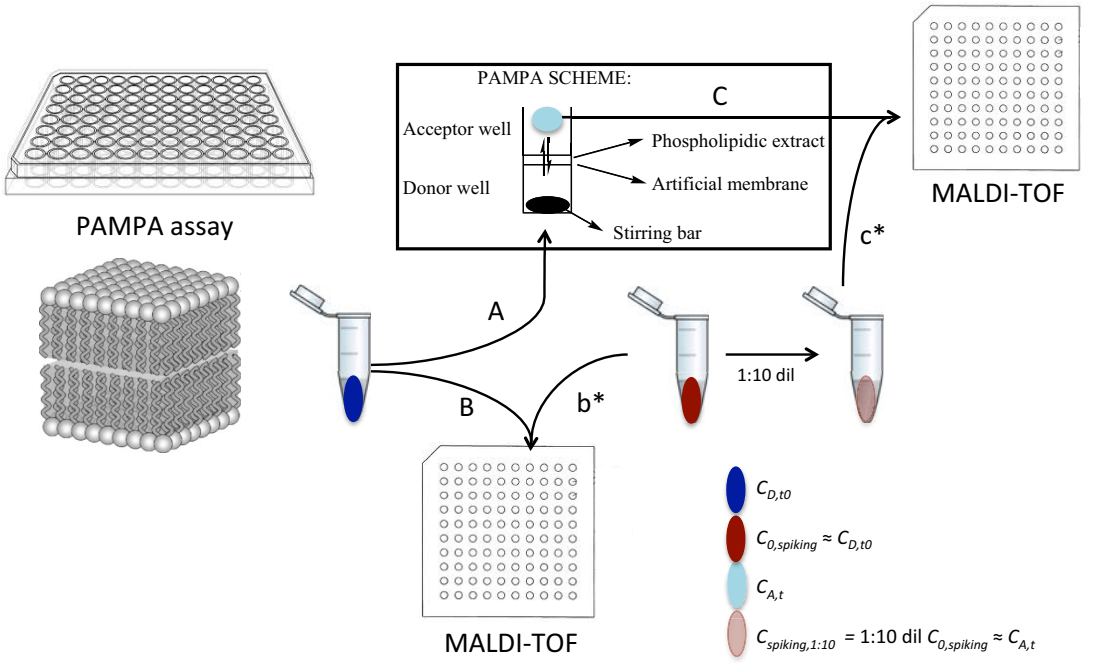


Figura 7. Esquema de la metodologia utilitzada en la quantificació per MALDI-TOF. Concentracions similars de patró i pèptid analitzat es barregen per determinar les concentracions relatives als pous donadors (t = 0) i acceptor (t = t).

Capítol 2. Modificació de llançadores peptídiques per millorar la seva capacitat de transport en models cel·lulars *in vitro* de BH

Com ja s'ha discutit anteriorment, la tècnica del *phage display* té un extraordinari potencial per estudiar les interaccions pèptid-proteïna i proteïna-proteïna, ja que permeten un anàlisi exhaustiu d'un espai químic que comprèn els aminoàcids proteinogènics. Tanmateix, normalment els pèptids seleccionats requereixen lleugeres modificacions per tal de millorar la solubilitat i/o la resistència a proteases, ja que l'eficiència d'una molècula es regeix per un compromís òptim entre activitat, solubilitat i estabilitat.

Al laboratori del Dr. Giralt, prèviament s'havia treballat amb un parell de pèptids descrits a la bibliografia capaços d'interaccionar amb el receptor de transferrina humana: H-HAIYPRH-NH_2 i H-THRPPMWSPVWP-NH_2. Treballs previs amb la segona molècula, per part del Dr. Roger Prades al si de la seva tesi al laboratori del Dr. Giralt, han permès desenvolupar una llançadora peptídica valuosa pel transport a través de models cel·lulars *in vitro* de BH. En aquest capítol doncs, hem explorat la millora en l'eficiència de transport a través de models cel·lulars *in vitro* de BH de H-HAIYPRH-NH_2.

Aquest primer objectiu del capítol s'ha abordat mitjançant la síntesi i avaluació d'una biblioteca d'anàlegs del pèptid Ac-HAIYPRH-NH_2 que ha permès explorar un espai químic més ampli que no pas l'explorat per *phage display* mitjançant l'ús de cadenes laterals no naturals. A més, s'ha dissenyat una biblioteca de 21 anàlegs amb l'esquelet peptídic amb estereoquímica D (11 anàlegs) i L (10 anàlegs). A més, s'ha utilitzat la unitat CH_3CO- en posició *N*-terminal per tal d'analitzar i validar el transport a través del model cel·lular *in vitro* de BH tant amb el mètode de quantificació amb HPLC-UV com amb el mètode simplificat de quantificació per MALDI-TOF.

Models de BH

Els models *in vitro* de BH encara són lluny de mimetitzar perfectament la BH *in vivo*. Tanmateix, són molts els esforços dedicats a dissenyar models *in vitro* basats en cocultius cel·lulars, que al cap i a la fi correlacionen força bé amb moltes de les característiques *in vivo* de la BH. Tot i que s'han descrit moltes combinacions de línies cel·lulars per mimetitzar la BH, la més acceptada i validada sembla ser la formada per cèl·lules endotelials de capil·lar de cervell boví i astròcits de rata. Aquest model implica treballar amb cultius de cèl·lules primàries, cosa que en complica la construcció, però alhora permet que el model expressi les principals característiques de la BH, tals com receptors específics, unions estretes o polarització de les cèl·lules endotelials. Tanmateix, la BH *in vivo* està subjecta a molts més estímuls mecànics o bioquímics que no queden representats al model *in vitro*.

Per controlar la semblança amb la BH, tenim unes eines que ens permeten determinar la qualitat del nostre model en certs paràmetres. Així doncs, disposem de la mesura de la resistència elèctrica (TEER) entre els dos compartiments del model separats per la membrana cel·lular que ens dóna informació sobre l'estanquitat de la membrana. També, l'assaig de transport del Lucifer Yellow (LY) ens dona informació sobre la formació d'aquestes unions estretes entre les cèl·lules endotelials característiques de la BH, així com possibles efectes citotòxics dels compostos assajats en el model. La presència de proteïnes especifiques, tals com ZO-1, occludines, P-gp o el receptor de transferrina, van ser avaluades pel Dr. Roger Prades en el si de la seva tesi doctoral.

Tots els anàlegs de la biblioteca van ser avaluats individualment a concentracions de 200 μM durant 2 hores a 37ºC en tampó Ringer/HEPES, coincubades amb LY (20 μM). Les mesures de TEER prèvies a l'assaig responien a valors de 116 ± 22 $\Omega \cdot cm^2$, superiors als valors mínims requerits (≈ 50 $\Omega \cdot cm^2$).

Quantificació

La quantificació del transport dels anàlegs assajats a través del model cel·lular *in vitro* de BH fou determinat mitjançant les tècniques de quantificació per HPLC-UV i MALDI-TOF aplicant les fórmules de permeabilitat aparent (P_{app}) i percentatge de transport ($\%T$).

$$P_{app} = \frac{\partial Q}{\partial t} \cdot \frac{1}{A \cdot C_0} \qquad\qquad \%T = \frac{Q_A(t)}{Q_D(t_0)} \cdot 100$$

on dQ/dt és el transport en funció del temps, A és l'àrea de la membrana cel·lular i C_0 és la concentració inicial al pou donador i on $Q_A(t)$ és la quantitat de compost al pou acceptor a temps t i $Q_D(t_0)$ és la quantitat de compost al pou donador a l'inici de l'assaig.

La quantificació per HPLC-UV es determina en base a l'àrea compresa sota els pics cromatogràfics. Tot i que, en la quantificació per MALDI-TOF, idealment seria convenient fer una recta de calibrat per a cada valor necessari per tal de calcular P_{app} i $\%T$, la quantificació d'una biblioteca de 21 anàlegs ens planteja la necessitat d'una simplificació del mètode que ens permeti un anàlisi rutinari més ràpid perdent el mínim de precisió, per finalment reanalitzar les mostres amb resultats més prometedors per tal de tenir un resultat de quantificació definitiu amb el mètode de la recta de calibrat.

Així, es va determinar el límit de quantificació per mescles de pèptids (pèptid:patró; 1:2) amb dilucions subsegüents 1:3 en dos equips MALDI-TOF disponibles. Així mateix, es va determinar el rang lineal en mescles amb diferents proporcions de pèptid i patró, compreses entre els valors de 0,2µM de pèptid i 20 µM de patró; i 20 µM de pèptid i 0,2 µM de patró. Finalment es va concloure que les concentracions de treball ($Q_D(t_0)$ ≈ 200 µM i $Q_A(t)$ ≈ 2 µM) no suposaven cap problema pel que fa al límit de detecció. Pel que fa a la diferència relativa entre pics per determinar de forma segura els valors de $Q_D(t_0)$ i $Q_A(t)$, es va concloure que era convenient treballar en diferències menors a 1:10, idealment menors a 1:3, amb l'equip MALDI-TOF (equip MALDI-4800).

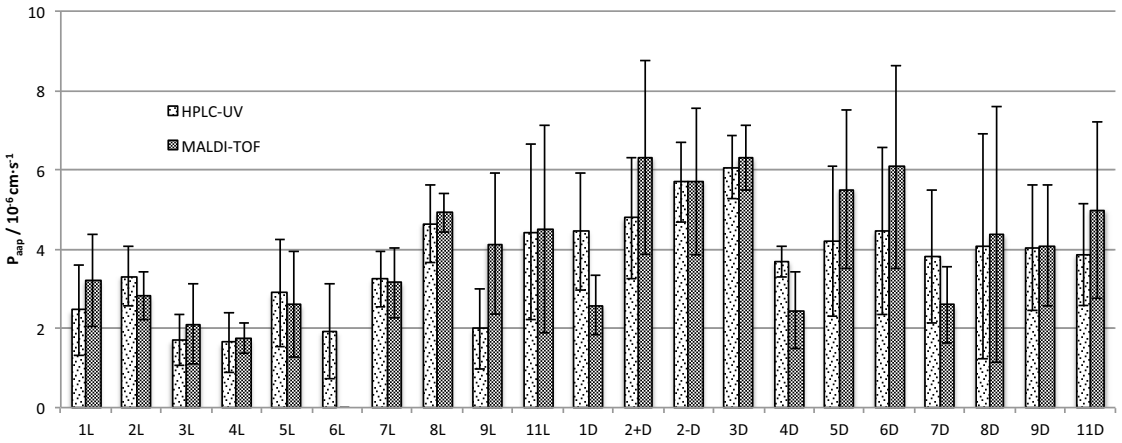


Figura 8. Resultats de quantificació dels anàlegs de la biblioteca HAIYRPH amb els dos mètodes de quantificació: HPLC-UV i MALDI-TOF.

Els resultats obtinguts tant amb el mètode de quantificació de HPLC-UV com amb MALDI-TOF no permeten determinar que hi hagi cap anàleg amb un transport significativament superior al pèptid base: Ac-HAIYPRH-NH_2. Malauradament, tampoc observem diferències significatives entre el transport del pèptid base en comparació a les formes *scrambled* (control negatiu). Aquest fet ens porta a pensar en dues hipòtesis. Primerament, que l'assaig cel·lular *in vitro* de BH no ha funcionat degudament (possiblement amb expressió massa limitada de receptors de transferrina, claus pel transport dels anàlegs assajats) tot i que els paràmetres de TEER i LY fossin correctes. Per

altra banda, pot ser que ens trobem en uns valors de selectivitat molt menors al que realment creiem estar i que els efectes d'uns canvis tan lleus entre els anàlegs no els puguem arribar a diferenciar. Tanmateix, en línies generals, els valors de transport dels pèptids amb esquelet D es mostren lleugerament superiors al dels anàlegs L, possiblement degut a la major estabilitat enzimàtica que presenten els primers.

Pel que fa a la valoració de la quantificació del transport entre els mètodes HPLC-UV i MALDI-TOF, els resultats tampoc ens permeten treure conclusions definitives. Si bé l'error de les mesures entre les dues tècniques de tots els anàlegs de la biblioteca entesa com a bloc és molt petit (0,59%), quan analitzem detalladament els resultats dels anàlegs individualment comprovem que el rànquing en funció del transport dels pèptids és diferent en cada tècnica.

Tot i que no podem concloure que la quantificació per MALDI-TOF és millor o comparable a la quantificació per HPLC-UV, si que és veritat que l'espectrometria de masses gaudeix d'una major sensibilitat, selectivitat i resolució, característiques que la fan més adequada i fiable en cas de treballar amb assajos cel·lulars *in vitro*, on molècules endògenes poden contribuir en l'àrea del pic cromatogràfic per HPLC-UV desviant les mesures per la quantificació. En canvi, en molts menys casos coincidiran en el senyal m/z de MALDI-TOF.

Tècniques de cribratge d'alt rendiment per descobrir noves llançadores de la BH

Capítol 3: Phage display: explorant biblioteques minimalistes

La feina presentada en aquest capítol és fruit d'una col·laboració i una estada del 22 de juliol de 2012 fins el 21 de gener de 2013 al laboratori de la Dra. Renata Pasqualini i el Dr. Wadih Arap al MD Anderson Cancer Center de Houston, TX, USA.

S'han descrit moltes variants de biblioteques de *phage display* a la bibliografia. Moltes d'elles dissenyades per explorar espais químics molt variats. Tanmateix, el disseny de la biblioteca de *phage display* proposada en el si d'aquest projecte busca explorar la que es considera la unitat peptídica mínima amb rellevància biològica, és a dir un tripèptid. Així, proposem la síntesi i avaluació *in vivo* (en ratolins BALB/c) de la biblioteca CX_3C, exposada a les proteïnes pIII dels fags. L'estructura de la biblioteca consta d'un pèptid cíclic format pel pont disulfur de les dues cisteïnes que constrenyen el tripèptid que expressa variabilitat (tots els aminoàcids proteinogènics).

Síntesi de la biblioteca de fags

Una biblioteca de les característiques de la CX_3C té una variabilitat de 8.000 individus, molts ordres de magnitud inferiors a les utilitzades més habitualment tals com CX_7C o CX_8C. Així doncs, és necessària una quantitat molt menor de vector d'ADN de fag (fUSE55) per tal de començar la síntesi de la biblioteca de fags. Començàrem la síntesi de la biblioteca amb 500 ng de fUSE55 i amb un protocol adaptat del descrit a la bibliografia. Inicialment, s'amplificà en bactèries DH5α, s'aïllà i es purificà el vector fUSE55. La purificació constà d'una doble purificació en kit *Qiagen Giga prep* i en una doble purificació per gradient de CsCl/EtBr. Simultàniament, s'acondicionà una biblioteca d'oligonucleòtids comercial responent al patró $(NNF)_3$, on *N* correspon a qualsevol dels nucleòtids mentre que *F* correspon a la guanina o la citosina, amb l'alineació amb els *primers* corresponents i l'elongació amb l'enzim, amb una purificació del material obtingut al final del procés. Tant la biblioteca d'oligonucleòtids com el vector fUSE55 varen ser digerits per l'enzim BGLI per tal d'inserir la biblioteca d'oligonucleòtids en el material genètic del fag. Abans de procedir amb la unió de la biblioteca d'oligonucleòtids i el vector fUSE55 i la ciclació del material genètic es van estudiar les proporcions idònies (fUSE55:biblioteca; 1:1, 1:3, 1:5, 1:10, 1:30) per fer

la reacció. Un cop conegudes les millors condicions (1:5) es va procedir a fer la reacció i a continuació el material genètic s'electroporà en cèl·lules bacterianes electrocompetents (MC-1061), que es van cultivar i van produir la biblioteca de fags. Finalment es van precipitar els fags amb NaCl/PEG.

A continuació es va procedir a validar la biblioteca produïda. Es va titular una mostra de la biblioteca per conèixer-ne la concentració i a més es van seqüenciar 192 fags corresponents a 192 colònies de bacteris infectats per cada fag. Els resultats de la seqüenciació van evidenciar que més d'un 95% dels fags contenien l'insert introduït. A mes, la variabilitat d'aminoàcids presents en la biblioteca representava tots els aminoàcids proteinogènics, tot i que en proporcions variables respecte les esperades. Aquest però, és un fet habitual ja que en la síntesi de la biblioteca de fags, hi ha factors que són difícilment controlables tals com la qualitat de la biblioteca d'oligonucleòtids o bé que la mostra seqüenciada no sigui completament significativa. Tanmateix, la biblioteca es va considerar apte per continuar el projecte.

La biblioteca de fags es va avaluar *in vivo* (ratolins BALB/c) buscant seqüències dirigides preferentment al cervell, potencials llançadores de la BH. El destí dels fags circulant pel flux sanguini pot resumir-se en quatre escenaris. Un, que creuin la BH i arribar al parènquima cerebral. Dos, que internalitzin les cèl·lules endotelials dels capil·lars del cervell, però que no arribin a creuar la BH. Tres, que interaccionin amb algun receptor, proteïna o membrana de la cara luminal de les cèl·lules endotelials dels capil·lars. Quatre, que no experimentin cap interacció als capil·lars cerebrals i que continuïn en el flux sanguini. Les llançadores peptídiques de la BH exigeixen el primer escenari.

Es va dissenyar una avaluació amb tres fases d'enriquiment (*panning*) focalitzat en el cervell. Així, es va injectar $5{\cdot}10^9$ TU amb injecció intravenosa a la vena lateral de la cua dels ratolins (triplicat). Al cap de 6 hores es va procedir a fer la perfusió dels animals amb PBS. Els òrgans d'interès (cervell, fetge i ronyó) es van seleccionar, triturar i lisar. Finalment, es va afavorir la infecció de bacteris pels fags retinguts en els òrgans estudiats per tal d'amplificar-los. Un cop amplificats durant tota la nit, els fags van ser precipitats i titulats per procedir a la segona fase de *panning* injectant altra vegada $5{\cdot}10^9$ TU de la biblioteca enriquida després de la primera fase. Així es van fer 3 rondes de *panning*, on en cada ronda es van seqüenciar 96 fags a l'atzar. Finalment es va procedir a fer una última fase de selecció per evitar desviacions degudes a la velocitat de propagació i replicació dels fags. Les dotze seqüències considerades més prometedores ja que havien aparegut en totes les rondes de *panning* o molt repetides en l'última ronda, van ser amplificades individualment i titulades. S'injectaren una barreja de $5{\cdot}10^9$ TU de cada fag en tres ratolins i es va procedir al mateix procés descrit anteriorment. D'aquesta última fase de selecció en sortiren tres candidats: CFLFC, CRWEC i CNSQC amb forta afinitat pel cervell. Tanmateix, amb aquests resultats no hi ha evidència que els pèptids siguin llançadores a la BH. Per això fou necessari un estudi exhaustiu de cada una de les seqüències seleccionades per estudiar la biodistribució amb q-PCR i la localització en el cervell amb immunofluorescència.

La biodistribució es va avaluar injectant el fag juntament amb el fag fd-Amp (control negatiu, fag sense insert) en un ratolí durant 6 hores. Després de perfondre el ratolí es van recollir els òrgans i teixits d'interès (cervell, pulmó, fetge, pàncrees, melsa, ronyó, múscul i greix) per procedir a l'extracció de l'ADN dels fags per fer q-PCR (quantificació del material genètic dels fags). Així, es va comparar quantitativament la presència dels fags seleccionats en cada òrgan i l'enriquiment en comparació al fag control (fd-Amp). Les tres seqüències estudiades es comporten en termes força semblants pel que fa a la biodistribució. S'observa una afinitat remarcable (fins a un ordre de magnitud) del les seqüències estudiades respecte el fag control al cervell. Tanmateix i com era d'esperar, els fags s'acumulen en quantitats superiors en altres òrgans (principalment el fetge).

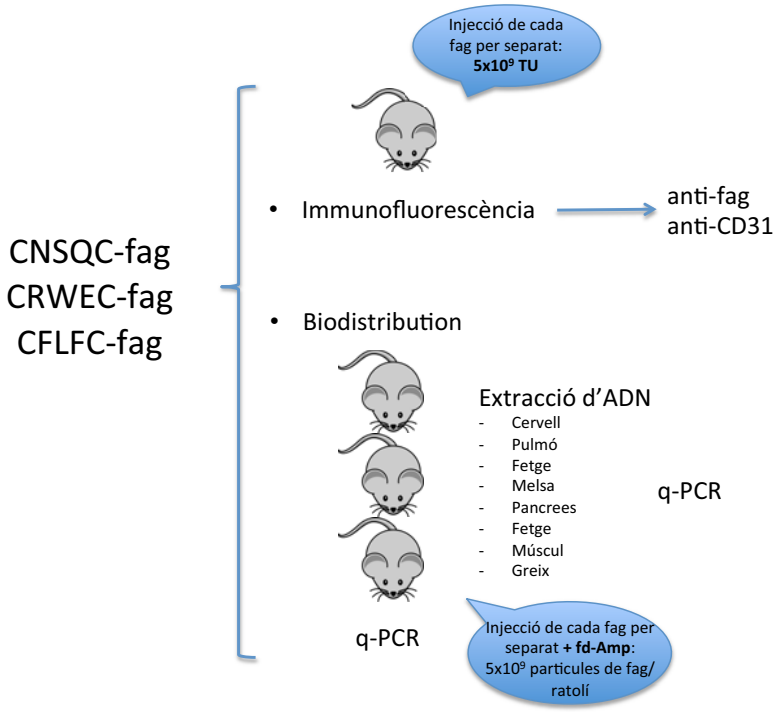


Figura 9. Esquema dels experiments d'immunofluorescència i biodisbribució *in vivo*.

Per altra banda, la localització es va estudiar injectant cada fag seleccionat ($5{\cdot}10^9$ TU) en un ratolí diferent. Al cap de 6 hores, es va perfondre l'animal i el cervell (així com els òrgans control, fetge i ronyó) es van recollir i fixar en 4% paraformaldehid i es van tractar amb sucrosa (amb solucions del 10%, 20% i finalment 30%). El cervell es va tallar en làmines fines de 10 µm amb un criòstat i es va procedir a la immunotinció amb anticòs anti-fag i anti-CD31, un marcador de les cèl·lules endotelials, per estudiar la colocalització i així poder visualitzar la localització dels fags dintre el cervell. Tot i que els experiments de colocaliztació van fallar degut a la no tinció dels marcadors CD-31, es van treure algunes conclusions amb les tincions dels fags i les marques del DAPI indicant els nuclis de les cèl·lules. Així es va deduir que en molts casos els fags semblen estar en el si dels capil·lars, però no es pot determinar si estan internalitzats o adherits a les cèl·lules endotelials o per contra les han travessat i estan localitzats als astròcits que abracen les cèl·lules endotelials que conformen els capil·lar sanguinis al cervell i la BH. Els resultats són semblants per les tres seqüències estudiades, però en el cas de CRWEC hi ha una imatge especialment prometedora mostrant tinció fora d'un capil·lar formant un triangle, forma típica dels peus astrocítics.

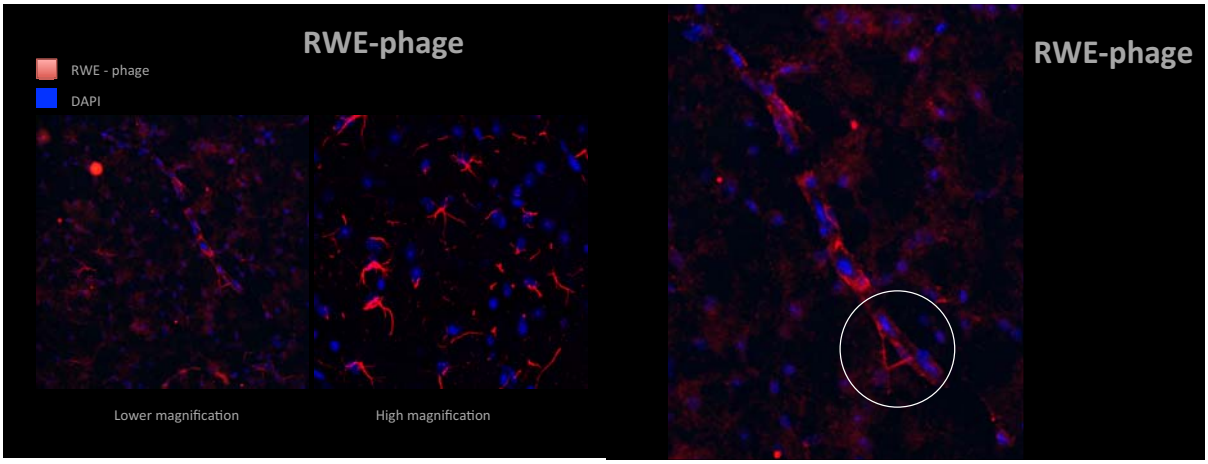


Figura 10. Imatges de la immunotinció (anticòs antifag, vermell; DAPI, blau) de cervell de ratolí injectat amb el fag CRWEC, vist al microscopi confocal.

Capítol 4: Més enllà del phage display: utilització de l'espectrometria de masses per la identificació i quantificació de biblioteques peptídiques a través de models de BH

Les tècniques de cribratge d'alt rendiment són utilitzades per identificar molècules amb un efecte determinat entre biblioteques amb grans quantitats de molècules. Hi ha tres paràmetres que defineixen cada tècnica tals com la naturalesa de la biblioteca que s'utilitza, el model per analitzar la biblioteca i finalment l'eina d'identificació apropiada a les necessitats dels dos primers paràmetres. En aquest capítol, es desenvolupa una tècnica basada en una biblioteca peptídica sintetitzada químicament, analitzada en models cel·lulars *in vitro* de BH i analitzada amb tècniques avançades d'espectrometria de masses.

Disseny i síntesi de la biblioteca

El disseny de la biblioteca es va fer en base a la viabilitat de la síntesi i intentant afavorir la detecció dels pèptids per espectrometria de masses. El mètode de síntesi utilitzat va ser la síntesi *separa i mescla* en fase sòlida. Aquest consisteix en dividir el polímer on es duu a terme la síntesi en diversos reactors per en cada un fer-hi un acoblament amb un aminoàcid diferent. Un cop acabada la reacció d'acoblament i desprotecció, la resina polimèrica es barreja homogèniament en un reactor mes gran. El procés es repeteix cíclicament en cada posició del pèptid on es vol introduir variabilitat. Aquest mètode doncs, permet la síntesi de biblioteques molt àmplies. Si bé una de les característiques és que en cada bola de resina només s'hi troba una seqüència determinada, després de l'escissió dels pèptids de la resina tots els pèptids de la biblioteca queden barrejats en un mateix recipient. Per tal de començar la síntesi amb garanties que totes les variants (pèptids) siguin sintetitzades va ser necessari definir el nombre mínim de boletes de resina per començar la síntesi. Aquests càlculs es van fer en base a altres estudis descrits en la bibliografia. Paràmetres com la llargada del pèptid i la variabilitat introduïda en cada posició defineixen finalment el nombre de boletes necessari per començar la síntesi amb garanties. Així es van considerar introduir una variabilitat de 7 aminoàcids en 5 posicions en hexapèptids. Un residu de D-Arg es va establir com a aminoàcid *N*-terminal de tots els pèptids per afavorir la ionització dels pèptids de la biblioteca ja que tots els pèptids de la biblioteca són acetilats en *N*-terminal i amidats en *C*-terminal i no permeten l'estabilització de càrregues en cap dels dos punts terminals. Els aminoàcids escollits per introduir variabilitat en les cinc posicions contigües al *C*-terminal són: D-Ala, D-Arg, D-Ile, D-Glu, D-Ser, D-Trp i D-Pro, que constitueixen una base reduïda d'aminoàcids que representa les propietats sintètiques, fisicoquímiques i paràmetres estadístics en funció de la seva presencia en centres actius de proteïnes amb rellevància biològica (Figura 8).

Un cop sintetitzada, la qualitat de la biblioteca es va validar amb dos mètodes. Primer, abans de l'escissió de tots els pèptids de la resina es va comprovar que fent l'escissió del pèptid present en una sola bola de resina, aquest fos pur i coincidís amb el patró Ac-D-Arg-XXXXX-NH_2, fet que vam comprovar en 15 boletes escollides a l'atzar. En segon lloc, després de procedir a l'escissió dels pèptids de la resina, una mostra de la biblioteca es va injectar amb infusió directa a l'equip d'espectrometria de masses FT-ICR i el perfil de l'espectre es va comparar amb el d'un espectre teòric considerant que la capacitat d'ionització de tots els pèptids fos igual. En aquest cas, es va concloure que el rang de masses de la biblioteca experimental era el correcte i que el perfil dels espectres teòrics i experimentals era significativament similar.

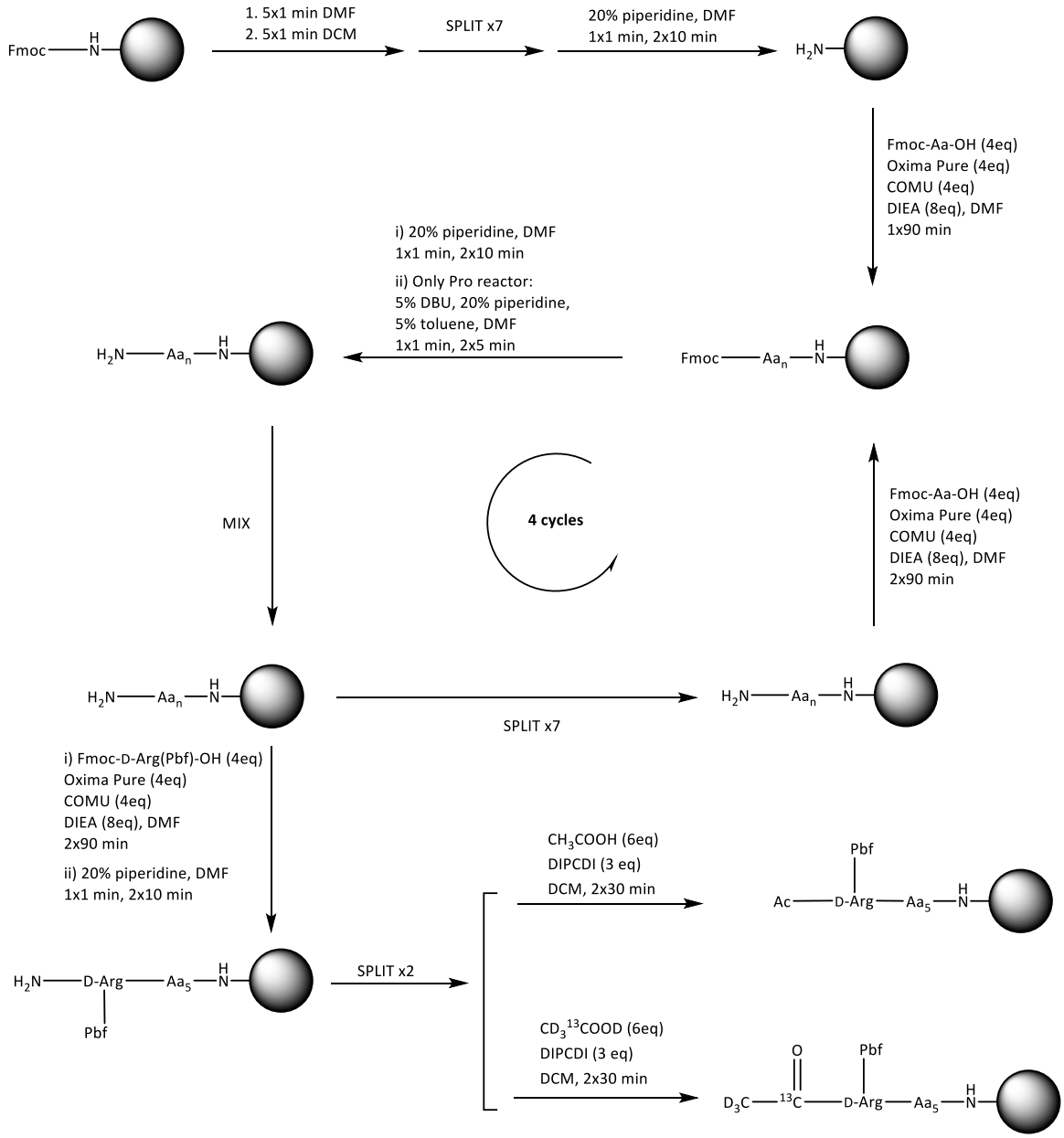


Figura 11. Esquema sintètic de la biblioteca Ac-D-Arg mitjançant el mètode *separa i mescla*.

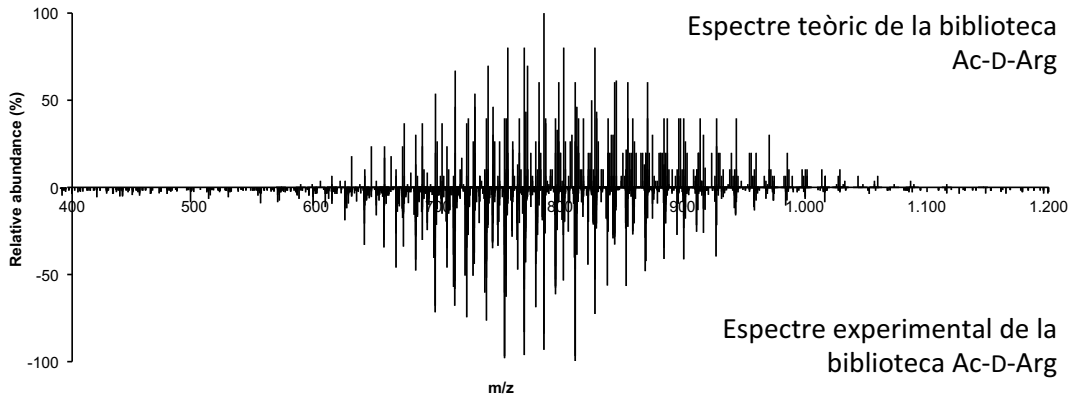


Figura 12. Perfil de l'espectre de masses de la biblioteca Ac-D-Arg obtingut en un FT-ICR (superior) i l'espectre teòric (inferior).

Propostes per a la identificació

La complexitat de les biblioteques sintetitzades amb el mètode *separa i mescla* rau en l'extrema similitud entre els pèptids sintetitzats. Així, en el cas descrit, tenim una biblioteca d'hexapèptids on en 5 de les posicions aminoacídiques hi ha variabilitat (7 aminoàcids). Específicament, aquesta biblioteca consta de 462 composicions diferents (famílies peptídiques), i cada una d'aquestes composicions o família estarà representada per una sèrie de seqüències que correspondran a totes les permutacions possibles segons els aminoàcids que la formin. Tot plegat, genera una biblioteca amb 16.807 pèptids teòrics. Podem parlar doncs, d'una biblioteca molt redundant pel que fa a la massa molecular dels compostos que la formen. Tot i així, l'assaig de la biblioteca a través de l'assaig cel·lular *in vitro* de BH hauria de reduir en gran mesura la quantitat de pèptids a analitzar finalment per espectrometria de masses.

La primera aproximació per analitzar la biblioteca assajada *in vitro* que vam considerar va ser basada en el model utilitzat en proteòmica *bottom-up* o *shotgun proteomics*. Si bé la nostra mescla de pèptids no ha estat generada per la digestió enzimàtica d'una proteïna, a l'hora d'injectar la nostra mostra problema tenim una mescla de pèptids complexa. L'estratègia *bottom-up* es basa en la injecció i anàlisi amb espectrometria de masses acoblat a cromatografia de líquids d'una mostra proteica digerida enzimàticament. La informació obtinguda experimentalment es correlaciona amb una sèrie de bases de dades teòriques generades a partir de la seqüenciació genòmica. Hi ha varies eines tals com el SEQUEST o el Mascot entre d'altres que permeten l'elucidació de la proteïna a partir de l'anàlisi de la mescla de pèptids generada enzimàticament mitjançant una sola injecció a l'espectròmetre de masses, on s'analitzen els espectres de pics moleculars o primera generació (MS1) i els espectres de fragmentació o de segona generació (MS/MS). Un dels equips capaços de dur a terme aquestes anàlisis és el LTQ-Orbitrap.

Tanmateix, aplicar aquesta aproximació al nostre cas no va servir per poder analitzar les mostres provinents de l'assaig cel·lular *in vitro* de BH amb la biblioteca Ac-D-Arg. Això fou degut bàsicament a dues raons. Primera, perquè els ions precursors de seqüències corresponents a una mateixa família es fragmenten i s'analitzen conjuntament, produint uns espectres MS/MS d'identificació quimèrica. Segona, perquè els controls de validació dels resultats obtinguts per part dels algoritmes anomenats anteriorment utilitzen seqüències control que en el nostre cas formen part de la mateixa biblioteca, ja que conté totes les permutacions possibles. Per tot això ens vam veure obligats a descartar aquesta aproximació.

Així doncs, es va considerar una altra aproximació menys estàndard basada en dues fases. La primera fase consistia en seleccionar les composicions (o famílies peptídiques) més abundants o que s'identifiquen més fàcilment al nivell d'espectre MS1 amb LTQ-Orbitrap. En aquest punt, i havent seleccionat les famílies més prometedores en aquest estadi, la segona fase d'identificació es desenvolupa amb la tecnologia SRM (*single reaction monitoring*) amb un equip de triple quadrupol, que permet amb gran sensibilitat monitoritzar les transicions (combinació de ió precursor i ió de fragmentació) desitjades. Així, de les famílies seleccionades es monitoritzen tots els fragments *b* i *y* dels ions precursors d'una família (mono- i bi-carregats), per definir quines de les seqüències concretes de cada família monitoritzada són inequívocament a la mostra. Com que cada injecció al triple quadrupol només pot monitoritzar 120 transicions i la mostra és limitada, és de vital importància tenir un mètode racional per seleccionar les famílies de més interès.

Els criteris de selecció de les famílies o composicions considerades per monitoritzar les transicions de tots els seus membres o seqüències són: primer, que aquella composició sigui present al pou acceptor dels tres triplicats de l'assaig cel·lular *in vitro* de BH. Segon, que els pics moleculars que representen una determinada composició tinguin el perfil isotòpic típic d'una espècie peptídica i que no sigui el pic isotòpic d'un altre pic molecular. Tercer, a partir d'una ponderació segons la intensitat de cada composició en cada triplicat es trien aquelles composicions més intenses i congruents en el si dels tres triplicats. Finalment, i quart, es descarten les famílies amb més de 30

membres ja que el mètode manual d'assignació de transicions fa que surti més a compte analitzar més varietat de famílies amb menys membres cada una.

Assaig *in vitro* de la biblioteca

En el si d'aquest projecte s'han utilitzat dos assajos *in vitro* de la BH per assajar la biblioteca. A més, s'han fet algunes proves *in vivo* amb ratolins BALB/c que es comentaran al final del capítol. Primerament, s'ha utilitzat l'assaig PAMPA ja descrit anteriorment. Es tracta d'un assaig *in vitro* basat en una membrana de fosfolípids de cervell porcí que mimetitza les propietats de difusió passiva presents a la BH. La biblioteca va ser assajada a concentració de saturació en aigua durant 4 hores. Per altra banda, l'assaig cel·lular *in vitro* de BH consta d'un cocultiu de cèl·lules endotelials de capil·lar de cervell boví i astròcits de rates. Aquest model mimetitza tan les condicions de transport actiu com les de transport passiu de la BH. La biblioteca va ser assajada a la concentració de 10% de saturació en tampó Ringer/HEPES durant 2 hores. Tot i que la mostra va ser assajada en dissolvents aquosos, la injecció al FT-ICR de la biblioteca dissolta en dissolvent orgànic o aquós no divergia en excés, assenyalant que no perdem la diversitat de la biblioteca en l'assaig. Les mostres assajades van ser dessalades amb una mini-columna C_{18} per poder injectar als espectròmetres de masses sense problemes pel funcionament de l'equip.

Primer nivell d'identificació

Les mostres dessalades es van injectar al LTQ-Orbitrap i la llista de pics moleculars es va extreure i analitzar amb un programa (*Bibliopepfinder*) dissenyat per Xavier Arroyo del laboratori del Dr. Giralt per filtrar aquells pics que poguessin correspondre a composicions presents a la biblioteca. Les dades van ser analitzades considerant un error en la detecció de fins a 10 ppm. Així doncs, el programa Bibliopepfinder genera una llista amb la composició aminoacídica de la família corresponent a cada senyal m/z, amb la intensitat i la càrrega del ió.

Els resultats obtinguts per les mostres assajades en PAMPA són una llista de 187 famílies candidates després d'haver restat els senyals que apareixien en una mostra del blanc. Així mateix, pel model cel·lular *in vitro* de BH, es va generar una llista amb 367 senyals corresponents a possibles famílies de la biblioteca Ac-D-Arg. D'aquestes, 286 es trobaven en els tres triplicats.

En aquest punt es va determinar un altre criteri per l'anàlisi de les dades amb la creació de dues llistes diferenciades de candidats. Per una banda, aquelles famílies presents en l'assaig PAMPA i en el cel·lular *in vitro* (155) van ser agrupades en una llista de candidats a llançadores de la BH per *difusió passiva*. Per altra banda, les famílies només presents en els triplicats del model cel·lular, però no presents en PAMPA es van incloure a la llista de candidats a llançadora per *altres tipus de transport*. Així i amb els criteris de selecció de famílies per la segona fase d'identificació es van seleccionar dos grups de famílies candidates.

- Transport per difusió passiva: I3-P2-R1, P3-S2-R1 i A3-R2-P1.
- Altres mecanismes de transport: I3-A1-P1-R1, E2-P2-A1-R1, P2-W2-I1-R1 i I4-P1-R1.

Segon nivell de quantificació

Un cop escollides les famílies candidates es van generar totes les possibles seqüencies de cada composició tenint en compte que tots els pèptids tenen una arginina com a residu *N*-terminal. Aquestes seqüències van ser introduïdes en el software lliure Skyline que permet generar les transicions desitjades i determinar les condicions d'anàlisi amb el triple quadrupol.

Després de la injecció de la mostra i l'anàlisi en el triple quadrupol el mateix software permet l'anàlisi i el tractament dels resultats obtinguts. Els espectres mostren les transicions monitoritzades amb una traça d'un color determinat per a cada una en funció del temps de retenció al cromatograma. La coelució de suficients transicions per definir una seqüència

determinada assegura la prèsencia d'aquesta a la mostra. Els espectres obtinguts en la injecció al triple quadrupol van ser més complexos del que esperàvem pel que fa a la quantitat de pics que hi apareixien. Tanmateix, no tots els pics presents a l'espectre mostren un perfil complet de traces de transicions coeluint i definint una seqüència determinada. Aquest fet l'explicaria que diverses seqüències d'una família presentessin temps de retenció diferents. Si fos així, seria lògic que apareguessin perfils incomplets de transicions en altres temps de retenció diferents al de la seqüència monitoritzada, ja que en una mateixa família molts dels pèptids compartiran moltes de les transicions. Per exemple, les seqüències RAIIPI i RAIIIP comparteixen totes les transicions a excepció de la corresponent als ions de fragmentació y_1 i b_1. Per tant, en aquest cas, aquesta transició es imprescindible per assignar quina de les dues seqüències hi ha a la mostra. Igualment, és possible que seqüències de la mateixa família coelueixin i per tant en un mateix temps de retenció poden aparèixer més d'una seqüència si és que coincideixen a la mostra.

Una altra observació tinguda en compte per a l'anàlisi dels resultats obtinguts és la congruència dels resultats en els triplicats: si els pics on coelueixen les transicions són en temps de retenció comparables, si els trobem en ions precursors diferents o si l'assignació és completa o en alguns casos deduïda (per exemple en el cas RAIIIP on les trancisions b_3-y_3 i b_4-y_2 no són estrictament necessàries per a l'assignació).

La complexitat de les mostres és accentuada per un altre factor determinant: la finestra de filtratge de ions precursors del segon quadrupol. Aquesta finestra es troba al voltant dels 0.7 u.m.a. i per tant, espècies que tinguin una massa dintre el rang $m/z_{seleccionat} \pm 0.7$ poden aparèixer a l'espectre SRM si hi ha algun ió de fragmentació que coincideixi amb els ions monitoritzats.

Other mechanisms transport list		*Passive diffusion transport list*		
I3-A1-P1-R1	I4-P1-R1	I3-P2-R1	P3-S2-R1	A3-R2-P1
RAIIIP	RIIIIP	RIIIPP	RPPPSS	RAAAPR
RAIIPI	RIIIPI	RIIPIP	RPPSPS	RAAARP
RAIPII	RIIPII	RIIPPI	RPPSSP	RAAPAR
RAPIII	RIPIII	RIPIIP	RPSPPS	RAAPRA
RIAIIP	RPIIII	RIPIPI	RPSPSP	RAARAP
RIAIPI		RIPPII	RPSSPP	RAARPA
RIAPII		RPIIIP	RSPPPS	RAPAAR
RIIAIP		RPIIPI	RSPPSP	RAPARA
RIIAPI		RPIPII	RSPSPP	RAPRAA
RIIIAP		RPPIII	RSSPPP	RARAAP
RIIIPA				RARAPA
RIIPAI				RARPAA
RIIPIA				RPAAAR
RIPAII				RPAARA
RIPIAI				RPARAA
RIPIIA				RPRAAA
RPAIII				RRAAAP
RPIAII				RRAAPA
RPIIAI				RRAPAA
RPIIIA				RRPAAA

Figura 13. Figura-resum dels resultats obtinguts. L'escala de colors a la part inferior de la figura indica la coherència de l'assignació dels espectres en els triplicats de cada mostra (vermell-marró, més; blanc, menys).

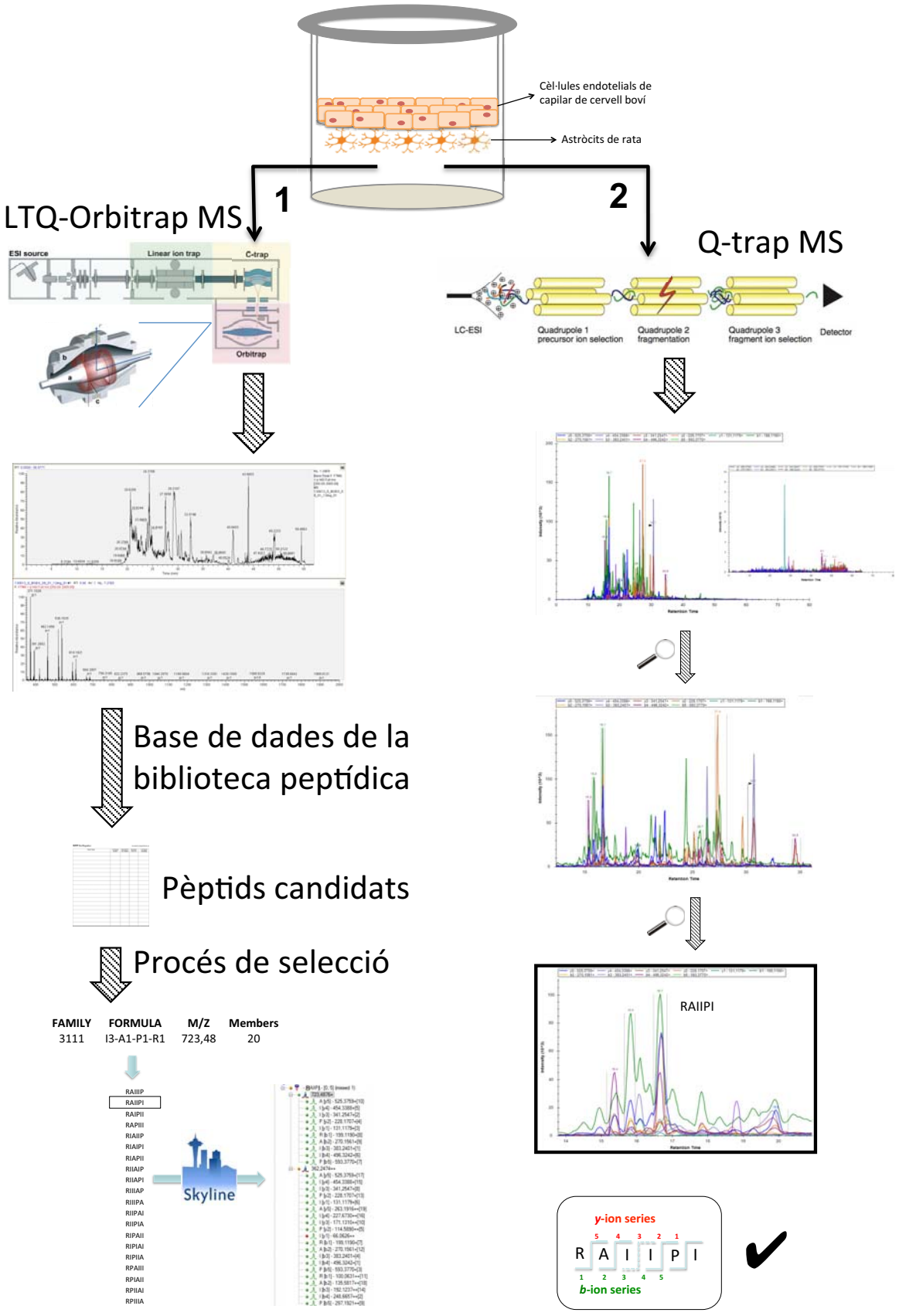


Figura 14. Esquema del procés d'identificació de pèptids de la biblioteca amb espectrometria de masses capaços de creuar l'assaig de barrera hematoencefàlica mitjançant espectrometria de masses.

Assaig *in vivo* de la biblioteca

Seguint la traça oberta per la Dra. Renata Pasqualini quan el 1996 va descriure l'aplicació *in vivo* del *phage display*, en el si de l'estada al seu laboratori vam explorar l'aplicació que podria tenir l'anàlisi de biblioteques peptídiques complexes sintetitzades químicament com la Ac-*D*-Arg en assajos *in vivo*.

Es van injectar a la vena lateral de la cua de ratolins BALB/c (duplicats) 200 µL en concentració saturada en PBS de la biblioteca Ac-*D*-Arg i es van estudiar 3 temps de circulació (30 minuts, 6 hores i 24 hores). Com a control es va injectar amb la mostra un 2% de blau d'Evans per controlar l'estanqueïtat de la BH. Els animals es van perfondre amb PBS i els òrgans i teixits d'interès (cervell, pulmó, cor, fetge, pàncrees, melsa, ronyó, múscul i greix) es van recollir, triturar i congelar.

La major dificultat trobada en aquesta aproximació sembla ser la neteja de la mostra per la injecció en els equips d'espectrometria de masses. Es van explorar diverses possibilitats en els casos del cervell, el fetge i el ronyó. Primerament per lisar el teixit dels òrgans es va considerar sonicar el teixit triturat en una dissolució de 4% SDS en tandes de 15 segons amb esperes de 15 segons en un bany de gel. A continuació es van avaluar tres possibilitats per eliminar proteïnes de la mostra: per una banda, la precipitació durant 16 hores a -20ºC amb un dissolvent orgànic (metanol i acetona) i posterior centrifugació. Per altra banda, es va filtrar la mostra en filtres de centrífuga per molècules menors a 10.000 Da. Tanmateix, la precipitació amb dissolvents orgànics no resultava en una mostra suficientment neta com per poder-la injectar als equips. Per tant es va considerar només el filtratge de les proteïnes amb els filtres de centrífuga. De totes maneres, aquests mètodes poden suposar pèrdues molt grans dels pèptids de la biblioteca.

Amb les mostres de cervell processades per injectar a l'espectròmetre de masses es van considerar dues aproximacions. Primer, es va injectar directament al triple quadrupol cercant les seqüències validades en els assajos *in vitro*. Si bé en les mostres de la llista de transport per *difusió passiva* no es van trobar transicions coeluint, sí que es van trobar per la família I4-A1-R1, corresponent a la llista d'*altres mecanismes de transport*. Tanmateix el temps de retenció on coeluïen les transicions era completament diferent (t_R = 45 min, mentre que *in vitro* era entre t_R = 15-30 min) i comprovant si hi havia alguna espècie amb pes molecular coincident en les llistes tant d'injeccions *in vitro* com *in vivo* en LTQ-Orbitrap, el pic corresponent a aquell m/z era un pic isotòpic. Per tant es va concloure que es tractava d'un artefacte o un fals positiu.

Per altra banda, es va intentar reproduir la mateixa aproximació utilitzada per les mostres *in vitro*. És a dir, un primer nivell d'identificació amb LTQ-Orbitrap amb selecció de les famílies més prometedores i un segon nivell d'identificació amb SRM i triple quadrupol. En aquest cas es va procedir a analitzar les tres mostres netejades (temps de circulació *in vivo* de 6 hores) de cervell, fetge i ronyó. En un primer nivell d'identificació, tant en cervell com en fetge, molt poques van ser els pics corresponents a possibles famílies de la biblioteca i per tant no es va poder progressar amb aquesta aproximació. Pel que fa al fetge, es van trobar algunes famílies candidates tals com P4-E1-R1, A4-R2, E3-W2-R1, E3-A2-R1 i E4-I1-R1, però cap d'aquestes seqüències va poder ser validada en el segon nivell d'identificació. Per això discutim un disseny lleugerament modificat de la biblioteca que permeti millorar les prestacions d'identificació *in vivo*, sobretot facilitant la recuperació dels pèptids del teixit.

Propostes per millorar l'aproximació *in vivo*: modificació del disseny de la biblioteca

Proposem dos modificacions principals per millorar la recuperació dels pèptids en el teixit i per diferenciar els pèptids dintre la biblioteca amb un espectre més ampli de propietats fisicoquímiques tot disminuint-ne lleugerament la complexitat. Així doncs es proposa que en la posició *N*-terminal hi aparegui una biotina per poder col·lectar els pèptids amb la interacció biotina-estreptavidina amb boletes magnètiques recobertes d'estreptavidina. Per altra banda, l'estructura dels pèptids estaria basada en tres posicions on s'introduiria variabilitat, separades

entre elles per dues glicines. A més, a l'extrem *N*-terminal els dos últims residus serien dues lisines encarregades de poder enllaçar un *cargo*, d'entrada un grup acetil. A l'extrem *C*-terminal, una lisina asseguraria la possibilitat d'estabilització d'una càrrega per poder identificar els pèptids per espectrometria de masses sense problemes tot i que l'extrem *C*-terminal estigui amidat.

Biotina--K(Ac/Ac*)-K(Ac/Ac*)-K-G-G-X_1-G-G-X_2-G-G-X_3-G-G-K-NH_2.

L'estructura del pèptid dissenyat té un esquelet tot-D. Pels tres residus on s'introdueix la variabilitat (X), proposem seleccions d'aminoàcids diferents per tal de donar un espectre més ampli pel que fa a les propietats fisicoquímiques de la biblioteca, cosa que facilitaria la separació per cromatografia (acoblada a l'espectrometre de masses). Així proposem, X_1: D-Arg, D-Glu, D-Ser, D-Phe, D-Asn i D-Ile; X_2: D-Lys, D-Asp, D-Ala, D-Tyr, D-Gln i D-Leu; X_3: D-His, D-Thr, D-Trp, D-Pro, D-Val i D-Gly.

Validació i quantificació

La validació i la quantificació del transport dels pèptids identificats es pot fer en base a diverses aproximacions. No obstant, n'hi ha una que permet la validació i la quantificació alhora. Es tracta de la síntesi de les versions amb isòtops pesants dels pèptids identificats. La coinjecció de la mostra assajada (ja sigui l'assaig de la biblioteca completa o un assaig amb el pèptid resintetitzat) i el patró intern sintetitzat permetria poder validar que les dues espècies coeluïssin en SRM, i que les intensitats de les transicions monitoritzades fossin proporcionals, i doncs validar que corresponen a la mateixa espècie. A més, la quantificació del transport es podria determinar amb la comparació de qualsevol de les transicions sempre i quan se sàpiga la quantitat de patró injectada i la concentració inicial del pèptid assajat.

Appendices

Appendix 1. *Bibliopepfinder* script (by Xavier Arroyo).

```
#!/usr/bin/perl -w
use strict;
use Data::Dumper;
#=head
my %aah = (
        "A" => 71.037110,
        "E" => 129.04259,
        "I" => 113.08406,
        "P" => 97.052760,
        "R" => 156.10110,
        "S" => 87.032030,
        "W" => 186.07931
);
#=cut
# ALTERNATIVE AA
=head
my %aah = (
        "V" => 117.1469,
        "H" => 155.1552,
        "Q" => 146.1451,
        "N" => 132.1184,
        "T" => 119.1197,
        "G" => 75.0669,
        "R" => 156.10110,
        "F" => 165.1900
);
=cut
open IN, shift or die $!;
my $intensitat = 100;
my @aa = qw(A E I P R S W);
my @pes= qw(71.03711 129.04259 113.08406 97.052760 156.1011 87.032030 186.07931);
my $suma;
my @a;
my @vector;
my $mida = 6;
my $ltotal = 5;
my $ntotal = 5;
my $pesBase = 43.01837 + 16.01871;
my $targetPes;
my $precisio = 10;
my $th = shift || 0.012; #0.105; #0.012;


my %resultat;
my $i=0;
my @finalOut;
my %finalOut2;
print "FAMILY\tFORMULA  \tM-THEORIC\tM-EXP    \tM/Z    \tZ    \tINT\n";
while (<IN>) {
        chomp;
        undef %resultat;
        my @arr = split (",",$_);
        if ($arr[3]>0 and $arr[1]>$intensitat) {
                $th = $arr[0]/100000;
```

```
                $targetPes = ($arr[0] * $arr[3])-(1.00782 * $arr[3]);
                &bibliopepfinder();
                foreach my $k (sort keys %resultat) {
                        $finalOut2{$k}{mteoric} = $resultat{$k};
                        $finalOut2{$k}{mexp} = $targetPes;
                        $finalOut2{$k}{mz} = $arr[0];
                        $finalOut2{$k}{z} = $arr[3];
                        $finalOut2{$k}{intensity} = $arr[1];

        push(@finalOut,"$k\t$resultat{$k}\t$targetPes\t$arr[0]\t$arr[3]\t$arr[1]\n");# if ($k=~/R/);
                }
        }
}

foreach my $k (
        map { $_->[0] }
        sort {
        $b->[0] cmp $a->[0]
                        ||
        $b->[1] <=> $a->[1]
                        ||
        $b->[2] <=> $a->[2]
        } map { [$_, $finalOut2{$_}{z}, $finalOut2{$_}{intensity}, $finalOut2{$_}{mz} ]}
keys %finalOut2)
{
        print
"$k\t$finalOut2{$k}{mteoric}\t$finalOut2{$k}{mexp}\t$finalOut2{$k}{mz}\t$finalOut2{$k}{z}\t$finalOu
t2{$k}{intensity}\n";
}


#
#  FUNCIONS
######################

sub generaPatrons {
        my $level = shift || 0;
        if ($#a>$ltotal) {
                $suma = 0;
                $suma += $_ foreach (@a);
                print @a,"\n" if ($suma == $ntotal);
        }
        else {
                for (my $i=0; $i<$ntotal; $i++) {
                        unshift @a, $i;
                        &generaPatrons($level+1);
                        shift @a;
                }
        }
}

sub patro {
        my $s=shift;
        my %f;
        my @arr = split("",$s);
        foreach (@arr) {
```

```
		$f{$_}=0 if (not defined $f{$_});
		$f{$_}++;
	}
	my $formula = "";
	my $familia = "";
	foreach (sort {$f{$b} <=> $f{$a}} sort keys %f) {
		$formula .= $_.$f{$_}."-";
		$familia .= $f{$_} if ($f{$_}!=0);
	}
	chop $formula;
	return "$familia\t$formula";
}

sub bibliopepfinder {
	my $level = shift || 0;
	if ($level == $mida) {
		my $suma = 0;
		$suma += $aah{$_} foreach(@vector);
		if (abs($suma - $targetPes + $pesBase) <= $th) {
			my $cmp = &patro(join("",@vector));
			$resultat{$cmp} = $suma + $pesBase;
		}
	}
	else {
		foreach(keys %aah) {
			unshift @vector, $_;
			&bibliopepfinder($level+1);
			shift @vector;
		}
	}
}

sub bibliopepfinder {
	my @resultat;
	my $level = shift || 0;
	if ($level == $mida) {
		$suma = 0;
		$suma += $pes[$_] foreach(@vector);
		if (abs($suma-$targetPes+$pesBase)<=$th) {
			my $cmp = &patro(join("",@vector));
			return $cmp; #$suma+$pesBase;
		}
	}
	else {
		foreach(keys %aah) {
		for(my $i=0; $i<=$#aa; $i++) {
			unshift @vector, $i;
			push (@resultat,&bibliopepfinder($level+1)) if ($level<$mida);
			shift @vector;
		}
		return @resultat;
	}
}
```

Appendix 2. Ac-D-Arg-library families found in PAMPA assay.

FAMILY	FORMULA	M-THEORIC	M-EXP	M/Z	Z
3111	P3-I1-R1-S1	706.41255	706.40928	354.21246	2
3111	I3-A1-R1-S1	712.4595	712.45928	357.23746	2
2211	A2-R2-I1-P1	723.45032	723.44992	362.73278	2
3111	A3-I1-R1-W1	727.41288	727.41442	364.71503	2
21111	I2-A1-E1-P1-R1	738.43876	738.43572	370.22568	2
2211	I2-S2-E1-R1	744.41295	744.41388	373.21476	2
321	S3-R2-I1	745.41943	745.41816	373.7169	2
3111	P3-E1-I1-R1	748.42311	748.4253	375.22047	2
21111	A2-I1-P1-R1-W1	753.42853	753.42958	377.72261	2
21111	I2-E1-P1-R1-S1	754.43368	754.43492	378.22528	2
411	I4-R1-S1	754.50645	754.50736	378.2615	2
321	R3-A2-S1	756.44663	756.45026	379.23295	2
3111	P3-A1-R1-W1	763.41288	763.4143	382.71497	2
2211	I2-R2-A1-P1	765.49727	765.4969	383.75627	2
321	P3-R2-I1	775.48162	775.48256	388.7491	2
321	R3-A2-I1	782.49866	782.49592	392.25578	2
21111	S2-I1-P1-R1-W1	785.41837	785.41882	393.71723	2
21111	A2-E1-I1-R1-W1	785.41836	785.41882	393.71723	2
2211	R2-S2-E1-I1	787.42999	787.43352	394.72458	2
222	I2-P2-R2	791.51292	791.51404	396.76484	2
21111	P2-A1-E1-R1-W1	795.40271	795.40348	398.70956	2
21111	I2-A1-P1-R1-W1	795.47548	795.4756	398.74562	2
2211	I2-R2-A1-E1	797.4871	797.48396	399.7498	2
3111	R3-A1-I1-S1	798.49358	798.49266	400.25415	2
111111	A1-E1-I1-R1-S1-W1	801.41328	801.41384	401.71474	2
2211	R2-S2-A1-W1	802.41976	802.41664	402.21614	2
3111	P3-I1-R1-W1	805.45983	805.46072	403.73818	2
2211	P2-R2-E1-I1	807.47145	807.47464	404.74514	2
321	R3-P2-S1	808.47793	808.48184	405.24874	2
21111	P2-E1-R1-S1-W1	811.39763	811.39864	406.70714	2
111111	A1-E1-I1-P1-R1-W1	811.43401	811.43426	406.72495	2
21111	R2-A1-P1-S1-W1	812.44049	812.43768	407.22666	2
3111	P3-E1-R1-W1	821.41836	821.41874	411.71719	2
2211	I2-P2-R1-W1	821.49113	821.491	411.75332	2
222	E2-P2-R2	823.42998	823.42976	412.7227	2
2211	I2-R2-E1-P1	823.50275	823.50296	412.7593	2
3111	R3-I1-P1-S1	824.50923	824.5055	413.26057	2
111111	E1-I1-P1-R1-S1-W1	827.42893	827.42916	414.7224	2
21111	I2-A1-E1-R1-W1	827.46531	827.46482	414.74023	2
321	I3-R2-E1	839.53405	839.53304	420.77434	2
3111	R3-A1-E1-I1	840.50414	840.50314	421.25939	2
321	R3-I2-S1	840.54053	840.54024	421.27794	2
21111	W2-A1-P1-R1-S1	842.4187	842.41864	422.21714	2
21111	I2-E1-R1-S1-W1	843.46023	843.46012	422.73788	2
321	P3-R2-W1	848.47687	848.47674	425.24619	2
321	R3-P2-E1	850.48849	850.485	426.25032	2
21111	R2-I1-P1-S1-W1	854.48744	854.4842	428.24992	2
3111	E3-A1-R1-W1	859.38237	859.38248	430.69906	2
21111	E2-I1-R1-S1-W1	859.41876	859.41814	430.71689	2
2211	P2-R2-I1-W1	864.50817	864.50644	433.26104	2
33	I3-R3	866.59256	866.59332	434.30448	2
2211	P2-W2-R1-S1	868.43435	868.43184	435.22374	2
21111	W2-A1-I1-P1-R1	868.47073	868.4707	435.24317	2
21111	E2-I1-P1-R1-W1	869.43949	869.43906	435.72735	2
3111	I3-E1-R1-W1	869.51226	869.50822	435.76193	2
21111	R2-E1-P1-S1-W1	870.44597	870.44236	436.229	2

2211	I2-R2-S1-W1	870.51874	870.51816	436.2669	2
3111	R3-A1-S1-W1	871.48883	871.49118	436.75341	2
21111	W2-A1-E1-R1-S1	874.40853	874.40856	438.2121	2
2211	P2-R2-E1-W1	880.4667	880.46576	441.2407	2
2211	I2-R2-P1-W1	880.53947	880.53822	441.27693	2
321	R3-E2-P1	882.47832	882.47434	442.24499	2
21111	R2-E1-I1-S1-W1	886.47727	886.47584	444.24574	2
21111	R2-E1-I1-P1-W1	896.498	896.49618	449.25591	2
21111	W2-E1-I1-P1-R1	926.47621	926.4777	464.24667	2
2211	E2-R2-I1-W1	928.48783	928.4852	465.25042	2
2211	E2-W2-R1-S1	932.41401	932.4158	467.21572	2
51	A5-R1	570.32373	570.32024	571.32806	1
411	A4-R1-S1	586.31865	586.3188	587.32662	1
321	A3-S2-R1	602.31357	602.31526	603.32308	1
3111	A3-P1-R1-S1	612.3343	612.33559	613.34341	1
411	A4-I1-R1	612.37068	612.37501	613.38283	1
321	S3-A2-R1	618.30849	618.30788	619.3157	1
321	A3-P2-R1	622.35503	622.35551	623.36333	1
3111	A3-I1-R1-S1	628.3656	628.3681	629.37592	1
3111	S3-A1-P1-R1	644.32414	644.32558	645.3334	1
3111	A3-E1-R1-S1	644.32413	644.32558	645.3334	1
321	P3-A2-R1	648.37068	648.37052	649.37834	1
3111	A3-E1-P1-R1	654.34486	654.34629	655.35411	1
2211	P2-S2-A1-R1	654.34487	654.34629	655.35411	1
321	A3-I2-R1	654.41763	654.42022	655.42804	1
411	S4-P1-R1	660.31906	660.31432	661.32214	1
2211	A2-S2-E1-R1	660.31905	660.31432	661.32214	1
3111	S3-A1-I1-R1	660.35544	660.35783	661.36565	1
2211	A2-P2-I1-R1	664.40198	664.40666	665.41448	1
321	S3-P2-R1	670.33979	670.34054	671.34836	1
21111	A2-E1-P1-R1-S1	670.33978	670.34054	671.34836	1
3111	A3-E1-I1-R1	670.37616	670.37764	671.38546	1
21111	S2-A1-I1-P1-R1	670.37617	670.37764	671.38546	1
321	P3-S2-R1	680.36052	680.362	681.36982	1
2211	A2-P2-E1-R1	680.36051	680.362	681.36982	1
21111	P2-A1-I1-R1-S1	680.3969	680.39311	681.40093	1
321	A3-R2-P1	681.40337	681.40563	682.41345	1
411	A4-R1-W1	685.36593	685.36402	686.37184	1
321	A3-E2-R1	686.33469	686.34045	687.34827	1
21111	S2-A1-E1-P1-R1	686.3347	686.34045	687.34827	1
3111	S3-I1-P1-R1	686.37109	686.37106	687.37888	1
21111	A2-E1-I1-R1-S1	686.37108	686.37106	687.37888	1
411	P4-R1-S1	690.38125	690.38314	691.39096	1
2211	P2-S2-I1-R1	696.39182	696.39307	697.40089	1
21111	A2-E1-I1-P1-R1	696.39181	696.39307	697.40089	1
21111	I2-A1-P1-R1-S1	696.4282	696.4257	697.43352	1
2211	A2-R2-P1-S1	697.39829	697.39646	698.40428	1
321	A3-R2-I1	697.43467	697.43233	698.44015	1
51	P5-R1	700.40198	700.4009	701.40872	1
3111	A3-R1-S1-W1	701.36085	701.36247	702.37029	1
3111	S3-E1-P1-R1	702.32962	702.33078	703.3386	1
2211	A2-E2-R1-S1	702.32961	702.33078	703.3386	1
21111	S2-A1-E1-I1-R1	702.366	702.36629	703.37411	1
321	S3-R2-A1	703.37248	703.36637	704.37419	1
2211	I2-P2-A1-R1	706.44893	706.44514	707.45296	1
2211	P2-S2-E1-R1	712.35035	712.35103	713.35885	1
2211	A2-E2-P1-R1	712.35034	712.35103	713.35885	1
111111	A1-E1-I1-P1-R1-S1	712.38673	712.38793	713.39575	1
2211	I2-S2-P1-R1	712.42312	712.42449	713.43231	1

2211	A2-I2-E1-R1	712.42311	712.42449	713.43231	1
321	A3-R2-E1	713.3932	713.38914	714.39696	1
2211	R2-S2-A1-P1	713.39321	713.38914	714.39696	1
2211	A2-R2-I1-S1	713.42959	713.42816	714.43598	1
411	P4-I1-R1	716.43328	716.43	717.43782	1
2211	A2-S2-R1-W1	717.35577	717.35412	718.36194	1
2211	E2-S2-A1-R1	718.32453	718.32502	719.33284	1
3111	P3-E1-R1-S1	722.37108	722.37301	723.38083	1
21111	P2-A1-E1-I1-R1	722.40746	722.40885	723.41667	1
2211	I2-P2-R1-S1	722.44385	722.44027	723.44809	1
2211	P2-R2-A1-S1	723.41394	723.41469	724.42251	1
21111	E2-A1-P1-R1-S1	728.34526	728.35117	729.35899	1
2211	A2-E2-I1-R1	728.38164	728.38254	729.39036	1
21111	S2-E1-I1-P1-R1	728.38165	728.38254	729.39036	1
21111	I2-A1-E1-R1-S1	728.41803	728.41942	729.42724	1
321	S3-R2-P1	729.38813	729.38433	730.39215	1
2211	A2-R2-E1-S1	729.38812	729.38433	730.39215	1
2211	R2-S2-A1-I1	729.42451	729.42369	730.43151	1
411	P4-E1-R1	732.39181	732.39337	733.40119	1
321	P3-I2-R1	732.46458	732.46175	733.46957	1
2211	A2-P2-R1-W1	737.39723	737.40011	738.40793	1
21111	P2-E1-I1-R1-S1	738.40238	738.40355	739.41137	1
3111	I3-P1-R1-S1	738.47515	738.47199	739.47981	1
222	P2-R2-S2	739.40886	739.40812	740.41594	1
2211	A2-R2-E1-P1	739.40885	739.40812	740.41594	1
21111	R2-A1-I1-P1-S1	739.44524	739.44467	740.45249	1
33	A3-R3	740.45171	740.44956	741.45738	1
321	I3-P2-R1	748.49588	748.4938	749.50162	1
321	P3-R2-S1	749.42959	749.42755	750.43537	1
21111	P2-A1-R1-S1-W1	753.39215	753.38998	754.3978	1
21111	E2-A1-I1-P1-R1	754.39729	754.39803	755.40585	1
21111	R2-A1-E1-P1-S1	755.40377	755.40137	756.40919	1
2211	I2-P2-E1-R1	764.45441	764.45631	765.46413	1
2211	P2-R2-A1-E1	765.4245	765.43169	766.43951	1
2211	P2-S2-R1-W1	769.38707	769.3893	770.39712	1
21111	A2-E1-P1-R1-W1	769.38706	769.3893	770.39712	1
111111	A1-I1-P1-R1-S1-W1	769.42345	769.42002	770.42784	1
21111	E2-I1-P1-R1-S1	770.39221	770.39257	771.40039	1
321	A3-R2-W1	770.42992	770.43108	771.4389	1
2211	E2-I2-A1-R1	770.42859	770.43108	771.4389	1
21111	R2-A1-E1-I1-S1	771.43507	771.43424	772.44206	1
321	R3-S2-A1	772.44155	772.43712	773.44494	1
3111	S3-I1-R1-W1	775.39764	775.40053	776.40835	1
3111	P3-R1-S1-W1	779.4078	779.40533	780.41315	1
21111	P2-A1-I1-R1-W1	779.44418	779.44163	780.44945	1
3111	I3-E1-P1-R1	780.48571	780.48217	781.48999	1
2211	P2-R2-E1-S1	781.41942	781.42709	782.43491	1
2211	I2-R2-P1-S1	781.49219	781.48474	782.49256	1
111111	A1-E1-P1-R1-S1-W1	785.38198	785.38374	786.39156	1
411	P4-R1-W1	789.42853	789.42467	790.43249	1
21111	P2-I1-R1-S1-W1	795.4391	795.43572	796.44354	1
2211	E2-I2-P1-R1	796.44424	796.44027	797.44809	1
2211	A2-R2-P1-W1	796.44557	796.44027	797.44809	1
321	R3-S2-P1	798.4572	798.46246	799.47028	1
321	R3-A2-E1	798.45719	798.46246	799.47028	1
21111	I2-P1-R1-S1-W1	811.4704	811.46675	812.47457	1
3111	E3-I1-P1-R1	812.40277	812.40263	813.41045	1
2211	I2-R2-E1-S1	813.48202	813.47645	814.48427	1
321	R3-S2-I1	814.4885	814.48166	815.48948	1

21111	S2-E1-I1-R1-W1	817.4082	817.40839	818.41621	1
321	R3-I2-A1	824.54561	824.53756	825.54538	1
21111	E2-A1-P1-R1-W1	827.39254	827.40044	828.40826	1
3111	I3-R1-S1-W1	827.5017	827.50533	828.51315	1
321	R3-S2-E1	830.44703	830.45315	831.46097	1
21111	P2-E1-I1-R1-W1	837.44966	837.44607	838.45389	1
21111	E2-A1-I1-R1-W1	843.42384	843.42476	844.43258	1
21111	R2-A1-E1-S1-W1	844.43032	844.43542	845.44324	1
21111	I2-E1-P1-R1-W1	853.48096	853.47789	854.48571	1
321	R3-E2-A1	856.46267	856.46682	857.47464	1
21111	R2-A1-E1-I1-W1	870.48235	870.47372	871.48154	1
42	E4-R2	887.40964	887.4151	888.42292	1
3111	W3-I1-P1-R1	983.51293	983.5092	984.51702	1

Appendix 3. Ac-D-Arg-library families found in cell-based *in vitro* BBB model.

FAMILY	FORMULA	M-THEORIC	M-EXP	M/Z	Z
321	S3-I2-R1	702.40239	702.40136	352.2085	2
3111	P3-I1-R1-S1	706.41255	706.41496	354.2153	2
2211	I2-P2-A1-R1	706.44893	706.45094	354.23329	2
222	A2-P2-R2	707.41902	707.4189	354.71727	2
2211	I2-S2-P1-R1	712.42312	712.42444	357.22004	2
2211	A2-I2-E1-R1	712.42311	712.42444	357.22004	2
3111	I3-A1-R1-S1	712.4595	712.45826	357.23695	2
2211	A2-R2-I1-S1	713.42959	713.42818	357.72191	2
2211	A2-S2-R1-W1	717.35577	717.35464	359.68514	2
21111	P2-A1-E1-I1-R1	722.40746	722.41006	362.21285	2
2211	P2-R2-A1-S1	723.41394	723.41296	362.7143	2
3111	A3-I1-R1-W1	727.41288	727.41488	364.71526	2
21111	E2-A1-P1-R1-S1	728.34526	728.34172	365.17868	2
321	I3-S2-R1	728.45442	728.45286	365.23425	2
2211	A2-P2-R1-W1	737.39723	737.39918	369.70741	2
3111	I3-P1-R1-S1	738.47515	738.47728	370.24646	2
222	A2-I2-R2	739.48162	739.4787	370.74717	2
2211	R2-S2-A1-E1	745.38304	745.38596	373.7008	2
321	P3-R2-S1	749.42959	749.42716	375.7214	2
2211	P2-R2-A1-I1	749.46597	749.46884	375.74224	2
21111	A2-I1-P1-R1-W1	753.42853	753.42972	377.72268	2
21111	E2-A1-I1-P1-R1	754.39729	754.39878	378.20721	2
21111	I2-E1-P1-R1-S1	754.43368	754.4347	378.22517	2
411	I4-R1-S1	754.50645	754.50666	378.26115	2
321	R3-A2-S1	756.44663	756.44992	379.23278	2
21111	S2-A1-I1-R1-W1	759.40272	759.4043	380.70997	2
42	P4-R2	759.45032	759.45272	380.73418	2
2211	P2-R2-A1-E1	765.4245	765.42544	383.72054	2
2211	I2-R2-A1-P1	765.49727	765.49836	383.757	2
2211	P2-S2-R1-W1	769.38707	769.3885	385.70207	2
21111	A2-E1-P1-R1-W1	769.38706	769.3885	385.70207	2
2211	A2-I2-R1-W1	769.45983	769.45852	385.73708	2
21111	E2-I1-P1-R1-S1	770.39221	770.39254	386.20409	2
321	A3-R2-W1	770.42992	770.42968	386.22266	2
2211	E2-I2-A1-R1	770.42859	770.42968	386.22266	2
3111	I3-E1-R1-S1	770.46498	770.464	386.23982	2
21111	R2-A1-E1-I1-S1	771.43507	771.4341	386.72487	2
222	I2-R2-S2	771.47146	771.46808	386.74186	2
321	R3-S2-A1	772.44155	772.4446	387.23012	2
321	P3-R2-I1	775.48162	775.48292	388.74928	2
321	E3-S2-R1	776.33001	776.32816	389.1719	2

3111	P3-R1-S1-W1	779.4078	779.40822	390.71193	2
21111	P2-A1-I1-R1-W1	779.44418	779.44516	390.7304	2
2211	E2-P2-I1-R1	780.41294	780.4135	391.21457	2
51	I5-R1	780.55848	780.55504	391.28534	2
2211	P2-R2-E1-S1	781.41942	781.41978	391.71771	2
321	I3-R2-A1	781.52857	781.52822	391.77193	2
3111	R3-A1-P1-S1	782.46228	782.46402	392.23983	2
21111	I2-A1-R1-S1-W1	785.45475	785.45482	393.73523	2
33	R3-S3	788.43647	788.43476	395.2252	2
321	P3-R2-E1	791.44015	791.4404	396.72802	2
222	I2-P2-R2	791.51292	791.51378	396.76471	2
21111	P2-A1-E1-R1-W1	795.40271	795.40352	398.70958	2
21111	P2-I1-R1-S1-W1	795.4391	795.439	398.72732	2
21111	I2-A1-P1-R1-W1	795.47548	795.47608	398.74586	2
2211	I2-R2-A1-E1	797.4871	797.4904	399.75302	2
321	I3-R2-S1	797.52349	797.52362	399.76963	2
321	R3-S2-P1	798.4572	798.45854	400.23709	2
321	R3-A2-E1	798.45719	798.45854	400.23709	2
3111	R3-A1-I1-S1	798.49358	798.49408	400.25486	2
321	A3-W2-R1	800.40813	800.40834	401.21199	2
2211	A2-E2-R1-W1	801.37689	801.37906	401.69735	2
21111	S2-E1-P1-R1-W1	801.3769	801.37906	401.69735	2
111111	A1-E1-I1-R1-S1-W1	801.41328	801.41428	401.71496	2
2211	I2-S2-R1-W1	801.44967	801.44908	401.73236	2
2211	P2-R2-E1-I1	807.47145	807.47548	404.74556	2
321	I3-R2-P1	807.54422	807.5449	404.78027	2
321	R3-P2-S1	808.47793	808.47914	405.24739	2
21111	P2-E1-R1-S1-W1	811.39763	811.39856	406.7071	2
111111	A1-E1-I1-P1-R1-W1	811.43401	811.4345	406.72507	2
21111	I2-P1-R1-S1-W1	811.4704	811.47066	406.74315	2
3111	I3-A1-R1-W1	811.50678	811.5057	406.76067	2
3111	E3-I1-P1-R1	812.40277	812.40226	407.20895	2
321	I3-E2-R1	812.47554	812.47564	407.24564	2
2211	A2-R2-I1-W1	812.47687	812.47564	407.24564	2
2211	I2-R2-E1-S1	813.48202	813.4799	407.74777	2
321	R3-S2-I1	814.4885	814.4889	408.25227	2
21111	S2-E1-I1-R1-W1	817.4082	817.4094	409.71252	2
321	S3-R2-W1	818.41468	818.41196	410.2138	2
2211	I2-P2-R1-W1	821.49113	821.49166	411.75365	2
2211	P2-R2-A1-W1	822.46122	822.4622	412.23892	2
2211	I2-R2-E1-P1	823.50275	823.50326	412.75945	2
42	I4-R2	823.57552	823.57564	412.79564	2
3111	R3-I1-P1-S1	824.50923	824.50822	413.26193	2
42	R4-A2	825.5157	825.51316	413.7644	2
2211	A2-W2-P1-R1	826.42378	826.42764	414.22164	2
111111	E1-I1-P1-R1-S1-W1	827.42893	827.4291	414.72237	2
21111	I2-A1-E1-R1-W1	827.46531	827.465	414.74032	2
3111	I3-R1-S1-W1	827.5017	827.50528	414.76046	2
21111	R2-A1-I1-S1-W1	828.47179	828.47144	415.24354	2
321	R3-S2-E1	830.44703	830.44844	416.23204	2
321	R3-P2-I1	834.52996	834.5296	418.27262	2
21111	P2-E1-I1-R1-W1	837.44966	837.44996	419.7328	2
21111	R2-A1-I1-P1-W1	838.49252	838.49232	420.25398	2
321	I3-R2-E1	839.53405	839.53698	420.77631	2
321	R3-I2-S1	840.54053	840.54054	421.27809	2
2211	A2-W2-I1-R1	842.45508	842.45418	422.23491	2
21111	E2-A1-I1-R1-W1	843.42384	843.42742	422.72153	2
21111	I2-E1-R1-S1-W1	843.46023	843.45932	422.73748	2
2211	R2-S2-I1-W1	844.46671	844.46636	423.241	2

321	P3-R2-W1	848.47687	848.47862	425.24713	2
2211	P2-W2-A1-R1	852.43943	852.43854	427.22709	2
21111	I2-E1-P1-R1-W1	853.48096	853.48106	427.74835	2
21111	R2-A1-E1-P1-W1	854.45105	854.45102	428.23333	2
21111	R2-I1-P1-S1-W1	854.48744	854.48714	428.25139	2
2211	I2-R2-A1-W1	854.52382	854.52288	428.26926	2
321	R3-A2-W1	855.49391	855.49116	428.7534	2
222	E2-I2-R2	855.49258	855.49116	428.7534	2
21111	W2-A1-I1-R1-S1	858.45	858.4461	430.23087	2
21111	E2-I1-R1-S1-W1	859.41876	859.4184	430.71702	2
2211	P2-R2-I1-W1	864.50817	864.50768	433.26166	2
3111	R3-E1-I1-P1	866.51979	866.51926	434.26745	2
2211	P2-W2-R1-S1	868.43435	868.43372	435.22468	2
21111	W2-A1-I1-P1-R1	868.47073	868.4695	435.24257	2
21111	E2-I1-P1-R1-W1	869.43949	869.43896	435.7273	2
21111	R2-A1-E1-I1-W1	870.48235	870.48188	436.24876	2
2211	I2-R2-S1-W1	870.51874	870.5178	436.26672	2
3111	R3-A1-S1-W1	871.48883	871.48518	436.75041	2
321	P3-W2-R1	878.45508	878.45418	440.23491	2
2211	P2-R2-E1-W1	880.4667	880.46978	441.24271	2
2211	I2-R2-P1-W1	880.53947	880.53864	441.27714	2
411	R4-A1-E1	883.52118	883.5237	442.76967	2
2211	I2-W2-A1-R1	884.50203	884.50282	443.25923	2
3111	E3-P1-R1-W1	885.39802	885.39744	443.70654	2
222	A2-R2-W2	885.47212	885.46938	443.74251	2
2211	E2-I2-R1-W1	885.47079	885.46938	443.74251	2
2211	S2-W2-E1-R1	890.40345	890.40228	446.20896	2
21111	R2-E1-I1-P1-W1	896.498	896.49364	449.25464	2
321	I3-R2-W1	896.57077	896.56946	449.29255	2
3111	R3-P1-S1-W1	897.50448	897.504	449.75982	2
321	R3-E2-I1	898.50962	898.50752	450.26158	2
21111	W2-E1-P1-R1-S1	900.42418	900.42316	451.2194	2
21111	W2-A1-E1-I1-R1	900.46056	900.45918	451.23741	2
2211	I2-W2-R1-S1	900.49695	900.49476	451.2552	2
2211	R2-W2-A1-S1	901.46704	901.46602	451.74083	2
321	R3-P2-W1	907.52521	907.52256	454.7691	2
2211	P2-W2-E1-R1	910.44491	910.44438	456.23001	2
2211	R2-W2-A1-P1	911.48777	911.48602	456.75083	2
2211	I2-R2-E1-W1	912.5293	912.52884	457.27224	2
3111	R3-I1-S1-W1	913.53578	913.53208	457.77386	2
33	E3-R3	914.46815	914.46634	458.24099	2
21111	W2-E1-I1-R1-S1	916.45548	916.45738	459.23651	2
222	R2-S2-W2	917.46196	917.46592	459.74078	2
3111	R3-I1-P1-W1	923.55651	923.5596	462.78762	2
21111	W2-E1-I1-P1-R1	926.47621	926.47366	464.24465	2
321	I3-W2-R1	926.54898	926.54572	464.28068	2
2211	R2-W2-P1-S1	927.48269	927.4784	464.74702	2
222	P2-R2-W2	937.50342	937.502	469.75882	2
321	R3-I2-W1	939.58781	939.5899	470.80277	2
3111	W3-A1-P1-R1	941.46598	941.46922	471.74243	2
42	R4-E2	941.52666	941.53012	471.77288	2
2211	E2-W2-P1-R1	942.43474	942.43276	472.2242	2
2211	R2-W2-A1-E1	943.4776	943.48064	472.74814	2
2211	R2-W2-I1-S1	943.51399	943.51098	472.76331	2
2211	R2-W2-I1-P1	953.53472	953.53302	477.77433	2
411	R4-S1-W1	956.55282	956.5497	479.28267	2
3111	W3-P1-R1-S1	957.4609	957.45786	479.73675	2
3111	W3-A1-I1-R1	957.49728	957.49312	479.75438	2
2211	E2-W2-I1-R1	958.46604	958.46376	480.2397	2

411	R4-P1-W1	966.57355	966.57142	484.29353	2
2211	R2-W2-E1-P1	969.49325	969.4908	485.75322	2
222	I2-R2-W2	969.56602	969.56858	485.79211	2
3111	W3-A1-E1-R1	973.45581	973.45288	487.73426	2
3111	W3-I1-R1-S1	973.4922	973.49118	487.75341	2
3111	W3-E1-R1-S1	989.45073	989.44672	495.73118	2
321	R3-W2-P1	996.55176	996.54758	499.28161	2
411	R4-E1-W1	998.56338	998.55982	500.28773	2
3111	W3-E1-P1-R1	999.47146	999.46784	500.74174	2
321	W3-I2-R1	999.54423	999.54008	500.77786	2
321	W3-R2-A1	1000.51432	1000.5147	501.26517	2
222	E2-R2-W2	1001.48308	1001.48146	501.74855	2
321	W3-E2-R1	1031.46129	1031.46158	516.73861	2
321	W3-R2-I1	1042.56127	1042.56244	522.28904	2
411	W4-R1-S1	1046.48745	1046.49128	524.25346	2
42	R4-W2	1055.6001	1055.60116	528.8084	2
321	W3-R2-E1	1058.5198	1058.5227	530.26917	2
411	W4-I1-R1	1072.53948	1072.54454	537.28009	2
51	A5-R1	570.32373	570.32506	571.33288	1
51	W5-R1	1145.53473	1145.53628	573.77596	2
411	A4-R1-S1	586.31865	586.32059	587.32841	1
411	A4-P1-R1	596.33938	596.34435	597.35217	1
321	A3-S2-R1	602.31357	602.31719	603.32501	1
411	A4-I1-R1	612.37068	612.37638	613.3842	1
321	S3-A2-R1	618.30849	618.31093	619.31875	1
321	A3-P2-R1	622.35503	622.36093	623.36875	1
2211	A2-P2-R1-S1	638.34995	638.3546	639.36242	1
3111	A3-I1-P1-R1	638.38633	638.39155	639.39937	1
3111	S3-A1-P1-R1	644.32414	644.32798	645.3358	1
3111	A3-E1-R1-S1	644.32413	644.32798	645.3358	1
2211	A2-S2-I1-R1	644.36052	644.35623	645.36405	1
321	P3-A2-R1	648.37068	648.36564	649.37346	1
3111	A3-E1-P1-R1	654.34486	654.35116	655.35898	1
2211	P2-S2-A1-R1	654.34487	654.35116	655.35898	1
21111	A2-I1-P1-R1-S1	654.38125	654.3844	655.39222	1
321	A3-I2-R1	654.41763	654.41374	655.42156	1
42	A4-R2	655.38772	655.3903	656.39812	1
411	S4-P1-R1	660.31906	660.32364	661.33146	1
2211	A2-S2-E1-R1	660.31905	660.32364	661.33146	1
3111	S3-A1-I1-R1	660.35544	660.35796	661.36578	1
3111	P3-A1-R1-S1	664.3656	664.36182	665.36964	1
321	S3-P2-R1	670.33979	670.34509	671.35291	1
21111	A2-E1-P1-R1-S1	670.33978	670.34509	671.35291	1
3111	A3-E1-I1-R1	670.37616	670.37781	671.38563	1
21111	S2-A1-I1-P1-R1	670.37617	670.37781	671.38563	1
2211	A2-I2-R1-S1	670.41255	670.40885	671.41667	1
321	A3-R2-S1	671.38264	671.38362	672.39144	1
411	P4-A1-R1	674.38633	674.38139	675.38921	1
3111	S3-A1-E1-R1	676.31397	676.31683	677.32465	1
411	S4-I1-R1	676.35036	676.35434	677.36216	1
321	P3-S2-R1	680.36052	680.36467	681.37249	1
2211	A2-P2-E1-R1	680.36051	680.36467	681.37249	1
21111	P2-A1-I1-R1-S1	680.3969	680.40176	681.40958	1
2211	A2-I2-P1-R1	680.43328	680.4307	681.43852	1
321	A3-R2-P1	681.40337	681.40612	682.41394	1
411	A4-R1-W1	685.36593	685.36242	686.37024	1
321	A3-E2-R1	686.33469	686.33666	687.34448	1
21111	S2-A1-E1-P1-R1	686.3347	686.33666	687.34448	1
3111	S3-I1-P1-R1	686.37109	686.37495	687.38277	1

21111	A2-E1-I1-R1-S1	686.37108	686.37495	687.38277	1
2211	I2-S2-A1-R1	686.40747	686.40307	687.41089	1
222	A2-R2-S2	687.37756	687.37856	688.38638	1
3111	P3-A1-I1-R1	690.41763	690.42062	691.42844	1
21111	P2-A1-E1-R1-S1	696.35543	696.3597	697.36752	1
21111	I2-A1-P1-R1-S1	696.4282	696.42706	697.43488	1
321	A3-R2-I1	697.43467	697.43809	698.44591	1
51	P5-R1	700.40198	700.40354	701.41136	1
3111	A3-R1-S1-W1	701.36085	701.36305	702.37087	1
3111	S3-E1-P1-R1	702.32962	702.33123	703.33905	1
2211	A2-E2-R1-S1	702.32961	702.33123	703.33905	1
21111	S2-A1-E1-I1-R1	702.366	702.36726	703.37508	1
3111	P3-A1-E1-R1	706.37616	706.37278	707.3806	1
3111	A3-P1-R1-W1	711.38158	711.38843	712.39625	1
2211	P2-S2-E1-R1	712.35035	712.35507	713.36289	1
2211	A2-E2-P1-R1	712.35034	712.35507	713.36289	1
111111	A1-E1-I1-P1-R1-S1	712.38673	712.38724	713.39506	1
321	A3-R2-E1	713.3932	713.3947	714.40252	1
2211	R2-S2-A1-P1	713.39321	713.3947	714.40252	1
411	P4-I1-R1	716.43328	716.43109	717.43891	1
2211	E2-S2-A1-R1	718.32453	718.32683	719.33465	1
3111	S3-E1-I1-R1	718.36092	718.36605	719.37387	1
42	S4-R2	719.3674	719.37131	720.37913	1
3111	P3-E1-R1-S1	722.37108	722.37489	723.38271	1
2211	I2-P2-R1-S1	722.44385	722.44347	723.45129	1
3111	I3-A1-P1-R1	722.48023	722.47936	723.48718	1
2211	A2-R2-I1-P1	723.45032	723.44642	724.45424	1
21111	A2-P1-R1-S1-W1	727.3765	727.37717	728.38499	1
2211	A2-E2-I1-R1	728.38164	728.38249	729.39031	1
21111	S2-E1-I1-P1-R1	728.38165	728.38249	729.39031	1
21111	I2-A1-E1-R1-S1	728.41803	728.42143	729.42925	1
321	S3-R2-P1	729.38813	729.39	730.39782	1
2211	A2-R2-E1-S1	729.38812	729.39	730.39782	1
2211	R2-S2-A1-I1	729.42451	729.42571	730.43353	1
411	P4-E1-R1	732.39181	732.39583	733.40365	1
321	P3-I2-R1	732.46458	732.46431	733.47213	1
3111	S3-A1-R1-W1	733.35069	733.34378	734.3516	1
321	P3-R2-A1	733.43467	733.43894	734.44676	1
2211	E2-P2-A1-R1	738.36599	738.37023	739.37805	1
21111	P2-E1-I1-R1-S1	738.40238	738.40644	739.41426	1
21111	I2-A1-E1-P1-R1	738.43876	738.44022	739.44804	1
222	P2-R2-S2	739.40886	739.41479	740.42261	1
2211	A2-R2-E1-P1	739.40885	739.41479	740.42261	1
21111	R2-A1-I1-P1-S1	739.44524	739.43918	740.447	1
3111	A3-E1-R1-W1	743.37141	743.37574	744.38356	1
21111	S2-A1-P1-R1-W1	743.37142	743.37574	744.38356	1
21111	A2-I1-R1-S1-W1	743.4078	743.40643	744.41425	1
321	E3-A2-R1	744.34017	744.34565	745.35347	1
2211	E2-S2-P1-R1	744.34018	744.34565	745.35347	1
2211	I2-S2-E1-R1	744.41295	744.41627	745.42409	1
321	S3-R2-I1	745.41943	745.4202	746.42802	1
3111	P3-E1-I1-R1	748.42311	748.42696	749.43478	1
321	I3-P2-R1	748.49588	748.49577	749.50359	1
411	S4-R1-W1	749.34561	749.35247	750.36029	1
21111	P2-A1-R1-S1-W1	753.39215	753.39055	754.39837	1
2211	E2-P2-R1-S1	754.36091	754.36572	755.37354	1
3111	I3-A1-E1-R1	754.47006	754.47181	755.47963	1
21111	R2-A1-E1-P1-S1	755.40377	755.41102	756.41884	1
2211	R2-S2-I1-P1	755.44016	755.44454	756.45236	1

2211	A2-R2-E1-I1	755.44015	755.44454	756.45236	1
2211	I2-R2-A1-S1	755.47654	755.47516	756.48298	1
3111	S3-P1-R1-W1	759.36634	759.36643	760.37425	1
21111	A2-E1-R1-S1-W1	759.36633	759.36643	760.37425	1
2211	E2-S2-I1-R1	760.37148	760.37847	761.38629	1
3111	P3-A1-R1-W1	763.41288	763.4133	764.42112	1
2211	I2-P2-E1-R1	764.45441	764.45608	765.4639	1
411	I4-P1-R1	764.52718	764.5263	765.53412	1
111111	A1-I1-P1-R1-S1-W1	769.42345	769.42108	770.4289	1
222	A2-E2-R2	771.39868	771.40457	772.41239	1
2211	R2-S2-E1-P1	771.39869	771.40457	772.41239	1
21111	S2-A1-E1-R1-W1	775.36125	775.36753	776.37535	1
3111	S3-I1-R1-W1	775.39764	775.396	776.40382	1
3111	I3-E1-P1-R1	780.48571	780.48512	781.49294	1
21111	R2-A1-E1-I1-P1	781.4558	781.46255	782.47037	1
2211	I2-R2-P1-S1	781.49219	781.49128	782.4991	1
321	R3-A2-I1	782.49866	782.49432	783.50214	1
111111	A1-E1-P1-R1-S1-W1	785.38198	785.38255	786.39037	1
21111	S2-I1-P1-R1-W1	785.41837	785.41982	786.42764	1
21111	A2-E1-I1-R1-W1	785.41836	785.41982	786.42764	1
2211	E2-I2-R1-S1	786.42351	786.4232	787.43102	1
2211	A2-R2-S1-W1	786.42484	786.4232	787.43102	1
2211	R2-S2-E1-I1	787.42999	787.42597	788.43379	1
411	P4-R1-W1	789.42853	789.43152	790.43934	1
3111	S3-E1-R1-W1	791.35617	791.35981	792.36763	1
321	E3-P2-R1	796.37147	796.3759	797.38372	1
2211	E2-I2-P1-R1	796.44424	796.44529	797.45311	1
2211	A2-R2-P1-W1	796.44557	796.44529	797.45311	1
2211	E2-R2-A1-P1	797.41433	797.41202	798.41984	1
21111	R2-E1-I1-P1-S1	797.45072	797.44674	798.45456	1
2211	R2-S2-A1-W1	802.41976	802.41786	803.42568	1
3111	P3-I1-R1-W1	805.45983	805.4596	806.46742	1
21111	R2-A1-P1-S1-W1	812.44049	812.4387	813.44652	1
2211	E2-R2-A1-I1	813.44563	813.44185	814.44967	1
3111	R3-A1-E1-S1	814.45211	814.44554	815.45336	1
2211	A2-W2-R1-S1	816.40305	816.40714	817.41496	1
21111	E2-A1-R1-S1-W1	817.37181	817.37963	818.38745	1
33	P3-R3	818.49866	818.49201	819.49983	1
3111	P3-E1-R1-W1	821.41836	821.42002	822.42784	1
222	E2-P2-R2	823.42998	823.42487	824.43269	1
3111	R3-A1-E1-P1	824.47284	824.47615	825.48397	1
21111	E2-A1-P1-R1-W1	827.39254	827.39358	828.4014	1
321	E3-I2-R1	828.43407	828.43365	829.44147	1
2211	R2-S2-P1-W1	828.43541	828.43365	829.44147	1
2211	A2-R2-E1-W1	828.4354	828.43365	829.44147	1
2211	E2-R2-I1-S1	829.44055	829.43549	830.44331	1
2211	S2-W2-A1-R1	832.39797	832.3987	833.40652	1
3111	I3-P1-R1-W1	837.52243	837.51956	838.52738	1
2211	P2-R2-S1-W1	838.45614	838.45463	839.46245	1
2211	E2-R2-I1-P1	839.46128	839.45746	840.46528	1
3111	R3-E1-P1-S1	840.46776	840.46162	841.46944	1
3111	R3-A1-E1-I1	840.50414	840.51104	841.51886	1
411	R4-A1-S1	841.51062	841.50243	842.51025	1
21111	W2-A1-P1-R1-S1	842.4187	842.41675	843.42457	1
21111	R2-A1-E1-S1-W1	844.43032	844.42757	845.43539	1
321	E3-R2-S1	845.39908	845.3908	846.39862	1
321	S3-W2-R1	848.39289	848.39144	849.39926	1
321	R3-P2-E1	850.48849	850.48114	851.48896	1
2211	E2-P2-R1-W1	853.40819	853.409	854.41682	1

321	E3-R2-P1	855.41981	855.41585	856.42367	1
321	R3-E2-A1	856.46267	856.47032	857.47814	1
3111	R3-E1-I1-S1	856.49906	856.4925	857.50032	1
2211	S2-W2-P1-R1	858.41362	858.40784	859.41566	1
2211	A2-W2-E1-R1	858.41361	858.40784	859.41566	1
2211	R2-S2-E1-W1	860.42524	860.42122	861.42904	1
3111	I3-E1-R1-W1	869.51226	869.51006	870.51788	1
21111	R2-E1-P1-S1-W1	870.44597	870.45413	871.46195	1
321	E3-R2-I1	871.45111	871.45738	872.4652	1
321	R3-E2-S1	872.45759	872.4573	873.46512	1
21111	W2-A1-E1-R1-S1	874.40853	874.40886	875.41668	1
2211	S2-W2-I1-R1	874.44492	874.44456	875.45238	1
321	R3-E2-P1	882.47832	882.4739	883.48172	1
21111	W2-A1-E1-P1-R1	884.42926	884.42836	885.43618	1
21111	W2-I1-P1-R1-S1	884.46565	884.4642	885.47202	1
21111	R2-E1-I1-S1-W1	886.47727	886.47988	887.4877	1
42	E4-R2	887.40964	887.41279	888.42061	1
321	R3-S2-W1	887.48375	887.48599	888.49381	1
2211	P2-W2-I1-R1	894.48638	894.48326	895.49108	1
3111	E3-I1-R1-W1	901.42932	901.42949	902.43731	1
2211	E2-R2-S1-W1	902.4358	902.44387	903.45169	1
2211	I2-W2-P1-R1	910.51768	910.51509	911.52291	1
3111	R3-A1-E1-W1	913.49939	913.5072	914.51502	1
321	W3-A2-R1	915.45033	915.44999	916.45781	1
2211	E2-W2-A1-R1	916.41909	916.42711	917.43493	1
3111	W3-A1-R1-S1	931.44525	931.44092	932.44874	1
3111	R3-E1-P1-W1	939.51504	939.51056	940.51838	1
2211	I2-W2-E1-R1	942.50751	942.5035	943.51132	1
321	E3-R2-W1	944.44636	944.45375	945.46157	1
321	W3-S2-R1	947.44017	947.44166	948.44948	1
321	R3-E2-W1	971.50487	971.4983	972.50612	1

Appendix 4. Complete family list and information filtered on *passive diffusion transport list. Green (Acc1), Blue (Acc2) and Red (Acc3).*

anking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY
1	2211	**A2-E2-P1-R1**	713.36289	1	1902422.6	1	2211	**I2-P2-R1-S1**	723.45137	1	1046728.7	1	21111	**I2-A1-P1-R1-W1**	398.74522	2	1609213.4
2	2211	**P2-S2-E1-R1**	713.36289	1	1902422.6	2	321	**I3-P2-R1**	749.50354	1	573285.3	2	2211	**I2-R2-P1-W1**	441.27738	2	1476024.6
3	2211	**I2-P2-R1-S1**	723.45129	1	1107997.8	3	21111	**E2-I1-P1-R1-W1**	435.72719	2	534796.3	3	21111	**I2-A1-P1-R1-S1**	697.43514	1	1247774.9
4	21111	**I2-A1-P1-R1-S1**	697.43488	1	963155.1	4	21111	**A2-E1-I1-R1-S1**	687.38276	1	499937.8	4	3111	**P3-I1-R1-W1**	806.46751	1	1046600.8
5	411	**A4-R1-S1**	587.32841	1	896967.2	5	3111	**S3-I1-P1-R1**	687.38276	1	499937.8	5	2211	**A2-P2-E1-R1**	681.37203	1	1021441.3
6	321	**P3-I2-R1**	733.47213	1	718637.3	6	2211	**A2-R2-E1-S1**	730.39767	1	469672.2	6	321	**P3-S2-R1**	681.37203	1	1021441.3
7	3111	**P3-I1-R1-W1**	806.46742	1	672709.1	7	321	**S3-R2-P1**	730.39767	1	469672.2	7	21111	**E2-I1-P1-R1-W1**	435.72734	2	986238.3
8	3111	**P3-A1-R1-W1**	764.42112	1	623944.8	8	3111	**I3-E1-P1-R1**	781.49314	1	425184.7	8	321	**I3-P2-R1**	749.50354	1	914690.4
9	321	**I3-P2-R1**	749.50359	1	611096.3	9	321	**A3-R2-E1**	714.4024	1	394233.5	9	3111	**I3-P1-R1-S1**	739.47774	1	829773.9
10	3111	**A3-E1-R1-S1**	645.3358	1	587510.6	10	2211	**R2-S2-A1-P1**	714.4024	1	394233.5	10	321	**A3-R2-P1**	682.41365	1	745850.7
11	3111	**S3-A1-P1-R1**	645.3358	1	587510.6	11	2211	**A2-P2-E1-R1**	681.37234	1	391231.6	11	2211	**A2-R2-E1-S1**	730.39791	1	745617.9
12	2211	**I2-R2-P1-W1**	441.27714	2	527168.8	12	321	**P3-S2-R1**	681.37234	1	391231.6	12	321	**S3-R2-P1**	730.39791	1	745617.9
13	2211	**I2-R2-S1-W1**	436.26672	2	509454.2	13	21111	**I2-A1-E1-R1-S1**	729.42896	1	356928.3	13	321	**A3-R2-E1**	714.40269	1	723142.4
14	21111	**E2-I1-P1-R1-W1**	435.7273	2	506436.4	14	21111	**A2-E1-I1-R1-W1**	786.42773	1	330281.1	14	2211	**R2-S2-A1-P1**	714.40269	1	723142.4
15	2211	**A2-R2-E1-S1**	730.39782	1	459916.4	15	21111	**S2-I1-P1-R1-W1**	786.42773	1	330281.1	15	21111	**W2-A1-I1-P1-R1**	869.47809	1	671754.1
16	321	**S3-R2-P1**	730.39782	1	459916.4	16	321	**R3-A2-S1**	379.23164	2	327893.3	16	21111	**I2-A1-E1-R1-S1**	729.42918	1	623426.6
17	21111	**A2-E1-I1-R1-S1**	687.38277	1	447375.3	17	321	**A3-R2-P1**	682.41375	1	309023.8	17	21111	**A2-E1-I1-R1-W1**	786.4279	1	617016.4
18	3111	**S3-I1-P1-R1**	687.38277	1	447375.3	18	2211	**A2-R2-I1-P1**	724.45431	1	302026.3	18	21111	**S2-I1-P1-R1-W1**	786.4279	1	617016.4
19	21111	**P2-I1-R1-S1-W1**	398.72732	2	435040.8	19	321	**A3-S2-R1**	603.32495	1	268850.2	19	21111	**P2-E1-I1-R1-S1**	739.41439	1	572105.6
20	3111	**I3-E1-P1-R1**	781.49294	1	430299.1	20	21111	**R2-E1-P1-S1-W1**	871.46191	1	241139.1	20	21111	**R2-E1-P1-S1-W1**	871.46232	1	506801.3
21	111111	**A1-I1-P1-R1-S1-W1**	770.4289	1	423203.1	21	321	**A3-R2-I1**	698.44661	1	203323.5	21	21111	**R2-A1-P1-S1-W1**	813.44638	1	429065.8
22	321	**A3-R2-E1**	714.40252	1	397672.3	22	21111	**R2-A1-P1-S1-W1**	813.44649	1	188642.4	22	3111	**R3-A1-I1-S1**	400.2552	2	351383.6
23	2211	**R2-S2-A1-P1**	714.40252	1	397672.3	23	3111	**R3-A1-I1-S1**	400.25481	2	178204.8	23	3111	**R3-A1-S1-W1**	436.75278	2	335362.6
24	21111	**I2-A1-E1-R1-S1**	729.42925	1	352833	24	2211	**I2-P2-E1-R1**	765.464	1	177861.8	24	321	**A3-R2-I1**	698.44896	1	332543.1
25	21111	**P2-E1-I1-R1-S1**	739.41426	1	348441.5	25	21111	**W2-A1-P1-R1-S1**	843.425	1	167684.8	25	321	**R3-A2-E1**	400.23718	2	325403.5
26	21111	**A2-E1-I1-R1-W1**	786.42764	1	336399.6	26	3111	**P3-E1-I1-R1**	749.43473	1	159513.6	26	321	**R3-S2-P1**	400.23718	2	325403.5
27	21111	**S2-I1-P1-R1-W1**	786.42764	1	336399.6	27	2211	**A2-R2-E1-P1**	740.42272	1	138882.7	27	21111	**W2-A1-P1-R1-S1**	843.42544	1	291223.3
28	2211	**A2-R2-I1-P1**	724.45424	1	327178.6	28	222	**P2-R2-S2**	740.42272	1	138882.7	28	321	**R3-A2-S1**	379.23172	2	281334.2
29	321	**A3-R2-P1**	682.41394	1	319622.1	29	321	**I3-R2-E1**	420.77629	2	138408.2	29	3111	**P3-E1-I1-R1**	749.43517	1	269755.8
30	2211	**A2-P2-E1-R1**	681.37249	1	319593.8	30	2211	**P2-W2-R1-S1**	869.44203	1	124171.8	30	21111	**R2-A1-E1-S1-W1**	845.43623	1	242065.6
31	321	**P3-S2-R1**	681.37249	1	319593.8	31	21111	R2-A1-E1-S1-W1	845.43629	1	121774.1	31	2211	**A2-R2-E1-P1**	740.42296	1	237888.8
32	2211	P2-R2-E1-W1	441.24271	2	313979.5	32	2211	E2-S2-A1-R1	719.33477	1	115176.8	32	222	**P2-R2-S2**	740.42296	1	237888.8
33	21111	**R2-E1-P1-S1-W1**	871.46195	1	304443	33	51	P5-R1	701.41151	1	114659.1	33	2211	**P2-W2-R1-S1**	869.4423	1	229703.4
34	321	**R3-A2-E1**	400.23709	2	275078.2	34	2211	R2-S2-A1-I1	730.43327	1	112566.8	34	2211	R2-S2-A1-I1	730.43452	1	208690.4
35	321	**R3-S2-P1**	400.23709	2	275078.2	35	3111	P3-R1-S1-W1	390.71193	2	108399.6	35	51	P5-R1	701.41339	1	184986.9
36	321	**A3-S2-R1**	603.32501	1	245828.1	36	321	**R3-A2-E1**	799.4726	1	62221.4	36	321	R3-P2-S1	405.24755	2	183442.2
37	21111	**R2-A1-P1-S1-W1**	813.44652	1	213268.9	37	321	**R3-S2-P1**	799.4726	1	62221.4	37	3111	**I3-E1-P1-R1**	781.49629	1	145063.1
38	321	**A3-R2-I1**	698.44591	1	211959.7	38	321	R3-S2-E1	416.22988	2	61252.4	38	2211	I2-S2-E1-R1	745.42283	1	138328.2
39	2211	P2-R2-E1-I1	404.74556	2	210387.7	39	3111	I3-A1-R1-S1	357.23734	2	58892.5	39	321	P3-R2-S1	750.4388	1	96929.3
40	2211	**I2-P2-E1-R1**	765.4639	1	183702.8	40	411	P4-E1-R1	733.40363	1	57861.8	40	3111	P3-E1-R1-W1	822.42811	1	94658.8
41	21111	E2-I1-R1-S1-W1	430.71702	2	180801.4	41	321	P3-R2-S1	750.43831	1	56504.3	41	321	R3-S2-E1	416.23215	2	92784.6

anking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY
42	21111	W2-A1-P1-R1-S1	843.42457	1	175786.7	42	2211	A2-R2-P1-W1	797.45327	1	56152.7	42	21111	R2-E1-I1-P1-W1	449.25613	2	91953.3
43	3111	**P3-E1-I1-R1**	749.43478	1	161181.3	43	2211	E2-I2-P1-R1	797.45327	1	56152.7	43	321	R3-S2-I1	408.25256	2	88110.6
44	3111	**R3-A1-I1-S1**	400.25486	2	160341.3	44	321	R3-I2-S1	421.27814	2	52714.9	44	2211	R2-S2-A1-W1	803.42642	1	79712
45	321	**I3-R2-E1**	420.77631	2	144231.2	45	3111	P3-E1-R1-W1	822.42782	1	52325.2	45	21111	E2-I1-R1-S1-W1	860.42723	1	68163.9
46	21111	**R2-A1-E1-S1-W1**	845.43539	1	128726.5	46	321	R3-E2-A1	857.47792	1	49823	46	2211	P2-R2-I1-W1	433.26165	2	65648.5
47	2211	**A2-R2-E1-P1**	740.42261	1	123696.2	47	2211	R2-S2-A1-W1	803.42593	1	43093	47	2211	I2-R2-P1-S1	782.49924	1	53854.3
48	222	**P2-R2-S2**	740.42261	1	123696.2	48	3111	I3-R1-S1-W1	414.75916	2	39574.4	48	411	**A4-R1-S1**	587.32796	1	48865
49	321	A3-E2-R1	687.34448	1	111953.8	49	2211	**A2-E2-P1-R1**	713.36353	1	38333.5	49	2211	R2-S2-E1-I1	788.4337	1	44368.8
50	21111	S2-A1-E1-P1-R1	687.34448	1	111953.8	50	2211	**P2-S2-E1-R1**	713.36353	1	38333.5	50	21111	R2-I1-P1-S1-W1	428.25133	2	41225.8
51	51	P5-R1	701.41136	1	111247.5	51	21111	R2-E1-I1-P1-W1	449.25479	2	36209.2	51	222	I2-P2-R2	396.76474	2	36818
52	2211	R2-S2-A1-I1	730.43353	1	108973.8	52	3111	A3-E1-P1-R1	655.35851	1	32118	52	21111	R2-E1-I1-S1-W1	887.48958	1	34145
53	2211	E2-S2-A1-R1	719.33465	1	104182.2	53	2211	P2-S2-A1-R1	655.35851	1	32118	53	2211	P2-R2-E1-W1	441.24001	2	32965
54	3111	P3-R1-S1-W1	390.71193	2	98039.8	54	21111	E2-I1-R1-S1-W1	860.42602	1	30844.8	54	2211	I2-R2-E1-P1	412.75947	2	31799.7
55	21111	W2-A1-E1-R1-S1	875.41668	1	86087.2	55	222	I2-P2-R2	396.76473	2	27524.7	55	21111	W2-A1-E1-R1-S1	875.41887	1	30538.8
56	321	R3-P2-S1	405.24739	2	77166.9	56	411	P4-R1-W1	790.43553	1	25853.3	56	21111	I2-A1-E1-P1-R1	739.44578	1	29518.5
57	2211	I2-S2-E1-R1	745.42409	1	71432.8	57	2211	R2-S2-E1-I1	788.43374	1	23924.2	57	21111	R2-A1-E1-I1-W1	436.24871	2	26692.9
58	3111	I3-A1-R1-S1	357.23695	2	66982.1	58	2211	A2-E2-R1-S1	703.34109	1	23252.1	58	411	A4-R1-W1	686.37241	1	26354.9
59	321	R3-E2-A1	857.47814	1	59562.9	59	3111	S3-E1-P1-R1	703.34109	1	23252.1	59	3111	E3-I1-P1-R1	813.41724	1	25706.3
60	321	R3-I2-S1	421.27809	2	57072.3	60	2211	I2-R2-E1-P1	412.75942	2	22607.4	60	2211	**A2-E2-P1-R1**	713.35882	1	23778.2
61	411	P4-E1-R1	733.40365	1	56688.4	61	321	S3-R2-I1	746.42762	1	22502.6	61	2211	**P2-S2-E1-R1**	713.35882	1	23778.2
62	321	R3-S2-E1	416.23204	2	55373	62	2211	P2-R2-E1-W1	441.24044	2	21972.3	62	21111	I2-E1-P1-R1-S1	378.22527	2	18653.3
63	3111	P3-E1-R1-W1	822.42784	1	54966.1	63	2211	I2-R2-P1-S1	782.4998	1	21031.5	63	2211	A2-I2-E1-R1	357.22017	2	17659.4
64	2211	R2-S2-A1-W1	803.42568	1	53515.4	64	21111	R2-A1-E1-I1-W1	436.24878	2	20761.1	64	2211	I2-S2-P1-R1	357.22017	2	17659.4
65	3111	I3-R1-S1-W1	414.76046	2	52525.1	65	21111	R2-I1-P1-S1-W1	428.2514	2	20238.9	65	3111	S3-A1-I1-R1	661.36623	1	17138.5
66	21111	I2-A1-E1-P1-R1	739.44804	1	50901	66	2211	P2-R2-I1-W1	433.26164	2	20014.6	66	2211	**I2-R2-S1-W1**	436.26536	2	16989.3
67	2211	A2-R2-P1-W1	797.45311	1	48795.7	67	321	A3-E2-R1	687.34203	1	16023.5	67	21111	P2-A1-E1-R1-W1	398.70967	2	16826.6
68	2211	E2-I2-P1-R1	797.45311	1	48795.7	68	21111	S2-A1-E1-P1-R1	687.34203	1	16023.5	68	2211	A2-S2-E1-R1	661.3315	1	16497.9
69	3111	E3-I1-P1-R1	407.20895	2	47584.7	69	21111	P2-A1-E1-I1-R1	362.21284	2	14300.1	69	411	S4-P1-R1	661.3315	1	16497.9
70	3111	A3-E1-P1-R1	655.35898	1	39781.1	70	411	**A4-R1-S1**	587.32957	1	14214.3	70	21111	P2-A1-E1-I1-R1	362.21287	2	16193.5
71	2211	P2-S2-A1-R1	655.35898	1	39781.1	71	411	A4-R1-W1	686.37091	1	12058.7	71	3111	A3-E1-P1-R1	655.35457	1	16012.9
72	321	R3-S2-I1	408.25227	2	33355.5	72	21111	R2-E1-I1-S1-W1	887.48458	1	11849	72	2211	P2-S2-A1-R1	655.35457	1	16012.9
73	21111	R2-E1-I1-P1-W1	449.25464	2	32201.7	73	3111	E3-I1-P1-R1	813.41647	1	11739.4	73	321	A3-E2-R1	687.34404	1	15807.9
74	2211	I2-R2-P1-S1	782.4991	1	29219.8	74	3111	S3-A1-I1-R1	661.3654	1	11620.6	74	21111	S2-A1-E1-P1-R1	687.34404	1	15807.9
75	222	I2-P2-R2	396.76471	2	26397.4	75	21111	I2-E1-P1-R1-W1	427.74999	2	10629.3	75	3111	**A3-E1-R1-S1**	645.33376	1	15787.3
76	2211	R2-S2-E1-I1	788.43379	1	25551.5	76	411	I4-R1-S1	378.26101	2	10394.1	76	3111	**S3-A1-P1-R1**	645.33376	1	15787.3
77	321	S3-R2-I1	746.42802	1	22813.5	77	21111	I2-A1-E1-P1-R1	370.22846	2	9763.9	77	2211	I2-P2-A1-R1	707.45568	1	15710.3
78	21111	R2-A1-I1-P1-S1	740.447	1	22462.8	78	21111	**I2-A1-P1-R1-S1**	697.43509	1	9722.8	78	2211	A2-E2-R1-S1	703.34101	1	15192.4
79	21111	A2-E1-P1-R1-S1	671.35291	1	21659.5	79	21111	I2-E1-P1-R1-S1	378.22511	2	7939.9	79	3111	S3-E1-P1-R1	703.34101	1	15192.4
80	321	S3-P2-R1	671.35291	1	21659.5	80	111111	A1-E1-I1-P1-R1-S1	713.39535	1	7852.6	80	321	P3-R2-W1	425.2462	2	15133
81	2211	I2-R2-E1-P1	412.75945	2	20531.2	81	3111	**R3-A1-S1-W1**	436.75041	2	7636.5	81	2211	**I2-P2-E1-R1**	765.46067	1	14522.6
82	21111	R2-A1-E1-I1-W1	436.24876	2	19428	82	2211	A2-S2-E1-R1	661.33041	1	7429.2	82	3111	I3-R1-S1-W1	414.75692	2	13916.6
83	2211	P2-R2-I1-W1	433.26166	2	18717.8	83	411	S4-P1-R1	661.33041	1	7429.2	83	2211	A2-R2-P1-W1	797.44961	1	12593
84	21111	R2-I1-P1-S1-W1	428.25139	2	18517.3	84	3111	P3-E1-R1-S1	723.38328	1	7293.4	84	2211	E2-I2-P1-R1	797.44961	1	12593
85	21111	R2-E1-I1-S1-W1	887.4877	1	18133.2	85	2211	**I2-R2-P1-W1**	441.27691	2	6900.1	85	3111	R3-A1-E1-I1	841.51812	1	12076.1

anking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY
86	21111	P2-A1-E1-I1-R1	362.21285	2	13810.1	86	3111	R3-I1-P1-S1	413.26193	2	6840.4	86	3111	A3-E1-I1-R1	671.38617	1	12074.4
87	411	A4-R1-W1	686.37024	1	11951.7	87	21111	A2-E1-P1-R1-W1	385.70202	2	6584.2	87	21111	S2-A1-I1-P1-R1	671.38617	1	12074.4
88	3111	S3-A1-I1-R1	661.36578	1	11023	88	2211	P2-S2-R1-W1	385.70202	2	6584.2	88	21111	**A2-E1-I1-R1-S1**	687.37278	1	11608.9
89	411	I4-R1-S1	378.26115	2	9950.6	89	2211	A2-I2-E1-R1	357.22006	2	6559.4	89	3111	**S3-I1-P1-R1**	687.37278	1	11608.9
90	2211	A2-S2-E1-R1	661.33146	1	9913.4	90	2211	I2-S2-P1-R1	357.22006	2	6559.4	90	111111	A1-E1-I1-P1-R1-S1	713.39574	1	11540.4
91	411	S4-P1-R1	661.33146	1	9913.4	91	21111	A2-E1-P1-R1-S1	671.34738	1	6335.6	91	2211	**I2-P2-R1-S1**	723.4498	1	10743.3
92	3111	P3-E1-R1-S1	723.38271	1	9237.8	92	321	S3-P2-R1	671.34738	1	6335.6	92	222	E2-P2-R2	824.43412	1	10603.8
93	321	P3-R2-S1	375.7214	2	7799.3	93	3111	A3-E1-R1-S1	645.33327	1	6078.7	93	321	R3-A2-I1	783.50187	1	10532.6
94	21111	I2-E1-P1-R1-S1	378.22517	2	7356.1	94	3111	**S3-A1-P1-R1**	645.33327	1	6078.7	94	2211	A2-E2-I1-R1	729.39206	1	10385.9
95	3111	**R3-A1-S1-W1**	436.75041	2	6997.7	95	2211	P2-R2-E1-I1	404.74522	2	6039.2	95	21111	S2-E1-I1-P1-R1	729.39206	1	10385.9
96	3111	R3-I1-P1-S1	413.26193	2	6502.2	96	2211	**I2-R2-S1-W1**	436.26551	2	6005.8	96	3111	R3-I1-P1-S1	413.26128	2	10301.8
97	222	E2-P2-R2	824.43269	1	6342.5	97	111111	A1-E1-I1-P1-R1-W1	812.44281	1	5823.6	97	321	A3-R2-W1	386.22272	2	10216.4
98	111111	A1-E1-I1-P1-R1-S1	713.39506	1	6203.4	98	2211	A2-E2-I1-R1	729.3908	1	5682.9	98	2211	E2-I2-A1-R1	386.22272	2	10216.4
99	2211	A2-I2-E1-R1	357.22004	2	6046.4	99	21111	S2-E1-I1-P1-R1	729.3908	1	5682.9	99	21111	A2-E1-P1-R1-S1	671.34956	1	9975.9
100	2211	I2-S2-P1-R1	357.22004	2	6046.4	100	2211	I2-P2-A1-R1	354.23325	2	5395.9	100	321	S3-P2-R1	671.34956	1	9975.9
101	321	R3-A2-I1	783.50214	1	5979.4	101	2211	I2-R2-E1-S1	407.74785	2	5134.8	101	21111	A2-E1-P1-R1-W1	385.70199	2	9906.6
102	3111	R3-A1-E1-I1	841.51886	1	5969.5	102	222	E2-P2-R2	824.4344	1	4980.6	102	2211	P2-S2-R1-W1	385.70199	2	9906.6
103	21111	A2-E1-P1-R1-W1	385.70207	2	5918.5	103	21111	P2-A1-I1-R1-W1	390.73036	2	4935.5	103	111111	A1-E1-I1-P1-R1-W1	406.72513	2	9894.7
104	2211	P2-S2-R1-W1	385.70207	2	5918.5	104	321	S3-A2-R1	619.31915	1	4885.7	104	321	**I3-R2-E1**	420.7748	2	9880.2
105	321	R3-E2-P1	883.48172	1	5362.7	105	21111	P2-A1-E1-R1-W1	398.70957	2	4744.9	105	321	**A3-S2-R1**	603.3231	1	9662.5
106	321	S3-A2-R1	619.31875	1	5276.7	106	3111	R3-A1-E1-I1	841.51835	1	4607.6	106	21111	R2-A1-I1-P1-S1	740.44846	1	9182.6
107	2211	I2-P2-A1-R1	354.23329	2	5237.5	107	21111	**P2-I1-R1-S1-W1**	796.44157	1	4533.6	107	21111	P2-A1-I1-R1-W1	390.73046	2	8973.8
108	21111	P2-A1-E1-R1-W1	398.70958	2	4738.9	108	21111	**I2-A1-P1-R1-W1**	398.74587	2	4307.5	108	321	R3-E2-A1	857.46514	1	8570
109	2211	I2-R2-E1-S1	407.74777	2	4652.6	109	21111	P2-E1-I1-R1-W1	419.73277	2	4203.7	109	321	R3-E2-P1	883.48077	1	8292.8
110	321	**R3-A2-S1**	379.23278	2	4176.1	110	321	A3-R2-W1	386.2227	2	4057.7	110	2211	A2-P2-R1-W1	369.70743	2	8156.4
111	3111	S3-I1-R1-W1	776.40382	1	4079.7	111	2211	E2-I2-A1-R1	386.2227	2	4057.7	111	321	S3-A2-R1	619.31958	1	8040.5
112	321	A3-R2-W1	386.22266	2	3898.3	112	321	R3-A2-I1	783.50202	1	3965.5	112	21111	P2-E1-I1-R1-W1	419.7328	2	7768.6
113	2211	E2-I2-A1-R1	386.22266	2	3898.3	113	21111	A2-I1-P1-R1-W1	377.72271	2	3682.5	113	21111	A2-I1-P1-R1-W1	377.72277	2	7021.4
114	111111	A1-E1-I1-P1-R1-W1	406.72507	2	3876.3	114	21111	R2-A1-I1-P1-S1	370.73	2	3487.9	114	111111	E1-I1-P1-R1-S1-W1	414.72243	2	6812.1
115	21111	P2-A1-I1-R1-W1	390.7304	2	3478.1	115	321	R3-E2-P1	883.48057	1	3403	115	21111	I2-E1-P1-R1-W1	427.74848	2	6453.8
116	2211	A2-P2-R1-W1	369.70741	2	3454.3	116	111111	E1-I1-P1-R1-S1-W1	414.72238	2	3357.9	116	3111	**P3-A1-R1-W1**	382.71506	2	6348
117	21111	**I2-A1-P1-R1-W1**	398.74586	2	3294.7	117	21111	**P2-E1-I1-R1-S1**	370.20974	2	3341.9	117	21111	P2-E1-R1-S1-W1	406.70709	2	6220.8
118	321	P3-R2-I1	388.74928	2	2833.2	118	2211	A2-P2-R1-W1	369.7074	2	3211.3	118	321	R3-S2-A1	387.23011	2	6028.4
119	21111	P2-E1-I1-R1-W1	419.7328	2	2827.2	119	321	P3-R2-I1	388.74932	2	3097.9	119	2211	I2-P2-R1-W1	411.75371	2	5982.6
120	2211	A2-E2-I1-R1	729.39031	1	2637.1	120	21111	W2-A1-E1-R1-S1	438.21271	2	3078.3	120	21111	I2-P1-R1-S1-W1	406.7433	2	5767
121	21111	S2-E1-I1-P1-R1	729.39031	1	2637.1	121	2211	I2-P2-R1-W1	411.75361	2	3053.4	121	21111	P2-A1-R1-S1-W1	754.39886	1	5731.7
122	21111	I2-E1-P1-R1-W1	427.74835	2	2557.6	122	3111	S3-I1-R1-W1	776.40445	1	3042.7	122	411	P4-E1-R1	733.3927	1	5500.7
123	111111	E1-I1-P1-R1-S1-W1	414.72237	2	2552.7	123	321	P3-R2-W1	425.24659	2	2933.4	123	2211	I2-R2-E1-S1	407.74796	2	4859.3
124	21111	**W2-A1-I1-P1-R1**	435.24257	2	2539.8	124	21111	I2-P1-R1-S1-W1	406.7432	2	2811.5	124	2211	A2-R2-I1-S1	357.72204	2	4318.5
125	21111	A2-I1-P1-R1-W1	377.72268	2	2509.6	125	321	**P3-I2-R1**	367.24088	2	2694.7	125	411	P4-I1-R1	717.43881	1	3662.3
126	321	P3-R2-W1	425.24713	2	2456.4	126	21111	**W2-A1-I1-P1-R1**	435.24262	2	2616.2	126	321	P3-A2-R1	649.37288	1	3551.2
127	321	P3-A2-R1	649.37346	1	2322.9	127	321	R3-S2-A1	387.22916	2	2287.2	127	21111	E2-A1-I1-P1-R1	378.20679	2	3376.3
128	2211	I2-P2-R1-W1	411.75365	2	2182.3	128	21111	E2-I1-P1-R1-S1	386.20412	2	2110.5	128	21111	E2-I1-P1-R1-S1	386.2041	2	3294.6
129	321	R3-S2-A1	387.23012	2	2121.3	129	321	R3-P2-S1	405.24726	2	2073.9	129	21111	W2-E1-I1-P1-R1	464.2446	2	3292.6

anking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY
130	21111	P2-A1-R1-S1-W1	754.39837	1	2063.7	130	21111	P2-A1-R1-S1-W1	754.39842	1	1960.8	130	321	P3-R2-I1	388.74917	2	3280.2
131	411	P4-R1-W1	790.43934	1	1927.5	131	3111	**P3-A1-R1-W1**	764.42453	1	1874.5	131	111111	**A1-I1-P1-R1-S1-W1**	770.43058	1	3121.9
132	21111	E2-I1-P1-R1-S1	386.20409	2	1861	132	21111	P2-E1-R1-S1-W1	406.70716	2	1755.9	132	21111	I2-A1-E1-R1-W1	414.74063	2	3096.5
133	2211	P2-R2-A1-E1	383.72054	2	1853.6	133	21111	E2-A1-I1-P1-R1	378.20708	2	1705.4	133	2211	E2-S2-A1-R1	719.3334	1	2970.4
134	21111	I2-P1-R1-S1-W1	406.74315	2	1822.9	134	2211	P2-R2-A1-E1	383.72056	2	1695.6	134	111111	A1-E1-I1-R1-S1-W1	401.71494	2	2786.1
135	21111	E2-A1-I1-P1-R1	378.20721	2	1809.6	135	321	P3-A2-R1	649.37325	1	1694.3	135	21111	**P2-I1-R1-S1-W1**	796.44547	1	2698.6
136	2211	A2-E2-R1-S1	703.33905	1	1615.2	136	21111	W2-E1-I1-P1-R1	464.24447	2	1686.9	136	2211	P2-R2-A1-E1	383.72048	2	2681.4
137	3111	S3-E1-P1-R1	703.33905	1	1615.2	137	321	R3-S2-I1	815.49122	1	1537.9	137	2211	**A2-R2-I1-P1**	724.4527	1	2595.8
138	21111	P2-E1-R1-S1-W1	406.7071	2	1597.2	138	2211	I2-S2-E1-R1	373.21485	2	1475.2	138	321	R3-I2-S1	421.27628	2	2243.6
139	411	P4-I1-R1	717.43891	1	1534.3	139	111111	**A1-I1-P1-R1-S1-W1**	770.42992	1	1429.7	139	21111	I2-E1-R1-S1-W1	422.73783	2	2210.7
140	21111	W2-E1-I1-P1-R1	464.24465	2	1433.1	140	2211	P2-R2-A1-S1	362.71424	2	1414	140	21111	E2-A1-I1-R1-W1	422.72001	2	2201
141	2211	P2-R2-A1-S1	362.7143	2	1305.4	141	3111	**P3-I1-R1-W1**	403.73815	2	1373.4	141	2211	P2-R2-A1-S1	362.71426	2	2093.5
142	2211	A2-R2-I1-S1	357.72191	2	1077.7	142	411	P4-I1-R1	717.43901	1	1370.1	142	3111	A3-I1-R1-W1	364.71526	2	1821
143	3111	**I3-P1-R1-S1**	370.24646	2	967.3	143	3111	**I3-P1-R1-S1**	370.24637	2	1264.5	143	3111	P3-R1-S1-W1	780.4141	1	1471.6
144	21111	I2-A1-E1-R1-W1	414.74032	2	876.2	144	21111	I2-A1-E1-R1-W1	414.7403	2	1238	144	3111	A3-R1-S1-W1	702.37113	1	1166.6
145	3111	A3-E1-I1-R1	671.38563	1	737.1	145	2211	A2-R2-I1-S1	357.72194	2	1177.5	145	21111	S2-E1-I1-R1-W1	409.71239	2	939.1
146	21111	S2-A1-I1-P1-R1	671.38563	1	737.1	146	111111	A1-E1-I1-R1-S1-W1	401.71499	2	1115.4	146	321	**P3-I2-R1**	733.46902	1	804.8
147	111111	A1-E1-I1-R1-S1-W1	401.71496	2	736.1	147	3111	A3-I1-R1-W1	364.71537	2	1070.9	147	2211	P2-R2-E1-I1	808.47158	1	765.8
148	3111	A3-I1-R1-W1	364.71526	2	685.3	148	3111	A3-E1-I1-R1	671.38539	1	1063.2	148	21111	S2-A1-E1-I1-R1	703.37464	1	724.2
149	21111	E2-A1-I1-R1-W1	422.72153	2	525.5	149	21111	S2-A1-I1-P1-R1	671.38539	1	1063.2	149	2211	P2-R2-E1-S1	391.71774	2	672.7
150	21111	I2-E1-R1-S1-W1	422.73748	2	500.2	150	21111	E2-A1-I1-R1-W1	422.72027	2	925	150	3111	I3-A1-R1-S1	357.23933	2	513.2
151	2211	P2-R2-E1-S1	391.71771	2	463.7	151	21111	I2-E1-R1-S1-W1	422.73787	2	823.5	151	411	I4-R1-S1	755.52129	1	432.4
152	2211	**P2-W2-R1-S1**	435.22468	2	448.2	152	3111	A3-R1-S1-W1	702.37022	1	509.8	152	321	S3-R2-I1	373.71642	2	266.8
153	3111	A3-R1-S1-W1	702.37087	1	352.1	153	2211	P2-R2-E1-S1	391.71775	2	425.3	153	411	P4-R1-W1	790.43239	1	228.4
154	21111	S2-A1-E1-I1-R1	703.37508	1	208.3	154	21111	S2-A1-E1-I1-R1	703.37417	1	250.5	154	3111	P3-E1-R1-S1	362.19179	2	159.6
155	21111	S2-E1-I1-R1-W1	409.71252	2	126.4	155	21111	S2-E1-I1-R1-W1	409.71197	2	166.6	155	3111	S3-I1-R1-W1	776.40411	1	110.2

Appendix 5. Complete family list and information filtered on *other mechanisms transport list. Green (Acc1), Blue (Acc2) and Red (Acc3).*

anking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY
1	2211	A2-I2-P1-R1	681.43852	1	1273789.9	1	2211	R2-W2-A1-P1	456.75089	2	725681.9	1	2211	E2-P2-A1-R1	739.37724	1	1426227.8
2	2211	R2-W2-A1-P1	456.75083	2	808534.6	2	3111	I3-A1-P1-R1	723.48722	1	654126.9	2	2211	I2-R2-A1-W1	428.26933	2	1213477.5
3	3111	I3-A1-P1-R1	723.48718	1	637952.9	3	2211	E2-P2-A1-R1	739.3778	1	634342.3	3	3111	I3-A1-P1-R1	723.48694	1	1211556.9
4	2211	I2-R2-A1-W1	428.26926	2	612973.1	4	2211	I2-R2-A1-W1	428.26916	2	616244.6	4	321	I3-R2-P1	404.77979	2	1193905.8
5	2211	E2-P2-A1-R1	739.37805	1	577002.1	5	51	I5-R1	391.28533	2	588927.3	5	2211	R2-W2-A1-P1	456.75081	2	1149578.4
6	51	I5-R1	391.28534	2	488915.5	6	2211	E2-S2-P1-R1	745.35377	1	342921.2	6	2211	I2-R2-E1-W1	457.27187	2	678188.3
7	222	P2-R2-W2	469.75882	2	448093.6	7	321	E3-A2-R1	745.35377	1	342921.2	7	51	I5-R1	391.28537	2	558333.6
8	2211	E2-S2-P1-R1	745.35347	1	310069.3	8	21111	W2-A1-E1-P1-R1	885.43586	1	264483	8	21111	W2-I1-P1-R1-S1	885.47333	1	423725.1
9	321	E3-A2-R1	745.35347	1	310069.3	9	2211	P2-W2-A1-R1	427.22717	2	250092.4	9	2211	P2-W2-A1-R1	427.22709	2	365455.2
10	21111	W2-A1-E1-P1-R1	885.43618	1	286786.5	10	2211	I2-R2-E1-W1	457.27205	2	241303.7	10	2211	P2-W2-E1-R1	456.22981	2	330502.5
11	2211	I2-R2-E1-W1	457.27224	2	254977	11	2211	P2-W2-E1-R1	456.2297	2	237608.6	11	321	A3-R2-S1	672.39139	1	315872.7
12	2211	P2-W2-A1-R1	427.22709	2	240854.7	12	321	I3-R2-A1	391.77209	2	224311.5	12	2211	P2-W2-I1-R1	448.25042	2	310675.2
13	21111	W2-I1-P1-R1-S1	885.47202	1	234699.9	13	321	A3-R2-S1	672.39134	1	203924.5	13	3111	R3-P1-S1-W1	449.7598	2	296410.8
14	2211	P2-W2-I1-R1	895.49108	1	224252.2	14	21111	W2-A1-E1-I1-R1	451.23744	2	194467.2	14	21111	W2-E1-P1-R1-S1	901.43251	1	270044.9
15	3111	A3-E1-R1-W1	744.38356	1	208466	15	3111	R3-P1-S1-W1	449.75977	2	189038.2	15	411	I4-P1-R1	765.53494	1	250934.1
16	21111	S2-A1-P1-R1-W1	744.38356	1	208466	16	2211	P2-W2-I1-R1	895.49138	1	175781.9	16	2211	A2-W2-P1-R1	414.21925	2	241697.4
17	21111	W2-A1-E1-I1-R1	451.23741	2	207860.7	17	2211	A2-W2-P1-R1	414.21839	2	158679.7	17	2211	A2-R2-S1-W1	787.43103	1	233755.1
18	321	A3-R2-S1	672.39144	1	199714.9	18	222	A2-R2-S2	688.38638	1	141664.8	18	2211	E2-I2-R1-S1	787.43103	1	233755.1
19	3111	R3-P1-S1-W1	449.75982	2	190823.4	19	2211	R2-W2-A1-S1	451.74084	2	137800.1	19	3111	I3-A1-E1-R1	755.47968	1	226174.8
20	321	I3-R2-A1	391.77193	2	178637.6	20	2211	A2-R2-S1-W1	787.43094	1	127322.8	20	321	P3-W2-R1	879.46292	1	223900.2
21	21111	E2-A1-R1-S1-W1	818.38745	1	177051.9	21	2211	E2-I2-R1-S1	787.43094	1	127322.8	21	3111	R3-A1-P1-S1	392.24016	2	218279.8
22	2211	A2-I2-R1-W1	385.73708	2	176890.6	22	411	I4-P1-R1	765.53364	1	123199.9	22	222	A2-R2-S2	688.38625	1	208702.8
23	3111	I3-E1-R1-S1	386.23982	2	172163.4	23	3111	I3-A1-E1-R1	755.47961	1	111961.1	23	2211	A2-W2-I1-R1	422.23479	2	190675
24	321	P3-W2-R1	440.23491	2	162025	24	222	E2-R2-W2	501.74859	2	107483.7	24	2211	E2-P2-I1-R1	391.21268	2	169262.6
25	2211	R2-W2-A1-S1	451.74083	2	145866.8	25	2211	A2-W2-I1-R1	422.23473	2	98183.4	25	222	E2-R2-W2	501.74861	2	159805.8
26	222	A2-R2-S2	688.38638	1	133963.2	26	3111	R3-A1-P1-S1	392.2398	2	94881.4	26	321	W3-R2-A1	501.26504	2	155619
27	2211	A2-R2-S1-W1	787.43102	1	130564.7	27	321	W3-R2-E1	530.26892	2	93368.3	27	2211	R2-W2-A1-S1	451.73912	2	153514.4
28	2211	E2-I2-R1-S1	787.43102	1	130564.7	28	321	W3-R2-A1	501.26461	2	93160	28	321	W3-R2-E1	530.26921	2	150038.3
29	222	E2-R2-W2	501.74855	2	125440.7	29	2211	I2-S2-R1-W1	401.73238	2	92312.4	29	2211	E2-P2-R1-S1	755.37185	1	143772.7
30	411	I4-P1-R1	765.53412	1	118242.3	30	2211	E2-W2-I1-R1	480.23975	2	78986.8	30	2211	R2-W2-A1-E1	472.74858	2	139966.9
31	3111	I3-A1-E1-R1	755.47963	1	117295.9	31	2211	R2-W2-A1-E1	472.74817	2	70256.6	31	2211	A2-R2-E1-W1	829.44158	1	130062.4
32	21111	P2-A1-E1-R1-S1	697.36752	1	108789.7	32	2211	A2-R2-E1-W1	829.4416	1	62819.5	32	321	E3-I2-R1	829.44158	1	130062.4
33	3111	R3-A1-P1-S1	392.23983	2	101282.8	33	321	E3-I2-R1	829.4416	1	62819.5	33	2211	R2-S2-P1-W1	829.44158	1	130062.4
34	2211	A2-W2-I1-R1	422.23491	2	98213.5	34	2211	R2-S2-P1-W1	829.4416	1	62819.5	34	2211	E2-W2-I1-R1	959.47422	1	113103.5
35	3111	I3-P1-R1-W1	838.52738	1	93999.6	35	2211	E2-P2-R1-S1	755.37334	1	60118.1	35	2211	P2-R2-S1-W1	839.46295	1	106778.3
36	321	W3-R2-E1	530.26917	2	89219.1	36	2211	A2-W2-R1-S1	817.40305	1	58196.1	36	222	A2-I2-R2	370.74731	2	104338.3
37	321	W3-R2-A1	501.26517	2	88531.7	37	2211	E2-P2-R1-W1	854.41692	1	57639.9	37	21111	R2-A1-I1-P1-W1	420.25391	2	101883.6
38	2211	I2-S2-R1-W1	401.73236	2	86866.2	38	42	R4-A2	413.76441	2	53574.2	38	321	E3-R2-I1	872.46546	1	98220.9
39	2211	E2-W2-I1-R1	480.2397	2	86748	39	222	A2-I2-R2	370.74716	2	52848.4	39	2211	E2-P2-R1-W1	854.41771	1	86736.3
40	2211	R2-W2-A1-E1	472.74814	2	83263	40	321	E3-R2-I1	872.46502	1	48675	40	42	R4-A2	413.76438	2	84575.1
41	2211	E2-P2-I1-R1	391.21457	2	78482.4	41	2211	P2-R2-A1-I1	375.74222	2	46827.1	41	2211	E2-R2-A1-I1	814.44964	1	79168.3
42	2211	A2-R2-E1-W1	829.44147	1	67755.2	42	2211	P2-R2-S1-W1	839.46283	1	43957.6	42	2211	I2-R2-A1-S1	756.48302	1	77176.7
43	321	E3-I2-R1	829.44147	1	67755.2	43	2211	I2-W2-P1-R1	911.52295	1	43569.9	43	2211	S2-W2-I1-R1	875.45293	1	75689.5
44	2211	R2-S2-P1-W1	829.44147	1	67755.2	44	3111	S3-A1-E1-R1	677.32463	1	43047.5	44	2211	R2-W2-I1-S1	472.76353	2	70916.5

anking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY
45	2211	I2-W2-P1-R1	911.52291	1	66812.8	45	321	S3-R2-W1	410.2138	2	40724.7	45	3111	S3-A1-E1-R1	677.32473	1	65876.7
46	321	E3-R2-I1	872.4652	1	62278.9	46	321	R3-E2-I1	450.26166	2	40243	46	321	R3-E2-I1	450.26159	2	63545.3
47	2211	P2-R2-S1-W1	839.46245	1	60264.5	47	2211	I2-R2-A1-S1	756.48294	1	38608.9	47	2211	A2-W2-E1-R1	859.42323	1	50812.5
48	42	R4-A2	413.7644	2	58818	48	2211	S2-W2-I1-R1	875.45271	1	38377.7	48	2211	S2-W2-P1-R1	859.42323	1	50812.5
49	2211	E2-P2-R1-S1	755.37354	1	56020.2	49	21111	R2-A1-I1-P1-W1	420.254	2	37514.6	49	2211	I2-W2-R1-S1	451.2556	2	46484.9
50	2211	E2-P2-R1-W1	854.41682	1	52106.8	50	2211	S2-W2-A1-R1	833.4062	1	30440.6	50	2211	P2-R2-A1-I1	375.74211	2	44529.2
51	222	A2-I2-R2	370.74717	2	47749.3	51	2211	R2-W2-I1-S1	472.76343	2	29009.6	51	2211	S2-W2-A1-R1	833.4065	1	40188.6
52	2211	S2-W2-I1-R1	875.45238	1	46698.6	52	2211	A2-P2-R1-S1	639.36215	1	25550.1	52	2211	A2-R2-E1-I1	756.45258	1	36923.1
53	2211	P2-R2-A1-I1	375.74224	2	45126.3	53	21111	R2-E1-I1-P1-S1	798.45515	1	23262.1	53	2211	R2-S2-I1-P1	756.45258	1	36923.1
54	2211	A2-W2-P1-R1	414.22164	2	42554.2	54	3111	W3-A1-P1-R1	942.46878	1	19782	54	21111	W2-A1-E1-I1-R1	901.47392	1	35436.5
55	2211	I2-R2-A1-S1	756.48298	1	41700.7	55	2211	A2-R2-E1-I1	756.45227	1	15878.8	55	21111	R2-E1-I1-P1-S1	798.45508	1	34364.9
56	222	I2-R2-S2	386.74186	2	40542	56	2211	R2-S2-I1-P1	756.45227	1	15878.8	56	2211	A2-P2-R1-S1	639.36276	1	33774.7
57	321	R3-E2-I1	450.26158	2	40506.8	57	2211	I2-W2-R1-S1	451.25561	2	15469	57	321	I3-S2-R1	365.23543	2	32317.2
58	2211	E2-R2-A1-I1	814.44967	1	40032.9	58	2211	A2-W2-E1-R1	859.42399	1	14605.6	58	2211	E2-R2-I1-S1	830.44342	1	31186
59	2211	R2-W2-I1-S1	472.76331	2	37465.3	59	2211	S2-W2-P1-R1	859.42399	1	14605.6	59	3111	R3-A1-E1-S1	815.45398	1	30731.6
60	3111	S3-A1-E1-R1	677.32465	1	36906.1	60	21111	W2-I1-P1-R1-S1	443.24275	2	14392.6	60	3111	S3-E1-I1-R1	719.37291	1	26508.8
61	21111	R2-A1-I1-P1-W1	420.25398	2	36192.2	61	2211	E2-R2-A1-P1	798.42057	1	14190.5	61	3111	W3-A1-R1-S1	932.45054	1	26362.9
62	21111	R2-A1-E1-I1-P1	782.47037	1	34633.5	62	2211	E2-R2-I1-S1	830.44336	1	14055.7	62	42	I4-R2	412.79534	2	25133
63	21111	S2-A1-E1-R1-W1	776.37535	1	34256.2	63	21111	W2-E1-P1-R1-S1	451.21913	2	13050.9	63	3111	W3-A1-P1-R1	471.74215	2	24569.7
64	2211	S2-W2-A1-R1	833.40652	1	27506.9	64	321	I3-S2-R1	365.23454	2	12932.8	64	2211	A2-I2-P1-R1	681.44383	1	23791.3
65	321	A3-W2-R1	401.21199	2	26034.5	65	2211	A2-I2-P1-R1	681.44325	1	12689.9	65	3111	I3-E1-R1-S1	771.47531	1	23709
66	2211	A2-P2-R1-S1	639.36242	1	24338.4	66	2211	A2-R2-I1-W1	407.24594	2	12651.3	66	2211	E2-R2-I1-P1	840.46564	1	22853.7
67	21111	R2-E1-I1-P1-S1	798.45456	1	22185.4	67	321	I3-E2-R1	407.24594	2	12651.3	67	2211	E2-R2-A1-P1	798.42329	1	22719.1
68	33	R3-S3	395.2252	2	20060.1	68	3111	S3-E1-I1-R1	719.3733	1	12548.6	68	3111	W3-P1-R1-S1	479.73749	2	22444.7
69	2211	A2-R2-E1-I1	756.45236	1	19003.3	69	2211	A2-I2-R1-W1	770.46283	1	12199.2	69	321	R3-E2-S1	873.46249	1	20005.6
70	2211	R2-S2-I1-P1	756.45236	1	19003.3	70	3111	I3-E1-R1-S1	771.47508	1	11871.4	70	411	S4-I1-R1	677.36259	1	18654.2
71	3111	W3-A1-R1-S1	932.44874	1	16261.9	71	3111	R3-A1-E1-S1	815.45377	1	11457.9	71	3111	R3-E1-P1-S1	841.46942	1	16857.7
72	2211	E2-R2-I1-S1	830.44331	1	15963	72	411	S4-I1-R1	677.36166	1	10974	72	33	R3-S3	395.22487	2	16096.4
73	3111	R3-A1-E1-S1	815.45336	1	14813.1	73	42	I4-R2	412.79525	2	10544.1	73	222	E2-I2-R2	428.7531	2	16008.1
74	21111	W2-E1-P1-R1-S1	451.2194	2	14040.3	74	3111	W3-A1-R1-S1	932.4471	1	10265	74	321	R3-A2-W1	428.7531	2	16008.1
75	2211	E2-R2-I1-P1	840.46528	1	13964.9	75	2211	E2-R2-I1-P1	840.46558	1	9702.5	75	21111	A2-E1-R1-S1-W1	760.36811	1	15667.4
76	2211	S2-W2-E1-R1	446.20896	2	13487.3	76	21111	R2-A1-I1-S1-W1	415.2435	2	9595.3	76	3111	S3-P1-R1-W1	760.36811	1	15667.4
77	2211	E2-R2-A1-P1	798.41984	1	13286.6	77	3111	S3-E1-R1-W1	792.36808	1	9370.5	77	21111	S2-A1-E1-R1-W1	776.36853	1	14461.6
78	411	S4-I1-R1	677.36216	1	13013.7	78	222	E2-I2-R2	428.75333	2	9170.3	78	21111	W2-A1-E1-P1-R1	443.22419	2	14427.8
79	42	A4-R2	656.39812	1	12162.4	79	321	R3-A2-W1	428.75333	2	9170.3	79	42	A4-R2	656.39964	1	14030.7
80	321	R3-E2-S1	873.46512	1	12065.3	80	42	A4-R2	656.39658	1	8883.9	80	2211	A2-R2-I1-W1	407.2467	2	13876.8
81	3111	W3-A1-P1-R1	471.74243	2	11516.7	81	321	R3-E2-S1	873.4622	1	8769.3	81	321	I3-E2-R1	407.2467	2	13876.8
82	2211	A2-R2-I1-W1	407.24564	2	11144.1	82	321	I3-R2-P1	404.78031	2	8582.4	82	2211	E2-S2-P1-R1	745.34782	1	13397.7
83	321	I3-E2-R1	407.24564	2	11144.1	83	21111	A2-E1-R1-S1-W1	760.36884	1	7875.1	83	321	E3-A2-R1	745.34782	1	13397.7
84	3111	R3-E1-P1-S1	841.46944	1	10079.8	84	3111	S3-P1-R1-W1	760.36884	1	7875.1	84	3111	R3-E1-I1-S1	857.50054	1	12954.9
85	321	I3-S2-R1	365.23425	2	9931.6	85	3111	R3-E1-P1-S1	841.46992	1	7366.2	85	3111	S3-E1-R1-W1	792.36984	1	11981.2
86	2211	I2-W2-R1-S1	451.2552	2	9111.9	86	2211	R2-W2-E1-P1	485.75364	2	6998.2	86	411	A4-P1-R1	597.35191	1	11829
87	42	I4-R2	412.79564	2	8678.9	87	411	A4-P1-R1	597.35251	1	6784	87	321	I3-R2-A1	391.77205	2	9718.3
88	222	E2-I2-R2	428.7534	2	8578.7	88	21111	W2-E1-I1-R1-S1	459.23669	2	5866.6	88	321	E3-R2-P1	856.42507	1	9711

anking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY	Ranking	FAMILY	FORMULA	M/Z	Z	INTENSITY
89	321	R3-A2-W1	428.7534	2	8578.7	89	321	S3-W2-R1	849.4006	1	5833	89	21111	R2-A1-I1-S1-W1	415.24342	2	9677.9
90	21111	R2-A1-I1-S1-W1	415.24354	2	8260.4	90	321	E3-R2-P1	856.42388	1	5721.5	90	3111	R3-A1-E1-W1	914.5147	1	9366.6
91	3111	R3-A1-E1-W1	914.51502	1	8249	91	3111	P3-A1-R1-S1	665.36931	1	5685.2	91	321	R3-S2-W1	888.49497	1	9027.6
92	321	I3-R2-P1	404.78027	2	8025.9	92	21111	P2-A1-E1-R1-S1	697.36565	1	5433.3	92	222	I2-R2-S2	772.47841	1	8940.6
93	3111	S3-E1-R1-W1	792.36763	1	7938.9	93	21111	S2-A1-E1-R1-W1	776.36844	1	4890.4	93	21111	A2-P1-R1-S1-W1	728.38405	1	8717.9
94	411	A4-P1-R1	597.35217	1	7132.9	94	21111	A2-I1-R1-S1-W1	744.41347	1	4769.1	94	2211	R2-W2-E1-P1	485.75361	2	8528.6
95	321	E3-R2-P1	856.42367	1	6721.4	95	21111	A2-P1-R1-S1-W1	728.38333	1	4682.3	95	222	P2-R2-W2	469.75823	2	8459.8
96	321	S3-W2-R1	849.39926	1	6352.1	96	222	I2-R2-S2	772.47786	1	4317.5	96	3111	P3-A1-R1-S1	665.37107	1	7929.6
97	3111	P3-A1-R1-S1	665.36964	1	6016.7	97	3111	R3-E1-I1-S1	857.50063	1	4139.4	97	321	S3-W2-R1	849.39918	1	7634
98	3111	R3-E1-I1-S1	857.50032	1	5948.3	98	2211	E2-P2-I1-R1	391.21533	2	3837.1	98	321	W3-A2-R1	916.46171	1	7447.1
99	21111	W2-E1-I1-R1-S1	459.23651	2	5910.2	99	222	P2-R2-W2	469.75832	2	3801.2	99	3111	R3-A1-E1-P1	825.48704	1	7258.2
100	21111	A2-I1-R1-S1-W1	744.41425	1	5663.2	100	321	W3-A2-R1	916.4597	1	3721.1	100	21111	W2-E1-I1-R1-S1	459.23527	2	6686.6
101	21111	A2-P1-R1-S1-W1	728.38499	1	5620.6	101	3111	W3-P1-R1-S1	479.73734	2	3454.2	101	21111	E2-A1-R1-S1-W1	818.37909	1	4417.3
102	2211	R2-W2-E1-P1	485.75322	2	5310.7	102	2211	R2-S2-I1-W1	423.24119	2	3085.3	102	3111	P3-A1-E1-R1	707.38697	1	4273.8
103	3111	R3-A1-E1-P1	825.48397	1	4311.6	103	321	P3-W2-R1	879.47139	1	2937.1	103	2211	A2-W2-R1-S1	817.41546	1	4045.1
104	321	R3-S2-W1	888.49381	1	4272.1	104	21111	R2-A1-E1-I1-P1	391.73595	2	2930.8	104	21111	P2-A1-E1-R1-S1	697.36544	1	3935.3
105	321	W3-A2-R1	916.45781	1	4104.3	105	3111	R3-A1-E1-P1	825.48317	1	2698.5	105	321	A3-W2-R1	401.21142	2	3783.1
106	3111	W3-P1-R1-S1	479.73675	2	3446.1	106	3111	R3-A1-E1-W1	914.51412	1	2693.3	106	222	A2-R2-W2	443.74297	2	3657.3
107	2211	A2-W2-E1-R1	859.41566	1	2087.6	107	321	R3-S2-W1	888.48825	1	2646.5	107	2211	E2-I2-R1-W1	443.74297	2	3657.3
108	2211	S2-W2-P1-R1	859.41566	1	2087.6	108	3111	A3-E1-R1-W1	744.38241	1	2405.7	108	42	S4-R2	720.38008	1	3360.4
109	2211	A2-W2-R1-S1	817.41496	1	1984.4	109	21111	S2-A1-P1-R1-W1	744.38241	1	2405.7	109	2211	R2-W2-P1-S1	464.74726	2	2871.8
110	3111	S3-E1-I1-R1	719.37387	1	1932.8	110	321	A3-W2-R1	401.21231	2	2287.8	110	2211	A2-I2-R1-W1	385.73826	2	2801.3
111	42	S4-R2	720.37913	1	1467.5	111	21111	E2-A1-R1-S1-W1	818.37803	1	1878.9	111	3111	R3-I1-P1-W1	462.7874	2	2644.1
112	3111	R3-I1-P1-W1	462.78762	2	1414.4	112	3111	R3-I1-P1-W1	462.78725	2	1822.8	112	21111	R2-A1-E1-I1-P1	391.73589	2	2226.7
113	42	P4-R2	380.73418	2	1159.8	113	3111	P3-A1-E1-R1	707.38749	1	1797.7	113	21111	I2-A1-R1-S1-W1	393.7357	2	2122.5
114	321	P3-R2-E1	396.72802	2	1104.7	114	42	S4-R2	720.37985	1	1326.9	114	321	P3-R2-A1	367.72637	2	1847
115	2211	R2-S2-I1-W1	423.241	2	1033.4	115	222	A2-R2-W2	443.74283	2	1319.2	115	2211	R2-S2-I1-W1	423.24033	2	1571
116	2211	R2-W2-P1-S1	464.74702	2	917	116	2211	E2-I2-R1-W1	443.74283	2	1319.2	116	321	P3-R2-E1	792.44985	1	1441.8
117	222	A2-R2-W2	443.74251	2	853.5	117	2211	R2-W2-P1-S1	464.74701	2	1288.8	117	2211	I2-W2-P1-R1	456.26537	2	1371.4
118	2211	E2-I2-R1-W1	443.74251	2	853.5	118	2211	S2-W2-E1-R1	891.41935	1	587.5	118	3111	A3-E1-R1-W1	744.37939	1	1310.7
119	321	P3-R2-A1	734.44676	1	648	119	21111	I2-A1-R1-S1-W1	393.73572	2	542.7	119	21111	S2-A1-P1-R1-W1	744.37939	1	1310.7
120	33	P3-R3	819.49983	1	500.4	120	33	R3-S3	395.22606	2	443.8	120	3111	I3-P1-R1-W1	419.76893	2	1211.9
121	2211	P2-W2-E1-R1	456.23001	2	379.2	121	321	P3-R2-E1	792.45031	1	383.2	121	21111	A2-I1-R1-S1-W1	744.41123	1	878.6
122	21111	I2-A1-R1-S1-W1	393.73523	2	356.2	122	3111	I3-P1-R1-W1	419.76871	2	332.3	122	2211	S2-W2-E1-R1	891.41693	1	797.9
123	21111	W2-A1-I1-R1-S1	430.23087	2	275.2	123	21111	W2-A1-I1-R1-S1	430.23129	2	312.8	123	2211	I2-S2-R1-W1	401.73253	2	791.9
124	2211	A2-E2-R1-W1	401.69735	2	207.6	124	2211	A2-E2-R1-W1	401.69707	2	286	124	42	P4-R2	380.73343	2	637.6
125	21111	S2-E1-P1-R1-W1	401.69735	2	207.6	125	21111	S2-E1-P1-R1-W1	401.69707	2	286	125	21111	W2-A1-I1-R1-S1	430.23217	2	606.5
126	321	R3-P2-I1	418.27262	2	169.7	126	2211	E2-R2-A1-I1	814.46086	1	246.5	126	321	R3-P2-I1	418.27395	2	364
127	321	S3-R2-W1	410.2138	2	113.8	127	321	R3-P2-I1	418.27267	2	187.6	127	321	S3-R2-W1	410.21387	2	363.3
128	21111	A2-E1-R1-S1-W1	760.37425	1	108	128	42	P4-R2	380.73232	2	157.6	128	2211	A2-E2-R1-W1	401.69827	2	359.2
129	3111	S3-P1-R1-W1	760.37425	1	108	129	222	A2-P2-R2	354.71755	2	136.4	129	21111	S2-E1-P1-R1-W1	401.69827	2	359.2
130	222	A2-P2-R2	354.71727	2	105.3	130	321	P3-R2-A1	367.72653	2	119.4	130	222	A2-P2-R2	354.7169	2	310.8
131	3111	P3-A1-E1-R1	707.3806	1	100.1	131	33	P3-R3	410.25752	2	118.2	131	33	P3-R3	410.25677	2	113.9

Appendix 6. Perl program script to generate permutations (by Xavier Arroyo).

```
#!/usr/bin/perl -w
use strict;
use Data::Dumper;

my %aah = (
        "A" => 71.037110,
        "E" => 129.04259,
        "I" => 113.08406,
        "P" => 97.052760,
        "R" => 156.10110,
        "S" => 87.032030,
        "W" => 186.07931
);

my @aa = qw(A E I P R S W);
my @pes= qw(71.03711 129.04259 113.08406 97.052760 156.1011 87.032030 186.07931);
my $suma;
my @a;
my @vector;
my $mida = 6;
my $ltotal = 5;
my $ntotal = 5;
my $pesBase = 43.01837 + 16.01871;
my $targetPes;
my $precisio = 10;
my $th = shift || 0.012;

sub patro {
        my $s=shift;
        my %f;
        my @arr = split("",$s);
        foreach (@arr) {
                $f{$_}=0 if (not defined $f{$_});
                $f{$_}++;
        }
        my $formula = "";
        my $familia = "";
        foreach (sort {$f{$b} <=> $f{$a}} sort keys %f) {
                $formula .= $_.$f{$_}."-";
        }
        chop $formula;
        return $formula;
}

sub bibliopepfinder {
        my $level = shift || 0;
        if ($level == $mida) {
                my $suma = 0;
                $suma += $aah{$_} foreach(@vector);
                print &patro(join("",@vector)),"\t",$suma,"\n";
        }
        else {
```

```
                foreach(keys %aah) {
                        unshift @vector, $_;
                        &bibliopepfinder($level+1);
                        shift @vector;
                }
        }
}
$mida = 2;
&bibliopepfinder();
$mida = 3;
&bibliopepfinder();
$mida = 4;
&bibliopepfinder();
$mida = 5;
&bibliopepfinder();
$mida = 6;
&bibliopepfinder();
```

Appendix 7. Detailed information of transitions found for sequences monitored by SRM.

	Acc1		Acc2		Acc3	
I3-P2-R1	+	++	+	++	+	++
RIIIPP	16.2	16.2		23	28	
RIIPIP				17	28	
RIIPPI		27.5				
RIPIIP		16.2	*25.8*			28
RIPIPI		27.5		27.2		28
RIPPII	26.1		25.8			26.6
RPIIIP	18.8		18.5	25.8	19.2	
RPIIPI	26.1		25.5			*26.6*
RPIPII	26.1		25.5		26.2	26.2
RPPIII	17.1/26.1		24.4		17.5/26.6	25.2

	Acc1		Acc2		Acc3	
P3-S2-R1	+	++	+	++	+	++
RPPPSS	14.6		14.6		15.4/15.9	
RPPSPS						
RPPSSP	23.4			11.5	23.8	12.3
RPSPPS					20.3	
RPSPSP				11.5		
RPSSPP		**21.0**				12.3
RSPPPS				11.5		
RSPPSP		11.5		11.5		12.3
RSPSPP		11.5		11.5		12.3
RSSPPP		11.5		11.5		12.3

	Acc1		Acc2		Acc3	
A3-R2-P1	+	++	+	++	+	++
RAAAPR		10.7				12.3
RAAARP	15.2		15.2		16	15.6
RAAPAR	26.4					
RAAPRA		11.5	17.1			
RAARAP		15.2		*11.2*		12.0
RAARPA	*26.4*			11.2		
RAPAAR	26.4			10.7		
RAPARA	26.4		*24.8*			
RAPRAA				11.2		
RARAAP		15.2		11.4		20.8
RARAPA	*23.5*		23.1			
RARPAA				10.5		12.2
RPAAAR			26.4			
RPAARA	25.5		26.0			
RPARAA	25.6		25.2		*12.2*	12.3
RPRAAA	17.4	11.5	23.0		23.8	12.3
RRAAAP	23.5		23.0		23.8	13.0/16.0/ 19.0
RRAAPA	23.5		23.0		20.7	
RRAPAA	23.4		23.0		23.8	12.2
RRPAAA	23.4		23.0		20.7	16.0

	Acc1		Acc2		Acc3	
I4-A1-R1	+	++	+	++	+	++
RIIIIP	**20.4**/**20.9**/ 21.7/22.5/ 24.2/**31.3**	20.4/**20.9**/ **21.7**/**22.5**/ 24.2	**19.9**/20.4/ 21.0/**22.0**/ 22.6/23.8/ **30.8**/31.1/ **31.9**	19.9/20.4/ **21.0**/**22.0**/ 27.5	**19.1**/19.5/ **20.4**/21.2/ 23.0	19.1/19.5/ 20.4/21.2/ **26.5**
RIIIPI	**22.5**/**23.5**	**20.4**/23.5/ 31.3/**32.4**	22.6/**22.9**	**19.9**/**20.4**/ 22.0/**30.8**/ 31.1/**31.9**	**22.2**	19.1/22.2/ **29.8**
RIIPII	**20.9**/**22.5**/ 31.3	**20.9**/21.7/ **23.5**	19.9/**20.4**/ **31.9**	**19.9**/**20.4**/ 21.0/**22.0**/ **22.6**/**23.8**/ **31.9**	19.1/19.5/ **22.2**/**31.3**	**19.1**/**19.5**/ **21.8**/**31.3**
RIPIII	23.0/24.2/ **32.0**	20.9/21.7/**22.5**/**23.0**/**32.0**	22.0/**22.6**/2 2.9/**31.1**	**21.0**/**22.0**/ **22.6**/**31.1**/ **31.9**	19.1/**22.2**	21.8/22.2/ **30.5**/**31.3**
RPIIII	**20.9**/**22.5**	20.9/**21.7**/ **22.5**/**32.0**	**20.4**/22.0/ 22.6/**22.9**/ **31.1**/31.9	**20.4**/22.0/ **22.6**/**31.1**	19.1/**19.5**/ **22.2**	19.5/**22.2**

	Acc1		Acc2		Acc3	
I3-A1-P1-R1	+	++	+	++	+	++
RAIIIP	**18.3/27.1**	*24.8/27.1/* **28.1**	15.8/18.3	15.8	**15.5/16.7/ 23.0**	*16.7/23.1*
RAIIPI	**27.1**	17.1/**21.7**	**16.7**	**27.7**	15.5/**26.2**	
RAIPII	**27.1**	**28.0**		**15.8/27.7**	**15.5**	*15.5*/26.6
RAPIII	**25.0**	**25.0**/27.1/ **28.0**	15.8/**16.2**/ 24.4	**15.8/16.6/ 24.4/27.2/ 27.7**	**15.0/23.1**	25.8
RIAIIP	**16.3/17.2/ 18.3/26.2/ 27.1**	**26.1/27.1**	*15.8*/**16.7**		**16.7/20.0**/ *25.5*	
RIAIPI	**27.1/28.0**	**26.2/27.1/ 28.0**	**16.7**		*20.0*/**26.2**	*20.0*
RIAPII	*16.5*/**27.0**	**16.3/28.0**	*15.8***/27.3**	**16.6/27.7**	*25.1*	**24.4**
RIIAIP	**16.2/17.2/ 27.1**	16.3/17.1/ 24.8/**27.1**	**15.8/16.6/ 26.5**	15.8/16.6/ 24.6/25.6	**14.6/**15.2/ 15.5/**22.3/** 23.1/**23.5/ 25.5**	15.2/15.5/ **22.3/**23.1
RIIAPI	*16.2*/17.2/ *25.2*/**27.1**	17.1/**27.1**	**15.8**	15.8/16.6/ 24.6	14.6/15.5	14.6/15.5
RIIIAP	**16.3/**18.3	16.0/16.3/ 18.3/24.1	**15.8/**17.8/ **23.5**	15.8/17.9/ **23.5/24.2**	*14.8*/16.7/ **22.3**	14.3/14.7/ 16.7/**22.3/** 23.1
RIIIPA	**24.1/***27.1*		*17.8/***23.5**	*15.5*	16.7/**22.3**	14.4/16.3/ **25.6**
RIIPAI	16.3/17.1/ **25.0/27.1**	**16.3**/17.1/ **27.1/28.0/ 28.6**	**15.8**	**15.8**/*16.1/* ***16.6****/26.6/* ***27.2***	*15.5/25.5*	14.6/*15.0*/ 15.5/**25.6**
RIIPIA	*17.1*/**27.1**	*28.6*	**16.7/27.2**	**26.4**	**25.1**/25.5	**15.2/25.6**
RIPAII	*27.0*/**27.8**	**16.3/28.0**	16.7/**27.3**	**27.7**	24.4	*25.8*
RIPIAI	**21.7/22.9/ 27.1**	**21.6/22.9/ 28.6**	*17.0/24.4*/ **27.3**	**21.2/27.2/ 27.7**	*20.0*/**26.2**	**20.0/21.3/ 25.8/26.2**
RIPIIA	**27.0**/*28.0*	17.9/**23.9/ 28.6**	**16.7**/17.8/ **21.2/27.3**	17.4/**26.6/ 28.0**	16.7/*26.0*	**27.0**
RPAIII	**16.3/16.5/ 25.0/27.1/** 29.4	21.6/**28.0**/ 28.6	**15.8/27.2**	*15.8*/**16.7/ 27.2/27.7**	*15.4*/26.2	**15.0/15.5/** 25.8
RPIAII	27.0	**16.3/28.0**	16.7/**27.3**	*15.8*/**27.2/ 27.7**	*27.5*	**25.8**
RPIIAI	16.3/16.5/ **22.9**/25.0/ 26.2/**27.0**	16.3/**17.1/ 21.6/22.9/ 28.6**	15.8/*16.6*/ **27.2**	**15.8**/16.7/ **21.1/22.4/ 27.2/27.7**	**15.5/21.3/** *26.2*/**27.4**	14.6/**15.0**/ 15.5/**20.0/ 21.3/25.8**
RPIIIA	16.3/**27.0**	**16.3**/17.9/ **23.9/28.6**	**16.7/17.1/ 27.2**	**17.4/23.3/ 25.0**/26.6/ 27.7	**15.5**/25.2/ 27.4	*16.3***/25.3**

Appendix 8. Complete list of family compositions with masses closer than 0.7 Da.

Family 1	Family 2	Mass 1	Mass 2	Δ
A1-I1	P1-S1	184.12117	184.08479	0.03638
I1-S1	A1-E1	200.11609	200.0797	0.03639
I2	E1-P1	226.16812	226.09535	0.07277
S2-P1	A2-E1	271.11682	271.11681	0.00001
P3-R1	S3-W1	447.25938	447.1754	0.08398
A2-R1-W1	E3-P1	484.25463	484.18053	0.0741
A2-R1-W1	E2-I2	484.25463	484.2533	0.00133
A1-R1-S1-W1	E3-I1	500.24955	500.21183	0.03772
S2-R1-W1	E4	516.24447	516.17036	0.07411
P4-R1	A2-E1-S1-W1	544.31214	544.22815	0.08399
P4-R1	S2-A1-I1-W1	544.31214	544.26454	0.0476
A3-R1-W1	E2-I1-P1-S1	555.29174	555.25403	0.03771
I3-E1-S1	A3-R1-W1	555.3268	555.29174	0.03506
I4-E1	A2-P1-R1-W1	581.37883	581.30739	0.07144
I3-E2	A1-P1-R1-S1-W1	597.33736	597.30231	0.03505
A1-I1-R1-S1-W1	E4-P1	613.33361	613.22312	0.11049
S2-P1-R1-W1	E3-I2	613.29723	613.29589	0.00134
A2-P2-E1-R1	S5-W1	621.32343	621.23946	0.08397
A4-R1-W1	E2-P2-S2	626.32885	626.25476	0.07409
A4-R1-W1	I2-S2-E1-P1	626.32885	626.32753	0.00132
I4-S2	A4-R1-W1	626.4003	626.32885	0.07145
P5-R1	A3-E1-I1-W1	641.3649	641.31729	0.04761
P5-R1	A2-I2-S1-W1	641.3649	641.35368	0.01122
I5-S1	A3-P1-R1-W1	652.45233	652.3445	0.10783
I5-E1	P2-A1-R1-S1-W1	694.46289	694.35507	0.10782
R4-A1	I2-W2-P1	695.44151	695.3795	0.06201
P4-R2	E3-I2-S1	700.41324	700.32792	0.08532
I4-E2	P2-S2-R1-W1	710.42142	710.34999	0.07143
R4-S1	W2-E1-I1-P1	711.43643	711.33803	0.0984
R4-S1	I3-W2	711.43643	711.4108	0.02563
R3-E2	W3-A1-P1	726.38848	726.3278	0.06068
R4-A2	P2-W2-I1-S1	766.47862	766.38023	0.09839
R3-S2-E1	A3-W3	771.40995	771.34926	0.06069
R3-E1-I1-S1	W3-A2-P1	797.46198	797.36491	0.09707
R4-S2	I2-W2-A1-E1	798.46846	798.40644	0.06202
R3-E2-S1	W3-A2-I1	813.42051	813.39621	0.0243
I4-W2	R4-A1-E1	824.49486	824.4841	0.01076
R3-E2-I1	W3-P2-S1	839.47254	839.37548	0.09706
E3-R3	W3-I1-P1-S1	855.43107	855.40678	0.02429
W3-I2-A1	E3-R3	855.44316	855.43107	0.01209

Appendix 9. Ac-D-Arg-library families found in kidney.

FAMILY	FORMULA	M-THEORIC	M-EXP	M/Z	Z
2211	P2-S2-I1-R1	696.39182	696.39107	697.39889	1
21111	A2-E1-I1-P1-R1	696.39181	696.39107	697.39889	1
2211	E2-P2-A1-R1	738.36599	738.36292	370.18928	2
411	P4-E1-R1	732.39181	732.39305	733.40087	1
321	P3-E2-R1	764.38164	764.37892	383.19728	2
2211	P2-R2-E1-S1	781.41942	781.41642	782.42424	1
42	A4-R2	655.38772	655.3916	656.39942	1
111111	A1-E1-I1-P1-R1-S1	712.38673	712.38771	713.39553	1
321	P3-S2-R1	680.36052	680.3611	681.36892	1
2211	A2-P2-E1-R1	680.36051	680.3611	681.36892	1
321	A3-R2-P1	681.40337	681.4078	682.41562	1
321	A3-R2-E1	713.3932	713.39887	714.40669	1
2211	R2-S2-A1-P1	713.39321	713.39887	714.40669	1
3111	E3-A1-I1-R1	786.38712	786.38464	394.20014	2
21111	P2-A1-E1-I1-R1	722.40746	722.40544	723.41326	1
21111	E2-A1-I1-P1-R1	754.39729	754.39397	755.40179	1
3111	P3-A1-I1-R1	690.41763	690.41723	691.42505	1
2211	E2-P2-I1-R1	780.41294	780.41044	391.21304	2
3111	E3-A1-P1-R1	770.35582	770.36178	771.3696	1
321	I3-R2-P1	807.54422	807.54701	808.55483	1
3111	P3-E1-R1-S1	722.37108	722.36798	362.19181	2
321	E3-W2-R1	974.42457	974.41502	975.42284	1
3111	P3-E1-I1-R1	748.42311	748.42486	749.43268	1
2211	P2-S2-E1-R1	712.35035	712.35578	713.3636	1
2211	A2-E2-P1-R1	712.35034	712.35578	713.3636	1
222	A2-R2-S2	687.37756	687.38094	688.38876	1
21111	A2-I1-P1-R1-W1	753.42853	753.42788	754.4357	1
2211	A2-R2-P1-S1	697.39829	697.40291	698.41073	1
321	E3-A2-R1	744.34017	744.34568	745.3535	1
2211	E2-S2-P1-R1	744.34018	744.34568	745.3535	1
3111	E3-I1-P1-R1	812.40277	812.39994	407.20779	2
411	R4-A1-P1	851.53135	851.53073	852.53855	1
2211	P2-R2-A1-S1	723.41394	723.40998	724.4178	1
411	E4-I1-R1	844.3926	844.39092	423.20328	2
321	A3-R2-S1	671.38264	671.38683	672.39465	1
2211	I2-P2-E1-R1	764.45441	764.45591	765.46373	1
2211	E2-P2-R1-S1	754.36091	754.35818	378.18691	2
51	A5-R1	570.32373	570.32937	571.33719	1
42	I4-R2	823.57552	823.57815	824.58597	1
21111	I2-E1-P1-R1-S1	754.43368	754.43228	755.4401	1
321	A3-E2-R1	686.33469	686.33984	687.34766	1
21111	S2-A1-E1-P1-R1	686.3347	686.33984	687.34766	1
321	P3-R2-S1	749.42959	749.42657	750.43439	1
21111	I2-A1-P1-R1-W1	795.47548	795.48067	796.48849	1
321	R3-P2-I1	834.52996	834.53057	835.53839	1
2211	E2-R2-I1-P1	839.46128	839.4651	840.47292	1
2211	E2-I2-R1-S1	786.42351	786.421	394.21832	2
2211	A2-R2-S1-W1	786.42484	786.421	394.21832	2
2211	P2-R2-A1-W1	822.46122	822.45958	412.23761	2
2211	I2-W2-R1-S1	900.49695	900.50171	901.50953	1
411	P4-I1-R1	716.43328	716.43194	717.43976	1
2211	P2-R2-E1-W1	880.4667	880.46678	881.4746	1
2211	A2-W2-R1-S1	816.40305	816.39531	817.40313	1
321	R3-E2-W1	971.50487	971.50754	486.76159	2

3111	R3-A1-I1-S1	798.49358	798.49372	400.25468	2
51	P5-R1	700.40198	700.40299	701.41081	1
3111	W3-A1-P1-R1	941.46598	941.46372	471.73968	2
3111	A3-P1-R1-S1	612.3343	612.3321	613.33992	1
3111	A3-E1-P1-R1	654.34486	654.34257	655.35039	1
2211	P2-S2-A1-R1	654.34487	654.34257	655.35039	1
21111	P2-A1-E1-R1-S1	696.35543	696.35624	697.36406	1
3111	S3-A1-I1-R1	660.35544	660.35834	661.36616	1
321	P3-A2-R1	648.37068	648.37091	649.37873	1
2211	I2-R2-A1-W1	854.52382	854.52322	428.26943	2
3111	R3-A1-I1-W1	897.54086	897.54216	898.54998	1
411	I4-A1-R1	738.51153	738.51584	739.52366	1
3111	I3-P1-R1-W1	837.52243	837.51666	838.52448	1
3111	W3-A1-E1-R1	973.45581	973.45124	487.73344	2
3111	S3-E1-P1-R1	702.32962	702.33393	703.34175	1
2211	A2-E2-R1-S1	702.32961	702.33393	703.34175	1
2211	I2-R2-A1-P1	765.49727	765.49986	766.50768	1
21111	W2-A1-I1-R1-S1	858.45	858.44356	859.45138	1
3111	R3-E1-P1-S1	840.46776	840.46323	841.47105	1
411	I4-E1-R1	796.51701	796.51019	797.51801	1
2211	A2-I2-R1-W1	769.45983	769.45955	770.46737	1
321	I3-R2-W1	896.57077	896.56591	897.57373	1
411	I4-R1-S1	754.50645	754.50033	755.50815	1
3111	I3-E1-P1-R1	780.48571	780.49215	781.49997	1
3111	E3-I1-R1-W1	901.42932	901.42213	902.42995	1
321	R3-P2-W1	907.52521	907.51631	908.52413	1
3111	A3-E1-I1-R1	670.37616	670.3808	671.38862	1
21111	S2-A1-I1-P1-R1	670.37617	670.3808	671.38862	1
411	A4-R1-W1	685.36593	685.36565	686.37347	1
2211	I2-R2-A1-E1	797.4871	797.48996	798.49778	1
411	P4-R1-S1	690.38125	690.38643	691.39425	1
42	R4-S2	857.50554	857.51112	858.51894	1
3111	A3-R1-S1-W1	701.36085	701.36126	702.36908	1
3111	R3-E1-I1-W1	955.54634	955.5416	956.54942	1
2211	R2-W2-I1-P1	953.53472	953.53692	954.54474	1
321	S3-R2-A1	703.37248	703.37572	704.38354	1
321	S3-I2-R1	702.40239	702.39949	703.40731	1
42	E4-R2	887.40964	887.41264	444.71414	2
21111	E2-I1-R1-S1-W1	859.41876	859.41574	430.71569	2
411	S4-A1-R1	634.30341	634.2991	635.30692	1
3111	R3-A1-P1-W1	881.50956	881.50442	882.51224	1
2211	E2-R2-A1-P1	797.41433	797.4112	399.71342	2
2211	I2-R2-P1-S1	781.49219	781.48892	782.49674	1
222	R2-S2-W2	917.46196	917.45912	459.73738	2
321	R3-S2-A1	772.44155	772.44509	773.45291	1
3111	P3-A1-E1-R1	706.37616	706.38031	707.38813	1
411	P4-A1-R1	674.38633	674.38707	675.39489	1
3111	S3-I1-P1-R1	686.37109	686.37309	687.38091	1
21111	A2-E1-I1-R1-S1	686.37108	686.37309	687.38091	1
411	R4-P1-W1	966.57355	966.57642	484.29603	2
2211	R2-S2-I1-W1	844.46671	844.46547	845.47329	1
2211	R2-S2-E1-W1	860.42524	860.42384	431.21974	2
2211	I2-P2-R1-S1	722.44385	722.45079	723.45861	1
21111	S2-A1-E1-I1-R1	702.366	702.3683	703.37612	1
42	P4-R2	759.45032	759.4504	760.45822	1
321	E3-S2-R1	776.33001	776.33504	777.34286	1
411	S4-R1-W1	749.34561	749.34572	375.68068	2
321	R3-S2-W1	887.48375	887.48802	888.49584	1
3111	R3-A1-E1-I1	840.50414	840.50309	841.51091	1

2211	S2-W2-A1-R1	832.39797	832.39486	833.40268	1
2211	P2-W2-I1-R1	894.48638	894.48302	448.24933	2
2211	E2-S2-R1-W1	833.36673	833.3631	834.37092	1
321	P3-R2-E1	791.44015	791.43822	396.72693	2
411	P4-R1-W1	789.42853	789.42562	395.72063	2
411	E4-R1-W1	917.38785	917.3806	918.38842	1
321	R3-W2-A1	970.53611	970.53262	486.27413	2
21111	E2-A1-I1-R1-S1	744.37656	744.37306	373.19435	2

Appendix 10. Ac-D-Arg-library families found in liver.

FAMILY	FORMULA	M-THEORIC	M-EXP	M/Z	Z
2211	P2-S2-I1-R1	696.39182	696.39107	697.39889	1
21111	A2-E1-I1-P1-R1	696.39181	696.39107	697.39889	1
2211	E2-P2-A1-R1	738.36599	738.36292	370.18928	2
411	P4-E1-R1	732.39181	732.39305	733.40087	1
321	P3-E2-R1	764.38164	764.37892	383.19728	2
2211	P2-R2-E1-S1	781.41942	781.41642	782.42424	1
42	A4-R2	655.38772	655.3916	656.39942	1
111111	A1-E1-I1-P1-R1-S1	712.38673	712.38771	713.39553	1
321	P3-S2-R1	680.36052	680.3611	681.36892	1
2211	A2-P2-E1-R1	680.36051	680.3611	681.36892	1
321	A3-R2-P1	681.40337	681.4078	682.41562	1
321	A3-R2-E1	713.3932	713.39887	714.40669	1
2211	R2-S2-A1-P1	713.39321	713.39887	714.40669	1
3111	E3-A1-I1-R1	786.38712	786.38464	394.20014	2
21111	P2-A1-E1-I1-R1	722.40746	722.40544	723.41326	1
21111	E2-A1-I1-P1-R1	754.39729	754.39397	755.40179	1
3111	P3-A1-I1-R1	690.41763	690.41723	691.42505	1
2211	E2-P2-I1-R1	780.41294	780.41044	391.21304	2
3111	E3-A1-P1-R1	770.35582	770.36178	771.3696	1
321	I3-R2-P1	807.54422	807.54701	808.55483	1
3111	P3-E1-R1-S1	722.37108	722.36798	362.19181	2
321	E3-W2-R1	974.42457	974.41502	975.42284	1
3111	P3-E1-I1-R1	748.42311	748.42486	749.43268	1
2211	P2-S2-E1-R1	712.35035	712.35578	713.3636	1
2211	A2-E2-P1-R1	712.35034	712.35578	713.3636	1
222	A2-R2-S2	687.37756	687.38094	688.38876	1
21111	A2-I1-P1-R1-W1	753.42853	753.42788	754.4357	1
2211	A2-R2-P1-S1	697.39829	697.40291	698.41073	1
321	E3-A2-R1	744.34017	744.34568	745.3535	1
2211	E2-S2-P1-R1	744.34018	744.34568	745.3535	1
3111	E3-I1-P1-R1	812.40277	812.39994	407.20779	2
411	R4-A1-P1	851.53135	851.53073	852.53855	1
2211	P2-R2-A1-S1	723.41394	723.40998	724.4178	1
411	E4-I1-R1	844.3926	844.39092	423.20328	2
321	A3-R2-S1	671.38264	671.38683	672.39465	1
2211	I2-P2-E1-R1	764.45441	764.45591	765.46373	1
2211	E2-P2-R1-S1	754.36091	754.35818	378.18691	2
51	A5-R1	570.32373	570.32937	571.33719	1
42	I4-R2	823.57552	823.57815	824.58597	1
21111	I2-E1-P1-R1-S1	754.43368	754.43228	755.4401	1
321	A3-E2-R1	686.33469	686.33984	687.34766	1
21111	S2-A1-E1-P1-R1	686.3347	686.33984	687.34766	1
321	P3-R2-S1	749.42959	749.42657	750.43439	1
21111	I2-A1-P1-R1-W1	795.47548	795.48067	796.48849	1
321	R3-P2-I1	834.52996	834.53057	835.53839	1
2211	E2-R2-I1-P1	839.46128	839.4651	840.47292	1

2211	E2-I2-R1-S1	786.42351	786.421	394.21832	2
2211	A2-R2-S1-W1	786.42484	786.421	394.21832	2
2211	P2-R2-A1-W1	822.46122	822.45958	412.23761	2
2211	I2-W2-R1-S1	900.49695	900.50171	901.50953	1
411	P4-I1-R1	716.43328	716.43194	717.43976	1
2211	P2-R2-E1-W1	880.4667	880.46678	881.4746	1
2211	A2-W2-R1-S1	816.40305	816.39531	817.40313	1
321	R3-E2-W1	971.50487	971.50754	486.76159	2
3111	R3-A1-I1-S1	798.49358	798.49372	400.25468	2
51	P5-R1	700.40198	700.40299	701.41081	1
3111	W3-A1-P1-R1	941.46598	941.46372	471.73968	2
3111	A3-P1-R1-S1	612.3343	612.3321	613.33992	1
3111	A3-E1-P1-R1	654.34486	654.34257	655.35039	1
2211	P2-S2-A1-R1	654.34487	654.34257	655.35039	1
21111	P2-A1-E1-R1-S1	696.35543	696.35624	697.36406	1
3111	S3-A1-I1-R1	660.35544	660.35834	661.36616	1
321	P3-A2-R1	648.37068	648.37091	649.37873	1
2211	I2-R2-A1-W1	854.52382	854.52322	428.26943	2
3111	R3-A1-I1-W1	897.54086	897.54216	898.54998	1
411	I4-A1-R1	738.51153	738.51584	739.52366	1
3111	I3-P1-R1-W1	837.52243	837.51666	838.52448	1
3111	W3-A1-E1-R1	973.45581	973.45124	487.73344	2
3111	S3-E1-P1-R1	702.32962	702.33393	703.34175	1
2211	A2-E2-R1-S1	702.32961	702.33393	703.34175	1
2211	I2-R2-A1-P1	765.49727	765.49986	766.50768	1
21111	W2-A1-I1-R1-S1	858.45	858.44356	859.45138	1
3111	R3-E1-P1-S1	840.46776	840.46323	841.47105	1
411	I4-E1-R1	796.51701	796.51019	797.51801	1
2211	A2-I2-R1-W1	769.45983	769.45955	770.46737	1
321	I3-R2-W1	896.57077	896.56591	897.57373	1
411	I4-R1-S1	754.50645	754.50033	755.50815	1
3111	I3-E1-P1-R1	780.48571	780.49215	781.49997	1
3111	E3-I1-R1-W1	901.42932	901.42213	902.42995	1
321	R3-P2-W1	907.52521	907.51631	908.52413	1
3111	A3-E1-I1-R1	670.37616	670.3808	671.38862	1
21111	S2-A1-I1-P1-R1	670.37617	670.3808	671.38862	1
411	A4-R1-W1	685.36593	685.36565	686.37347	1
2211	I2-R2-A1-E1	797.4871	797.48996	798.49778	1
411	P4-R1-S1	690.38125	690.38643	691.39425	1
42	R4-S2	857.50554	857.51112	858.51894	1
3111	A3-R1-S1-W1	701.36085	701.36126	702.36908	1
3111	R3-E1-I1-W1	955.54634	955.5416	956.54942	1
2211	R2-W2-I1-P1	953.53472	953.53692	954.54474	1
321	S3-R2-A1	703.37248	703.37572	704.38354	1
321	S3-I2-R1	702.40239	702.39949	703.40731	1
42	E4-R2	887.40964	887.41264	444.71414	2
21111	E2-I1-R1-S1-W1	859.41876	859.41574	430.71569	2
411	S4-A1-R1	634.30341	634.2991	635.30692	1
3111	R3-A1-P1-W1	881.50956	881.50442	882.51224	1
2211	E2-R2-A1-P1	797.41433	797.4112	399.71342	2
2211	I2-R2-P1-S1	781.49219	781.48892	782.49674	1
222	R2-S2-W2	917.46196	917.45912	459.73738	2
321	R3-S2-A1	772.44155	772.44509	773.45291	1
3111	P3-A1-E1-R1	706.37616	706.38031	707.38813	1
411	P4-A1-R1	674.38633	674.38707	675.39489	1
3111	S3-I1-P1-R1	686.37109	686.37309	687.38091	1
21111	A2-E1-I1-R1-S1	686.37108	686.37309	687.38091	1
411	R4-P1-W1	966.57355	966.57642	484.29603	2
2211	R2-S2-I1-W1	844.46671	844.46547	845.47329	1

2211	R2-S2-E1-W1	860.42524	860.42384	431.21974	2
2211	I2-P2-R1-S1	722.44385	722.45079	723.45861	1
21111	S2-A1-E1-I1-R1	702.366	702.3683	703.37612	1
42	P4-R2	759.45032	759.4504	760.45822	1
321	E3-S2-R1	776.33001	776.33504	777.34286	1
411	S4-R1-W1	749.34561	749.34572	375.68068	2
321	R3-S2-W1	887.48375	887.48802	888.49584	1
3111	R3-A1-E1-I1	840.50414	840.50309	841.51091	1
2211	S2-W2-A1-R1	832.39797	832.39486	833.40268	1
2211	P2-W2-I1-R1	894.48638	894.48302	448.24933	2
2211	E2-S2-R1-W1	833.36673	833.3631	834.37092	1
321	P3-R2-E1	791.44015	791.43822	396.72693	2
411	P4-R1-W1	789.42853	789.42562	395.72063	2
411	E4-R1-W1	917.38785	917.3806	918.38842	1
321	R3-W2-A1	970.53611	970.53262	486.27413	2
21111	E2-A1-I1-R1-S1	744.37656	744.37306	373.19435	2

Appendix 11. Ac-D-Arg-library families found in brain.

FAMILY	FORMULA	M-THEORIC	M-EXP	M/Z	Z
3111	P3-I1-R1-S1	706.41255	706.41943	707.42725	1
21111	E2-A1-P1-R1-S1	728.34526	728.34498	365.18031	2
2211	A2-P2-I1-R1	664.40198	664.39975	665.40757	1
222	A2-R2-S2	687.37756	687.37959	688.38741	1
3111	P3-A1-I1-R1	690.41763	690.42234	691.43016	1
321	P3-A2-R1	648.37068	648.37476	649.38258	1
222	A2-P2-R2	707.41902	707.42289	708.43071	1
21111	P2-I1-R1-S1-W1	795.4391	795.43546	398.72555	2
2211	P2-W2-E1-R1	910.44491	910.43915	911.44697	1
21111	E2-A1-I1-P1-R1	754.39729	754.40064	378.20814	2
411	S4-I1-R1	676.35036	676.35167	677.35949	1
2211	*A2-E2-I1-R1*	*728.38164*	*728.38472*	*365.20018*	*2*
21111	*S2-E1-I1-P1-R1*	*728.38165*	*728.38472*	*365.20018*	*2*
2211	A2-P2-R1-S1	638.34995	638.35556	639.36338	1
2211	A2-R2-I1-P1	723.45032	723.44393	724.45175	1
321	S3-R2-A1	703.37248	703.37345	704.38127	1
2211	P2-R2-A1-E1	765.4245	765.42098	383.71831	2
21111	I2-E1-P1-R1-S1	754.43368	754.42981	755.43763	1
411	R4-A1-P1	851.53135	851.52741	852.53523	1
411	P4-R1-S1	690.38125	690.38564	691.39346	1
2211	I2-S2-E1-R1	744.41295	744.41648	373.21606	2
21111	R2-A1-P1-S1-W1	812.44049	812.44212	407.22888	2
21111	W2-A1-E1-P1-R1	884.42926	884.43232	443.22398	2
3111	S3-E1-I1-R1	718.36092	718.36423	719.37205	1
2211	E2-R2-I1-P1	839.46128	839.46667	840.47449	1
321	R3-W2-E1	1028.54159	1028.5501	1029.55792	1
411	R4-A1-E1	883.52118	883.52464	442.77014	2
2211	E2-R2-A1-W1	886.44088	886.43906	444.22735	2
321	E3-R2-W1	944.44636	944.44208	473.22886	2
2211	I2-S2-R1-W1	801.44967	801.45678	802.4646	1
21111	I2-A1-P1-R1-W1	795.47548	795.47334	796.48116	1
21111	R2-A1-E1-P1-S1	755.40377	755.4042	378.70992	2
321	S3-R2-I1	745.41943	745.41976	373.7177	2
222	*A2-R2-W2*	*885.47212*	*885.46914*	*443.74239*	*2*
2211	*E2-I2-R1-W1*	*885.47079*	*885.46914*	*443.74239*	*2*
321	W3-E2-R1	1031.46129	1031.45836	516.737	2
411	E4-P1-R1	828.3613	828.36414	415.18989	2
321	*S3-R2-P1*	*729.38813*	*729.38874*	*365.70219*	*2*

2211	*A2-R2-E1-S1*	*729.38812*	*729.38874*	*365.70219*	*2*
321	*S3-P2-R1*	*670.33979*	*670.33655*	*671.34437*	*1*
21111	*A2-E1-P1-R1-S1*	*670.33978*	*670.33655*	*671.34437*	*1*
321	A3-P2-R1	622.35503	622.3576	623.36542	1
3111	R3-E1-P1-S1	840.46776	840.47087	841.47869	1
2211	S2-W2-E1-R1	890.40345	890.40654	446.21109	2
21111	R2-E1-I1-S1-W1	886.47727	886.47402	444.24483	2
3111	W3-E1-P1-R1	999.47146	999.47522	500.74543	2
321	A3-R2-W1	770.42992	770.43086	386.22325	2
2211	E2-I2-A1-R1	770.42859	770.43086	386.22325	2
42	S4-R2	719.3674	719.37024	720.37806	1
321	R3-W2-I1	1012.58306	1012.585	507.30032	2
42	E4-R2	887.40964	887.41194	444.71379	2
3111	A3-P1-R1-S1	612.3343	612.33682	613.34464	1
2211	E2-S2-I1-R1	760.37148	760.3746	381.19512	2
3111	E3-I1-P1-R1	812.40277	812.40376	407.2097	2
3111	A3-R1-S1-W1	701.36085	701.36348	702.3713	1
3111	P3-A1-E1-R1	706.37616	706.38132	707.38914	1
2211	*P2-S2-E1-R1*	*712.35035*	*712.34962*	*713.35744*	*1*
2211	*A2-E2-P1-R1*	*712.35034*	*712.34962*	*713.35744*	*1*
3111	R3-A1-E1-S1	814.45211	814.45178	408.23371	2
51	P5-R1	700.40198	700.40583	701.41365	1
21111	R2-A1-E1-S1-W1	844.43032	844.43266	845.44048	1
321	P3-W2-R1	878.45508	878.4546	440.23512	2
2211	E2-W2-P1-R1	942.43474	942.43116	472.2234	2
51	W5-R1	1145.53473	1145.5359	573.77577	2
321	S3-R2-E1	761.37796	761.38422	762.39204	1
3111	A3-I1-R1-S1	628.3656	628.35952	629.36734	1
321	S3-W2-R1	848.39289	848.39633	849.40415	1
21111	P2-E1-I1-R1-S1	738.40238	738.40178	370.20871	2