**Iana Lychko**

Licenciada em Bioquímica

# Design and production of peptide-based scaffolds for bioengineering applications

Dissertação para obtenção do Grau de Mestre em

Biotecnologia

Orientadora: Prof.ª Doutora Ana Cecília Afonso Roque, FCT-NOVA, Portugal

Co-orientadora: Doutora Olga Iranzo Casanova, iSm2, Aix-Marseille Univesité, França

Júri

Presidente: Prof. Doutor Rui Manuel Freitas Oliveira, FCT-NOVA

Arguente: Doutora Ana Luísa Moreira de Carvalho, FCT-NOVA

Vogal: Prof.ª Doutora Ana Cecília Afonso Roque, FCT-NOVA

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro de 2018

**LOMBADA**

**Design and production of peptide-based scaffolds for bioengineering applications**

**Iana Lychko**

**2018**

**Iana Lychko**

Licenciada em Bioquímica

# Design and production of peptide-based scaffolds for bioengineering applications

Dissertação para obtenção do Grau de Mestre em

Biotecnologia

Orientadora: Prof.ª Doutora Ana Cecília Afonso Roque, FCT-NOVA, Portugal

Co-orientadora: Doutora Olga Iranzo Casanova, iSm2, Aix-Marseille Univesité, França

Júri

Presidente: Prof. Doutor Rui Manuel Freitas Oliveira, FCT-NOVA

Arguente: Doutora Ana Luísa Moreira de Carvalho, FCT-NOVA

Vogal: Prof.ª Doutora Ana Cecília Afonso Roque, FCT-NOVA

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro de 2018

# Design and production of peptide-based scaffolds for bioengineering applications

**“Copyright”**

Iana Lychko

Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

# Agradecimentos

Em primeiro lugar gostaria de agradecer à minha orientadora Professora Cecília Roque, por me ter recebido neste projeto, por toda a orientação e o apoio durante este ano. Obrigada pelas palavras encorajadas que não me deixaram desistir nos momentos mais difíceis e pela confiança que depositaram em mim. Agradeço ainda pelas oportunidades que me proporcionou e que me ajudaram a crescer tanto ao nível profissional como pessoal.

Um grande agradecimento à minha co-orientadora, Dra. Olga Iranzo que me ensinou a síntese química dos péptidos. Obrigada pela sua constante disponibilidade durante a minha estadia em Marselha, pelos seus conselhos, sugestões e todas as questões que me colocava e que permitiram ter uma atitude autocrítica ao meu trabalho e melhorar, assim, o meu desempenho no laboratório.

Às pessoas que me ajudaram durante a minha estadia em Marselha, por me terem recebido com um bom ambiente e por me terem mostrado sítios bonitos dentro da cidade e arredores.

Um agradecimento enorme aos meus colegas do Biomolecular Engineering Group, pelo ambiente acolhedor, todos conselhos e toda ajuda com que contribuíram para a realização deste projeto. Em partícula à Raquel dos Santos, à Cláudia Fernandes e à Dra. Ana Pina, muito obrigada por tudo o que me ensinaram.

Gostaria de agradecer também ao Dr. Arménio Barbosa por todo suporte na parte computacional, foi um mundo novo para mim. Obrigada por todas as explicações e sugestões que ajudaram bastante na realização deste trabalho.

Um muito obrigado ao Gonçalo Teixeira que me acompanhou desde o início até ao fim deste projeto. Obrigada por toda a preocupação, todos conselhos e principalmente pelo suporte durante a expressão das proteínas, que sem dúvida foi indispensável.

Às minhas amigas Daniela, Beatriz, Carolina Francisco, Carolina Ferreira, Patrícia e à minha afilhada Beatriz pelos convívios animados, almoços produtivos, pelas aventuras e pela amizade. À Sara por todos os bons momentos e companhia, dentro e fora do laboratório. Obrigada pelas palavras de apoio sempre que estava com incertezas. Ao Fábio, pela companhia nos laboratórios quando ficava até tarde, pelos cafés, copos, conversas animadas e bons momentos cheios de risos.

À minha melhor amiga Vanessa, por me ter acompanhado durante todos anos os académicos e principalmente pelo todo teu apoio no último ano. Agradeço por estares sempre presente, pelas conversas descontraídas às tantas de manha, pelo suporte nos momentos mais difíceis e pela companhia nas corridas. Sem ti isto não ia ser tão fácil e animado.

Ao Tiago que me acompanhou durante maior parte do meu percurso académico. Obrigada pela paciência, compreensão, pelo teu esforço e dedicação para que tudo corresse bem. Obrigada pelo carinho e tua companhia que ajudam a melhorar o meu dia-a-dia.

Aos meus pais, à minha irmã e a toda a família, por acreditarem sempre em mim, por toda ajuda e esforço que me permitiu chegar até aqui. Obrigada pela compreensão e motivação para ser melhor e nunca desistir dos desafios.

Por fim, agradeço à minha avó, por tudo o que fez por mim.

# Abstract

Protein- and peptide-based affinity reagents have demonstrated a great potential in different bioengineering fields, including the identification and capture of target molecules with applications in purification and sensing.

This work focused on the study and production of cyclic β-hairpin peptides and Odorant-Binding Proteins (OBPs) as affinity reagents for application in bioseparation and biosensing, respectively.

Two cyclic β-hairpin peptides (cyclic-M3 and cyclic-M9) were previously designed by docking, as potential affinity reagents for phosphorylated peptides. Here, cyclic-M3 and cyclic-M9, as well as a control peptide cyclic-M0 were chemically synthetized and characterized through Mass Spectrometry, analytical HPLC and Circular Dichroism. To evaluate the binding affinity of cyclic peptides towards several phosphorylated peptides, binding studies were performed in solution, by the MicroScale Thermophoresis technique. Cyclic-M3 and cyclic-M9 interact with a phosphorylated peptide GK14P with $K_A$ of 1.0 $mM^{-1}$ and 1.34 $mM^{-1}$, respectively. In addition, the cyclic peptides were selective for the phosphorylated moieties.

Two rat OBPs (OBP2 and OBP3) were selected as experimental models for developing affinity reagents capable to detect specific volatile organic compounds (VOCs). Binding studies published until May 2018 reporting proteins selectivity and structural information were used to analyze structural characteristics involved in the natural binding of VOCs. Due to the lack in structural information for OBP2, homology modeling was employed to set a 3D structure. OBPs bind molecules with variable chemical and structural features mostly though hydrophobic interactions. However, the presence of determinant amino acid residues in the binding pockets (Lys113 for OBP2; Glu120 and Tyr122 for OBP3) increase the specificity of these proteins against VOCs. Both OBPs were successfully produced as soluble proteins using the *E. coli* expression system for further purification and biochemical characterization.

**Keywords**: Affinity reagents, synthesis, β-hairpin, OBP, scaffold

# Resumo

Reagentes de afinidade baseados em péptidos e proteínas, têm demostrado o seu grande potencial ao serem aplicados em distintas áreas da bioengenharia incluindo identificação e captura das moléculas alvo para aplicações em purificação e biossensores.

Este trabalho focou-se na produção e analise dos péptidos cíclicos com conformação em *β-hairpin* e proteínas ligadoras de odor (OBPs) como reagentes de afinidade para aplicações em biosseparação e em biossensores, respetivamente.

Dois péptidos (cyclic-M3 e cyclic-M9), foram previamente desenhados com técnicas de *molecular docking*, para a captura dos péptidos fosforilados. Os *β-hairpins* cyclic-M3, cyclic-M9 e o controlo, cyclic-M0, foram quimicamente sintetizados tendo sido caracterizados por Espectrometria de Massa, HPLC analítico e Dicroísmo Circular. De forma a avaliar a afinidade dos péptidos cíclicos para os diferentes péptidos fosforilados, efetuou-se uma análise com *MicroScale Thermophoresis*. Tendo se verificado que os cyclic-M3 e cyclic-M9 ligam-se ao péptido fosforilado GK14P com as $K_A$ de 1.0 $mM^{-1}$ e 1.34 $mM^{-1}$ respetivamente. Além disso, os peptídeos cíclicos foram seletivos para os alvos fosforilados.

Duas OBPs de rato (OBP2 e OBP3) foram selecionadas como modelos experimentais com o objetivo de desenvolver os reagentes de afinidade, capazes de detetar compostos orgânicos voláteis (VOCs). Estudos de ligação das OBPs aos diferentes VOCs publicados até a Maio de 2018 em conjunto com a informação estrutural disponível foram analisados de forma a compreender a especificidade e o mecanismo envolvido na ligação. O modelo da estrutura tridimensional da OBP2 foi construído recorrendo à modelação por homologia. As OBPs ligam VOCs com diferentes caraterísticas químicas e estruturais maioritariamente por ligações hidrofóbicas. No entanto, ambas as proteínas contêm resíduos de aminoácidos capazes de estabelecer pontos de hidrogénio (Lys113 em OBP2; Glu120 and Tyr122 em OBP3) aumentando assim a sua especificidade contra VOCs. A OBP2 e a OBP3 foram produzidas em frações solúveis via biológica, utilizando como o sistema de expressão *E. coli* para posterior purificação e caracterização bioquímica.

**Palavras-chaves**: Reagentes de afinidade, síntese, *β-hairpin*, OBP, *scaffold*

# Index of contents

# Index of figures

# Index of tables

# Abbreviations

1,8-ANS - 1-anilino-8-naphthalenesulfonate

1-AMA - 1-aminoanthracene

ACN – Acetonitrile

Amp – Ampicillin

BRCA1 – Breast Cancer 1 protein

BRCT – BRCA1 C-terminal

CD – Circular Dichroism

Cy5 – Cyanine 5 maleimide dye

Dap – Diaminopropionic acid

DCM – Dichloromethane

dd water – Distilled deionized water

DIEA – N,N-Diisopropylethylamine

DMF – N,N-Dimethylformamide

DTNB – 5,5'-dithio-bis-[2-nitrobenzoic acid

EDT – 1,2-Ethanedithiol

EDTA - Ethylenediaminetetraacetic acid

Ellman's reagente – 5,5'-Dithiobis(2-nitrobenzoic acid)

ESI – Electrospray Ionization

Fmoc - fluorenylmethyloxycarbonyl protecting group

HATU – O-(7-azabenzotriazol-1-yl)-N N N' N'-tetramethyluronium hexafluorophosphate

HBTU – 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate

HEPES - 4-(2-hydroxyethyl)-1-piperazine-1-ethanesulfonic acid

HPLC – High-performance liquid chromatography

IMAC – Immobilized Metal Affinity Chromatography

IPTG- Isopropyl β-D-1-thiogalactopyranoside

$K_A$ – Association equilibrium constant

$K_D$ – Dissociation equilibrium constant

LB – Luria Broth

MeOH – Methanol

MS – Mass Spectrometry

MST – MicroScale Thermophoresis

NMP – N-methyl-2-pyrrolidone

NMR – Nuclear Magnetic Resonance

OBP – Odorant Binding Protein

PBD – Phosphoprotein-binding domain

PBS - Phosphate Buffered Saline

pDNA – plasmid DNA

PSA – Ammonium persulfate

pSer – Phosphoserine

pThr – Phosphothreonine

pTyr- Phosphotyrosine

Rt – Retention time

RT – Room temperature

SDS – Sodium Dodecyl Sulfate

SPPS – Solid Phase Peptide Synthesis

TAE - Tris-acetate-EDTA

TCEP – tris(2-carboxyethyl) phosphin

TEMED – Tetramethylethylenediamine

TFA – Tetrafluoroacetic acid

UV – Ultraviolet

Vis – Visible

VOCs – Volatile Organic Compoun

# CHAPTER 1

# Literature Review

## Affinity ligands

Affinity reagents are well known due to their ability to bind a target with high specificity and high affinity. They have demonstrated their importance in identification and capture of target biomolecules, with applications in purification and detection/biosensing.

Antibodies are the mostly used affinity reagents in research, biotechnology and biopharmaceutical industries [1] since they can be developed to bind different target molecules with high specificity. They are also frequently used for diagnosis and therapy of some types of cancer and autoimmune diseases [2]. However, antibodies like other natural proteins, have several drawbacks associated with their application. Their high molecular weight limits tissue penetration when used in clinical treatments, thus higher doses are needed for efficient therapy. Furthermore, there are high costs associated with their production mostly due to glycosylation and folding issues which require mammalian systems [1]. Moreover, their application in bioseparation and biosensing is challenging due to the low stability under adverse chemical conditions (e.g. organic solvents, extreme pH, high salt concentration) and difficulties related to their immobilization onto solid supports [3]. Protein engineering uses a wide range of methods and technologies to develop alternative affinity ligands as engineered protein scaffolds and biomimetic ligands with improved features.

Three different approaches are generally used in protein engineering:

- The combinatorial approach, which involves the creation of large and random libraries that are further experimentally screened against targets, to select potential ligands. Libraries can be generated by chemical synthesis (for small molecules), by phage, ribosome or yeast display systems (for peptides and protein scaffolds), and by SELEX techniques ("systematic evolution of ligands by exponential enrichment" for oligonucleotide-ligands) [4].
- The rational approach, involves the rational design of ligands. The main purpose is to mimic the natural interaction between a biological ligand and the target molecule. This method is focused on the available functional and structural (NMR, X-ray

crystallography and homology data) information about the complex of interest, and usually results in biomimetic ligand [4].

- A combined approach uses both, rational design and combinatorial methods, to develop novel affinity ligands based on defined protein scaffolds.

## 1.1 Biomimetic ligands

The biological activity of naturally occurring proteins is defined by their three-dimensional structure. Their binding site consists mostly, of well-ordered motifs (α-helix, β-turn, γ-turns and β-strands), which allow to display in an appropriate way the hot spot amino acid residues. It is estimated that on an average 9.5% of proteins interfacial residues are hot spots that contribute more significantly to binding affinity by making a major input to the binding free energy [5].

Biomimetic ligands tend to mimic the recognition motifs involved in protein-protein or protein-receptor interactions and consist of small synthetic and engineered biological molecules.

Synthetic ligands, as the one based on Triazine-scaffold or Ugi-scaffold (Figure 1.1), are well recognized as affinity reagents for protein purification [6–8]. Rational design combined with synthesis and screening of combinatorial libraries, is the most common strategy adopted to develop such type of reagents. Ligands are designed *de novo* by incorporating novel functionalities on appropriate small scaffolds (e.g. triazine ring or di-amide derived from Ugi reaction) [9]. The libraries are synthetized by solid-phase combinatorial synthesis and screened against target to identify lead ligands [10]. The main advantages of completely synthetic ligands include the high stability towards chromatographic operation conditions, low-cost of production and easy immobilization onto solid supports [11]. However, their smaller contact area with the target molecule, may lead to lower affinity and specificity in comparison to biological ligands.

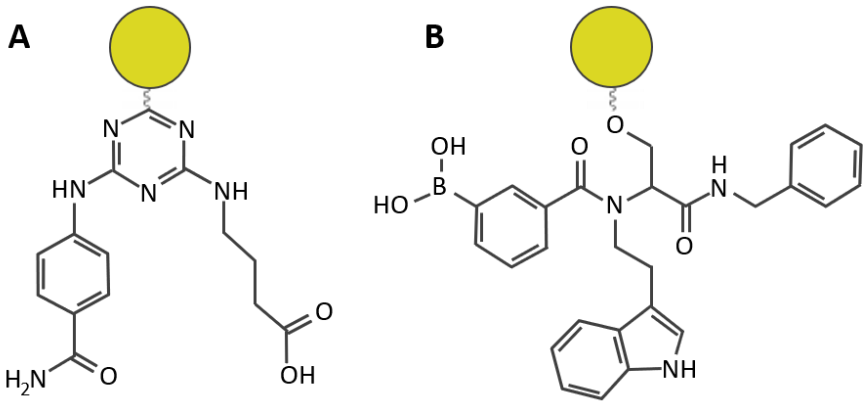


**Figure 1.1:** Examples of synthetic affinity ligands immobilized onto resin support (yellow circles). (A) Triazine-based protein L biomimetic ligand 8/7[12]; (B) Glycoprotein-binding Ugi ligand.A13C24I8 [13]

Biological biomimetic ligands are composed by peptides and oligonucleotides, that also are designed in order to mimic natural interactions. Some examples of natural biomimetic ligands are described in the following section.

## 1.2 Synthetic peptides

Peptide-based ligands have gained an interest as affinity reagents due to their wide functional and structural diversity as well as possibility of cost-effective production. The combinatorial approach allows to construct peptide libraries using, biological display systems [14] and chemical synthesis in solid and liquid phases [15]. Biological libraries have the advantage to generate a great number of potential ligands, nonetheless, they are limited by the diversity of natural amino acids. On the other hand, chemical synthesis allows the incorporation of a wide range of different synthetic groups and amino acid in the sequence, but the library size is reduced [16]. Still, both methodologies can be combined for ligand discovery and optimization [17].

Peptide-based affinity ligands can be designed *de novo* using rational approaches. Here, the binding mode between a target and a natural receptor is reproduced in a stable and appropriate scaffold that mimic the right spatial orientation of the side chains responsible for interaction.

Once the sequences of the lead peptides are identified, they can be modified to improve the biochemical features and binding to target [18]. Most of the modification strategies are based on the backbone and amino acid side chain variations which can stabilize or induce secondary structure and display in a correct way the energetically important groups [19]. For instance, the incorporation of non-natural amino acid residues, non-peptidic structural elements and N- and C-terminal, can increase ligands' stability in solution and protect them against degradation by proteases in physiological environment [19]. Several studies have demonstrated that linear peptides have lower biological activity in comparison to their cyclic counterparts [20–22], as a result of high natural conformational flexibility [19,23]. Thus, the restriction of peptides conformation is achieved through peptide cyclization or by incorporation of restricted structures [24]. The cyclization can be performed in four different manners (i) side chain-to-side chain; (ii) head-to-side chain; (iii) side chain-to tail and (iv) head-to-tail (C-terminus to N-terminus) [25].

The most common peptide-based structures are turns, α-helix, β-sheet [26], the β-hairpin folding (Figure 1.2-A) is a key recognition motif involved in protein-biomolecule interactions. It consists of two antiparallel β-strands stabilized by hydrogen bonds and connected by a turn/loop. β-hairpin mimetic ligands can be developed using a combinatorial method [27], however one of the simplest approaches consists in transplanting the protein β-hairpin loop onto a hairpin-inducing template. Such templates generally have β-turn inducing structural elements in the center of the peptide sequence [28,29]. Dipeptide $^{D}$Pro-$^{L}$Pro unit is frequently used as a template since it forms a rigid

type-ll´ β-turn which promote interaction between strands and consequently β-hairpin formation [30]. This approach was already successfully applied to design β-hairpin mimetic ligands to inhibit a serine protease [31] and to bind the bovine immunodeficiency virus (BIV) Tat protein [32].

There are several reports that show the potential of peptide ligands to be used in a wide range of fields, namely, as therapeutic agents [18,23,33], some of which are already commercially available [24]. Furthermore, they can be used in biomedicine [34–36] for developing new biomaterials for tissue engineering, in bionanotechnology for creating nanoparticles and nanotubes [37]. In addition, some of the peptide-based ligands are used in affinity-based purification techniques [38–40].

## 1.3 Aptamers

Aptamers are single chain DNA or RNA oligonucleotide ligands (Figure 1.2-B), that can fold into a unique three-dimensional structure to enable target molecular recognition with high specificity and affinity. They are selected from combinatorial libraries of synthetic nucleic acids using Systematic Evolution of Ligands by Exponential Enrichment (SELEX) technique. SELEX is based on the repeating of the selection rounds until the ligand with the desirable affinity will be obtained. Each round consists of three main steps: selection, separation and amplification [41].

Aptamers can be developed for any kind of targets such as small molecules [42,43] proteins [44,45] or cells [46]. Aptamers have several advantages in comparison to protein-based ligands: they are stable to proteases degradation, severe conditions (high pH, temperature and salt concentration), non-toxic, non-immunogenic and have reversible denaturation. Besides, their production, comparing to antibodies, is less expensive and less complex. Once the sequence is identified it can be produced via chemical synthesis. These ligands are also amenable to various chemical modifications, without affecting significantly their binding capacity, allowing their oriented immobilization onto solid supports, [47] increasing their stability, as well as introducing reporter molecules [48,49].

Despite multiple advantages, there are some drawbacks. Namely, the rapid degradation of aptamers by nucleases limits their application in biological samples. Ligands stability can be improved with some structural modifications, however this may increase the production costs. Another disadvantage is the restricted chemical diversity since aptamers contain only four different nucleic acids when comparing with the twenty natural amino acid residues used to develop protein- and peptide-based ligands [50].

## 1.4 Engineered protein scaffolds

Engineered protein scaffolds are tailor-made, small proteins (less than 20kDa[9]) that are characterized by containing a robust and rigid core to maintain protein folding, associated with

variable regions where the amino acids can be randomized without affecting stability [9]. These variable domains are generally loop regions of the scaffold that vary in sequence and length [1]. This type of ligands also has higher affinity and selectivity in contrast to small synthetic ligands, since an extended interface, formed by several exposed loops, allows to stabilize a higher number of interactions with the target. A recent review on the protein scaffolds used so far can be found [51–55,9].

The design of protein scaffolds includes two main steps: (i) the selection of the appropriate core protein and (ii) the generation of combinatorial libraries with randomized variable domains. Protein scaffolds are robust and have high thermal and chemical stability. They are less complex than antibodies, and their production should be possible in bacterial expression systems (as *Escherichia coli*), which is less expensive when comparing with eukaryotic systems. The absence of post-translational modifications, disulfide bonds or free cysteine residues is also favored [56] to achieve high expression levels of functional protein [57]. The modifications on variable loops can be performed randomly using different combinations of amino acid residues [58,59] or following rational approach with the available functional and structural information [60]. The combinatorial libraries are created, and the lead ligands are selected through biological display systems. Furthermore, the selected scaffold could be modified to improve their properties (stability, solubility, specificity) either through mutation of critical amino acid residues or side-chain structural modifications.

Owing to the diversity of proteins scaffolds and their advantages they can be used in a wide range of different applications. The possibility of their production by chemical synthesis, enables the introduction of structural modifications and new functionalities improving their immobilization and potential to be used for purification and bioseparation. The side-specific conjugation of labeling molecules [61] allow their application in biosensors, imaging and diagnosis.

Protein scaffolds can be derived from Immunoglobulin (Ig) fragments or non-Immunoglobulin proteins (non-Ig), and divided in globular proteins and repeat proteins. Some examples of protein-based scaffolds and their applications are described next.

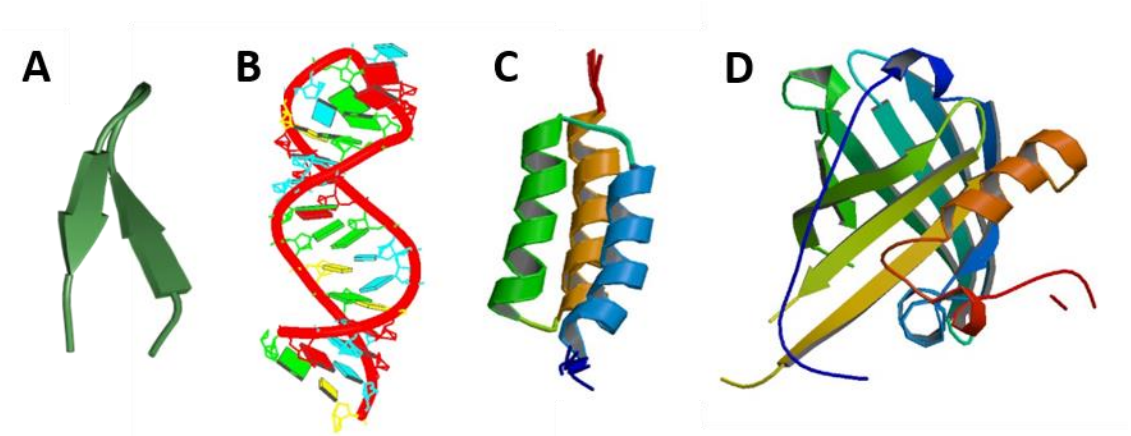


**Figure 1.2**: Examples of affinity ligands. (A) β-hairpin peptide (PDB:1N09 [62]); (B) Structure of RNA-based aptamer (PDB: 5OB3 [63]); (C) affibody (PDB:2B89 [64]); (D) lipocalin (PDB:1KXO [65])

### 1.4.1 Affibodies

Affibodies are non-Ig affinity ligands, derived from the Z-domain that is an engineered version of an Ig-binding domain (B-domain) of Protein A of *Staphylococcus aureus.* The Z-domain was chosen as template protein for engineering due to its attractive biophysical properties. It is a soluble, relatively short (58 amino acid residues, ~6.5kDa) and cysteine-free peptide, folded into a structure with three α-helices (Figure 1.2-C) [66]. The Z-domain has high thermal stability, reversible and fast folding. Additionally, it can be produced with high expression levels in *E. coli*. Affibody combinatorial libraries were already created using phage-display [67] and ribosome-display [68]. In this case, genetic randomization of 13 surface located residues situated in helix 1 and helix 2 were performed. The selection of residues was based on the available structural information of the complex B-domains of Protein A and human IgG. Affibodies with high affinity and specificity have been reported [67,69]. The small size of these ligands allows their production through solid-phase peptide synthesis, which enable the incorporation of different chemical groups [61].

Affibodies have shown promising results with protein capture microarrays [70]. Due to small size and possibility of side-specific radiolabeling they were tested as tracer molecules for imaging tumor cells [71,72]. In therapeutic applications, they can act by blocking the interactions between molecules [73] or can be fused to effector molecules and promote the target drug delivery [74]. Affibodies have shown a potential to be used for bioseparation as well [75,76].

### 1.4.2 Lipocalins

Lipocalins are soluble, extracellular proteins with a single polypeptide chain of 150-190 amino acids and high thermal stability [77]. Naturally, lipocalins occur in vertebrates, insects, plants and bacteria, and are involved in storage and transport of small, mostly hydrophobic organic

molecules in their barrel or loop regions [78]. Although lipocalins have a very low sequence homology, they share a very conserved three-dimensional folding that consists of a C-terminal α-helix and eight antiparallel β-strands connected by four loops (Figure 1.2-D). The β-strands form a rigid β-barrel scaffold that is highly conserved in different lipocalins and define a central polar cavity called calyx, while loops are hypervariable in sequence, conformation and length [79,80]. Some of the lipocalins were found to have post-translational modification (glycosylation), however it has been demonstrated that these modifications are not essential for production of fully functional proteins [81].

The first library that used a lipocalin scaffold was based on the bilin-binding protein (BBP) of *Pieris brassicae* and was produced by randomization of 16 amino acid residues positioned in loops and form its binding site (Figure 1.3) [82]. These modifications lead to the reshaping of the binding site and as a result, using phage-display, proteins with novel ligand-binding functions were selected. These engineered lipocalin-based ligands are also known as anticalins and can be produced using bacterial expression systems. Nowadays, anticalins derived from the human lipocalin 2 (Lcn2) scaffold are extensively explored as recognition elements. Anticalins with high affinity and selectivity for medically relevant protein targets have been demonstrated a potential to be used in therapy [83,84], imaging and diagnosis [85,86].

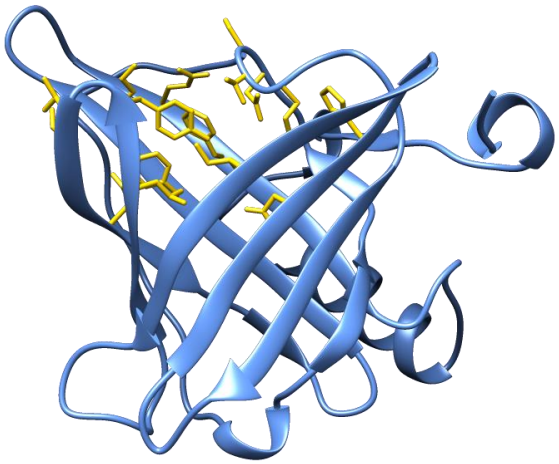

**Figure 1.3**: BBP with the 16 side-chains at the entrance of the ligand pocket that were subjected to the random mutagenesis highlighted (yellow) (PDB: 1BBP) [87].

## 1.5 Objectives

Affinity reagents have been used in a wide range of bioengineering applications, due to their versatile properties, of tunable affinity and specificity, as well as a capacity to reversibly bind to the target. The main goal of this dissertation was to explore the potential of peptide-based and protein-based affinity reagents to be used for bioseparation and for biosensing. The work plan is schematically represented in Figure 1.4.

The first part focused on the chemical production of small synthetic peptides, their characterization and study of binding affinity against phosphorylated target peptides, for a final bioseparation application. The synthetic peptides based on a β-hairpin scaffold, were previously designed to mimic the binding site of BRCA1-BRCT domain, which is known to bind phosphorylated peptides and proteins with a specific consensus sequence.

The aim of the second part was to analyze the potential of rat OBPs as scaffolds to generate affinity ligands able to bind different volatile organic compounds (VOCs). A literature search was performed to understand the binding mechanism and specificity of OBPs towards VOCs as well as the possibility to tune their binding features. Rat OBPs were also produced in bacterial expression systems to further study they affinity towards specific VOCs.

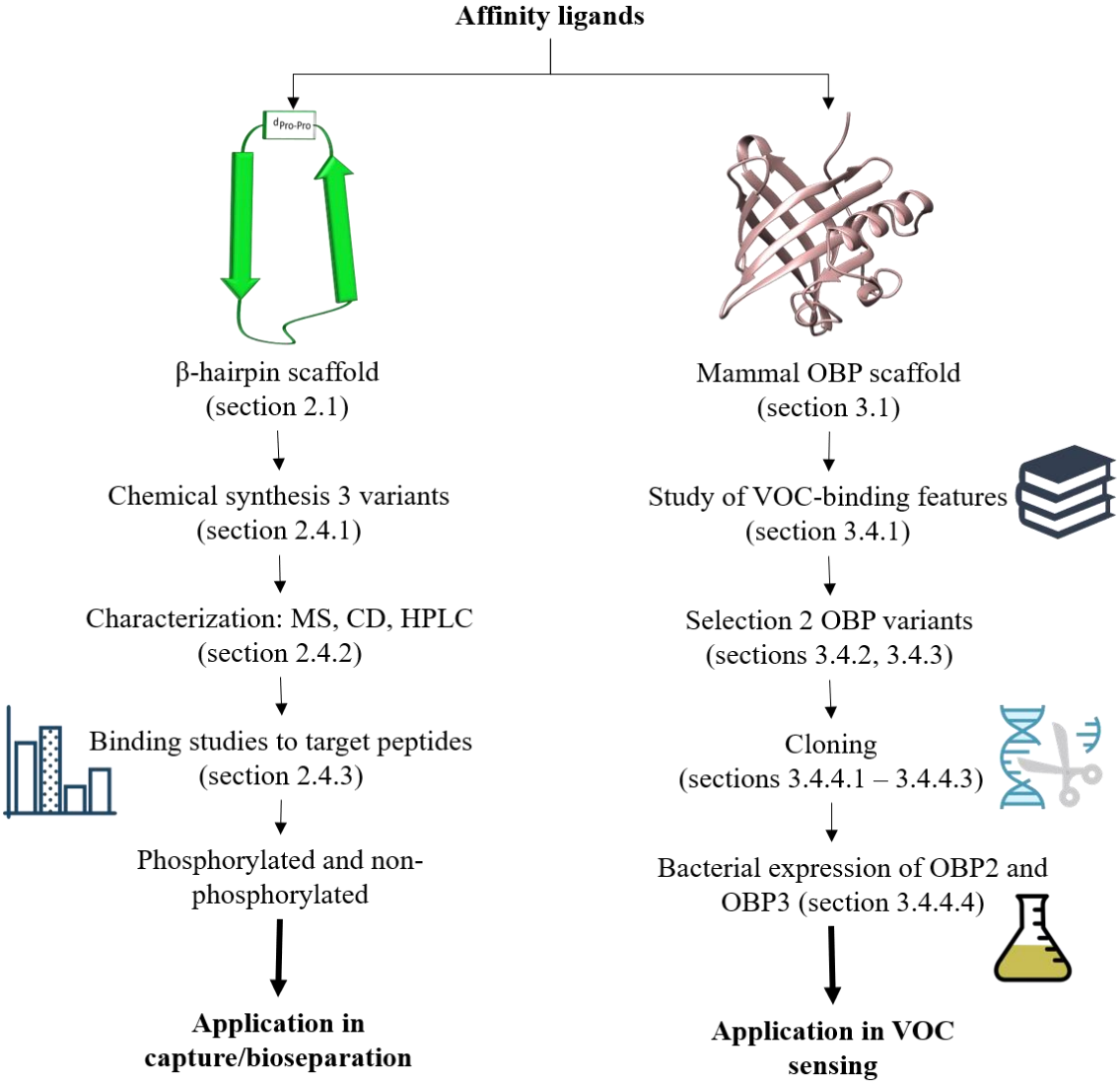


**Figure 1.4:** Schematic representation of the work planning.

# CHAPTER 2

# Chemical production and characterization of cyclic β-hairpin peptides as ligands for bioseparation

## 2.1 State of the art

Protein phosphorylation is an essential post-translational modification which regulates complex signaling networks and is involved in numerous basic cellular processes (gene expression, cell cycle, cell growth) [88]. Abnormal phosphorylation may lead to the development of several diseases namely cancer, metabolic and neurodegenerative disorders [89–91]. Therefore, the structural and genetic information about phosphorylated proteins is necessary to understand several disease mechanisms and develop tools for diagnosis and therapy. Currently, Mass Spectrometry strategies are widely used in the identification and characterization of phosphoproteins [92–94]. However, it is often necessary to enrich sample prior to analysis [94]. Conventional methods for phosphoprotein enrichment are based on the proteins' features resulting from the presence of a phosphate group. These include affinity-based techniques, selective chemical modification of the phosphate moieties, and immunoprecipitation [95,96]. Most part of these methods have low specificity and/or affinity and may involve addition steps leading to sample loss and increased costs [97], thus there is an interest in developing new solutions for the proteome enrichment.

As mentioned in Chapter 1, synthetic peptides have demonstrated a great potential to be used as affinity ligands for chromatographic purifications. The use of cyclic β-hairpin template to create affinity ligands has already been described in literature. Using phage-display, Delano and co-workers were able to identify a peptide-based ligand, Fc-III, that binds human antibody Fc domain [27]. The ligand FC-III had a β-hairpin conformation and was cyclized by a disulfide bridge. Lately, Dias and co-workers, increased significantly (80 folds) the affinity of this ligand by grafting the Fc-III loop onto a more stable β-hairpin inducing template containing a $^{D}$Pro-$^{L}$Pro unit [16].

In our laboratory, the cyclic β-hairpin peptides (Table 2.1), were previously rationally designed, as peptidomimetics of the C-terminal (BRCT) domain of the Breast Cancer Gene 1 (BRCA1) protein that recognizes the consensus sequence pSer-X-X-Phe (X-corresponded to any amino acid

residue). Mutations in this protein increase significantly the risk for breast and ovarian cancers [98] since BRCA1 is known to play an important role in the DNA damage response.

The interaction between the BRCT domain and the target phosphorylated peptide sequence occurs in a two-point binding mode (Figure 2.1). Firstly, the phosphate group interacts with the amino acid residues Gly1656 and Ser1655 through hydrogen bonds, and with Lys1702 by a salt-bridge. The phenylalanine residue, from the consensus sequence pSer-X-X-Phe, sits in a hydrophobic pocket at BRCT domain, composed by Phe1704, Met1775 and, is stabilized through interactions with Arg1699 (electrostatic interaction between Arg1699 side chain guanidium group and carbonyl of Phe residue and hydrogen bond between carbonyl group of Arg1699 and main chain NH of Phe) [98].

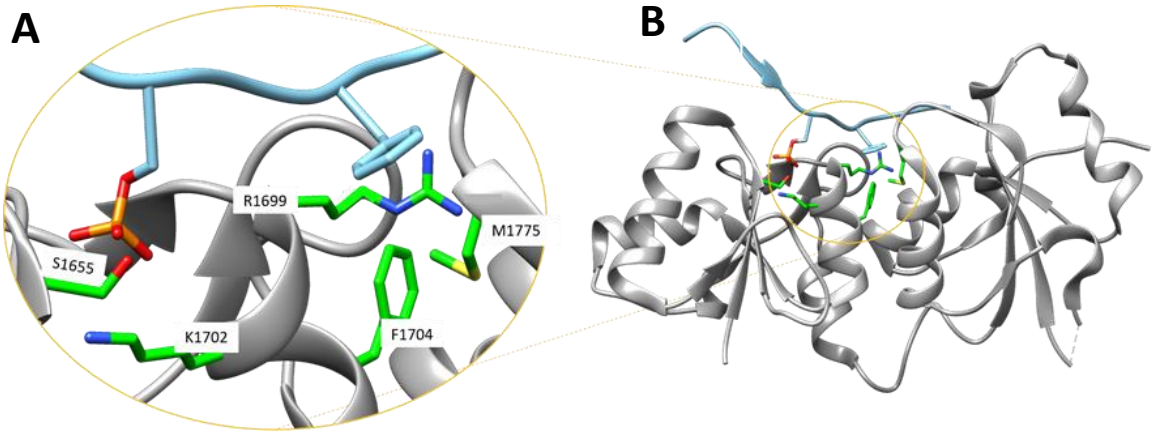


**Figure 2.1:** (A) Close-up view of the BRCA1-BRCT binding site with residues involved in the phosphor-peptide binding represented in green. (B) The BRCA1-BRCT domains (in grey ribbon) in complex with a phosphorylated GK14P (in blue ribbon) (PDB:1T2V [98]). Note: G1656 that stablish a hydrogen bond with phosphate group is not represented.

A 14-mer β-hairpin cyclic peptide (PDB:2NS4[79]; Figure 2.2) with a $^{D}$Pro-$^{L}$Pro unit, was used as a template where point mutations were introduced to mimic the binding mode of the BRCT domain with the phosphorylated consensus sequence. Lys at position 1, situated in the opposite site of the $^{D}$Pro-$^{L}$Pro, was mutated to a Cys to promote ligand immobilization onto solid support through sulfhydryl group. Arg at position 3 was mutated to Phe to mimic Phe1704 from BRCT. Arg at positions 4, 6, 10, 12 and 14 were mutated to Gly residues: Arg at position 10 and 12 for cyclic-M3 and cyclic-M9 respectively, were conserved to mimic the electrostatic and hydrogen bond interactions which are mainly derived from the crystallographic structure of BRCA1-BRCT domain between the backbone atoms of Ser1655 and Gly1656, as well as the side chain of Lys1702 and pSer moiety. Two Ile residues at position 5 and 7 were mutated to the non-natural amino acid diaminopropionic acid (Dap) to mimic the side chain interaction of Lys1702 and phosphate group of the target peptide.

**Table 2.1:** Sequences of the template and the designed peptides (cyclic-M3 and Cyclic-M9, cyclic-M0 was used as a control peptide.

| Peptide ID | Sequence | MW (g $mol^{-1}$) |
|---|---|---|
| Template | Cyclic($K_1$ $G_2$ $R_3$ $R_4$ $I_5$ $R_6$ $I_7$ $^{D}P_8$ $P_9$ $R_{10}$ $V_{11}$ $R_{12}$ $T_{13}$ $R_{14}$) | (PDB:2NS4) |
| Cyclic-M0 | Cyclic($C_1$ $G_2$ $G_3$ $G_4$ $G_5$ $G_6$ $G_7$ $^{D}P_8$ $P_9$ $G_{10}$ $V_{11}$ $G_{12}$ $T_{13}$ $G_{14}$) | 1010.4 |
| Cyclic-M3 | Cyclic($C_1$ $G_2$ **$F_3$** $G_4$ $I_5$ $G_6$ $I_7$ $^{D}P_8$ $P_9$ **$R_{10}$** $V_{11}$ $G_{12}$ $T_{13}$ $G_{14}$) | 1311.7 |
| Cyclic-M9 | Cyclic($C_1$ $G_2$ **$F_3$** $G_4$ **$Dap_5$** $G_6$ **$Dap_7$** $^{D}P_8$ $P_9$ $G_{10}$ $V_{11}$ **$R_{12}$** $T_{13}$ $G_{14}$) | 1257.6 |

Having in hand the *in silico* candidates (cyclic-M3, cyclic-M9 and cyclic-M0) in this work their production and characterization was performed. The chemical production of the cyclic peptidomimetics of BRCT domain is depicted in Figure 2.3. Fmoc-based solid-phase peptide synthesis (SPPS) strategy was used. It consists on the sequential addition of amino acid residues to a solid support. These amino acid residues have a Fmoc-protecting group at the N-terminal, as well as orthogonal side chain protecting groups to avoid side-chain reactions. The peptide assembly was performed from the C-terminal to the N-terminal. Linear precursors of peptides were synthetized, then they were cleaved from the resin, cyclized in a head-to-tail manner and deprotected by removing the side chain protecting groups. Crude peptides were then purified by preparative reversed-phase HPLC (RP-HPLC) and their purity determined by analytical RP-HPLC. They were characterized by Mass Spectrometry (MS), and, Circular Dichroism (CD). Finally, binding studies against phosphorylated and non-phosphorylated target peptides were performed by MicroScale Thermophoresis (MST) technique.

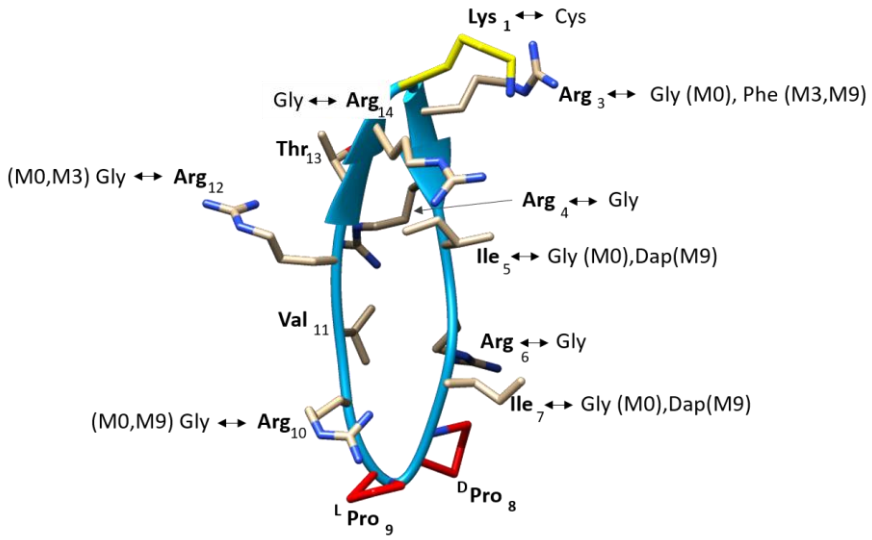


**Figure 2.2:** Structure representation of cyclic β-hairpin peptide used as a template (PDF: 2NS499) with labeled amino acid residues of the template (in bold) and residues used to mimic the BRCA1 binding site. Lys1 residue considered as anchoring point is colored yellow.

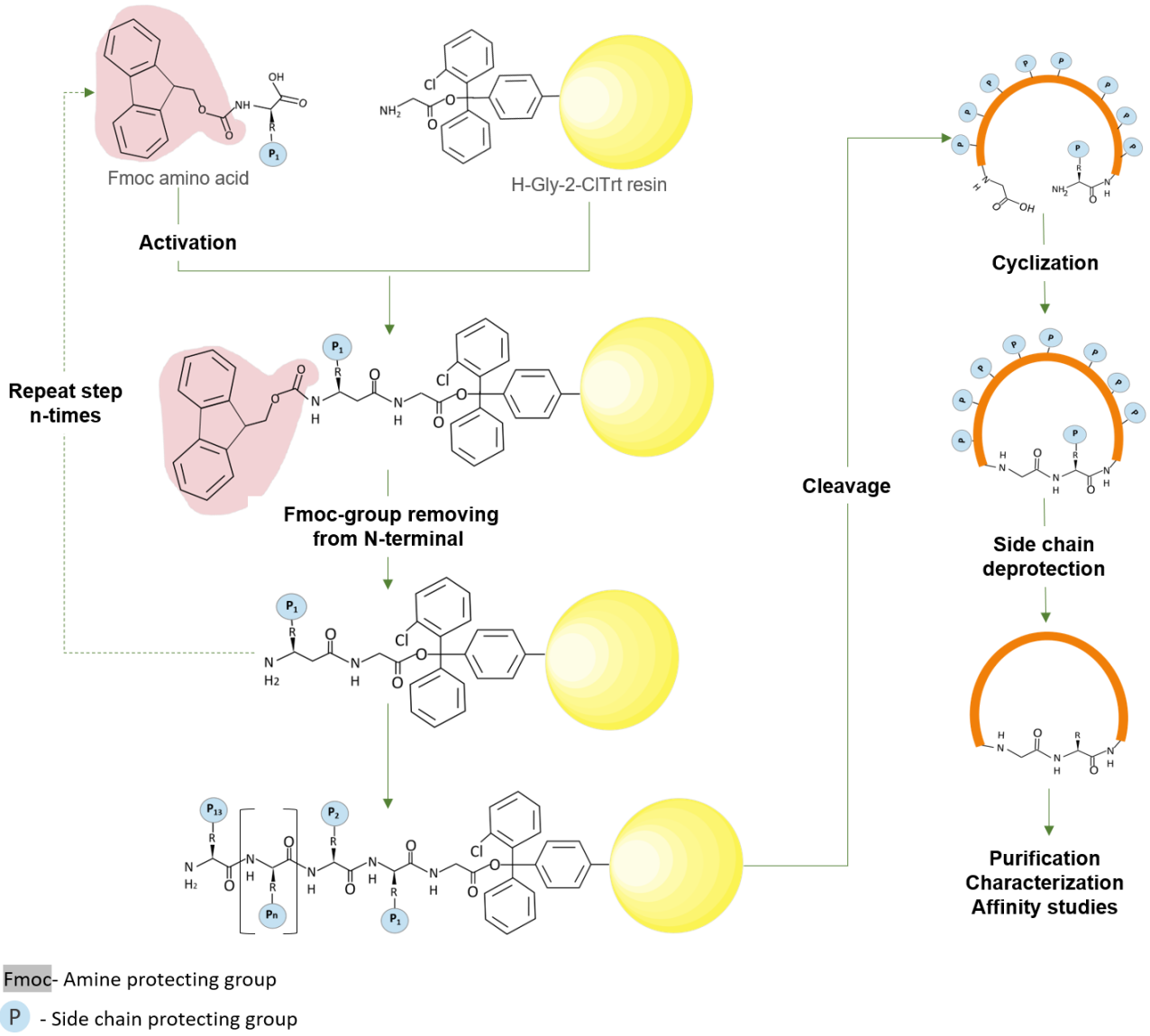


**Figure 2.3**: Schematic representation of synthesis of the cyclic β-hairpin peptidomimetics.

## 2.2 Materials

### 2.2.1 Chemical Reagents

All reagents were of the highest purity available and the solvents were HPLC-gradient. For chemical synthesis N-fluorenylmethoxycarbonyl (Fmoc)-protected amino acids (Fmoc-Arg(Pbf)-OH, Fmoc-Ala-OH, Fmoc-Cys(Trt)-OH, Fmoc-Dpr(Boc)-OH (Dap), Fmoc-D-Pro-OH, Fmoc-Gly-OH, Fmoc-Ile-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Thr(tBu)-OH, Fmoc-Val-OH), O-(7-azabenzotriazole-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HATU) and H-Gly-2-ClTrt resin were purchased from Novabiochem. 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), piperidine anisole, thioanisole, 1,2-ethanedithiol, acetonitrile (ACN), ethanol, N-methyl-2-pyrrolidone (NMP), chloroform, dichloromethane (DCM), methanol (MeOH) and diethyl-ether were acquired from Sigma-Aldrich. Trifluoroacetic acid (TFA) and N,N-dimethylformamide (DMF) were purchased from Roth, Acros Organics. N,N-diisopropylethylamine (DIEA) from Fisher Scientific. For Kaiser test ninhydrin, phenol, potassium cyanide(KCN) and pyridine were obtained from Sigma-Aldrich.

For CD 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) and sodium chloride (NaCl) were obtained from Sigma-Aldrich and PanReac AppliChem respectively. For MST assays and peptide labelling, cyanine 5-maleimide was obtained from Lumiprobe, tris(hydroxymethyl)aminomethane (TRIS) from NZYTech, bond-breaker tris(2-carboxyethyl)phosphine (TCEP) solution from Fisher Scientific. Ethylenediaminetetraacetic acid (EDTA), sodium phosphate dibasic heptahydrate and sodium phosphate monobasic monohydrate from Sigma-Aldrich. Sodium hydroxide (NaOH) and hydrochloric acid 37% were obtained from VWR and PanReac AppliChem respectively. Gly-Ala-Ala-Tyr-Asp-Ile-Ser-Gln-Val-Phe-Pro-Phe-Ala-Lys (GK14), Gly-Ala-Ala-Tyr-Asp-Ile-**p**Ser-Gln-Val-Phe-Pro-Phe-Ala-Lys (GK14P), **p**Ser-Gln-Val-Phe-Pro-Trp (SW6P), Ser-Gln-Val-Phe-Pro-Trp (SW6), **p**Tyr-Gly-Gly-Ile-Pro-Trp (YW6P), Tyr-Gly-Gly-Ile-Pro-Trp (YW6), Tyr-Ala-Gly-**p**Ser-**p**Thr-Asp-Glu-Asn-pThr-Asp-Ser-Glu-Trp (YW13P) and Tyr-Ala-Gly-Ser-Thr-Asp-Glu-Asn-Thr-Asp-Ser-Glu-Trp (YW13) peptides (**p**- is phosphorylated group) with N- and C- free terminals, were > 98% pure and were obtained from Genecust.

### 2.2.2 Equipment

For peptides synthesis an Automated Microwave Peptide Synthesizer Biotage® Initiator+ Alstra™ was used. Mass Spectrometry (MS) measurements were performed using a Water Synapt G2 HDMS (Manchester, UK) equipped with an ESI source at Spectropole, the Analytical Facility of Aix-Marseille Université. CD spectra were recorded using a Chirascan qCD spectropolarimeter

and MST measurements were done using the Monolith NT.115 (Nanotemper) instrument, at Laboratório de Análises, REQUIMTE and Biolab, Ucibio respectivly.

## 2.3 Methods

### 2.3.1 Mass spectrometry analysis

MS spectra were acquired, using a time-of-flight (TOF) mass spectrometer with electrospray ionisation (ESI). During the sample preparation crude linear peptides and pure cyclic peptides were dissolved in 300 μL of MeOH and then diluted (1:10) in solution of formic acid (1% in MeOH). Samples were ionized in positive electrospray mode and following parameters were used: ESI capillary voltage: +2.8 kV; extraction cone voltage: +20 V; desolvation gas ($N_2$) flow: 100 L $h^{-1}$.

### 2.3.2 Solid-Phase Peptide Synthesis of linear peptidomimetic precursors

#### 2.3.2.1 Synthesis of M0 and M9

The linear precursors of Cyclic-M0 and Cyclic-M9 (Table 2.1) were synthetized using Fmoc chemistry [100] on Automated Microwave Peptide Synthesizer. The resin used in both cases was 2-Chlorotrityl chloride resin (H-Gly-2-ClTrt) with loading capacity 0.63 mmol $g^{-1}$. Synthesis were performed in a 10 mL reactor vial with 0.25 mmol scale. All coupling reactions were carried out at 50°C for 6 min 30 sec to avoid the cleavage of the resin linker and, hence, a lower yield. Coupling reactions were carried out in DMF using 4 equivalents (equiv. relative to peptide) of Fmoc protected amino acids, 3.98 equiv. of HBTU (activation agent) and 8 equiv. of DIEA (activation base). Double coupling was used for the proline and d-proline residues. Two Fmoc removing reactions were performed using 20% piperidine in DMF (v/v) at room temperature (RT) for 3 and 10 min each time.

#### 2.3.2.2 Synthesis of M3

The linear precursors of Cyclic-M3 (Table 2.1) was synthetized manually, using also Fmoc chemistry and H-Gly-2-ClTrt resin. Synthesis was performed in a glass reactor with 0.63 mmol scale and all reactions were carried out at RT. All solutions were removed by vacuum filtration. Kaiser test (section 2.2.2.3) was performed after each coupling and deprotection reaction. Synthesis starts from pre-swelling of 1 g of dry resin with DCM (2 x 30 min) and washed with DMF (2 x 15 min). The following protocol was employed for each coupling cycle: (i) coupling mixture (2 equiv Fmoc protected amino acid, 1.95 equiv of HBTU, 4 equiv. of DIEA) was added and the reaction was stirred for 1 h; (ii) solution was removed and resin was washed with DMF (3 x 5 min); (iii) Fmoc removal reactions were performed using 20% piperidine in DMF (v/v) (2 x 15 min) and, the presence of Fmoc in the filtrate was verified by thin-layer chromatography

(TLC) with the system diethyl ether: hexane (3:2 v/v). Steps from (i) to (iii) were repeated until the last amino acid residue was coupled.

#### 2.3.2.3 Kaiser test

The following reagents were used: 5% ninhydrin in ethanol (w/v), 80% phenol in ethanol (w/v) and, 0.001 M potassium cyanide in pyridine (2 mL KCN in 98 mL pyridine). The standard protocol was employed for the Kaiser test [101]. Namely, 2 drops of each reagent were added to the, previously ethanol washed, resin beads. Then the sample was mixed well and heated to 120˚C for 5-6 min. Blue resin beads would indicate free amine. The test is based on the reaction between ninhydrin and primary amines.

### 2.3.3 Peptides cleavage and cyclization

Before proceeding with peptides cleavage and cyclization, the confirmation of peptides molecular weight was performed by electrospray ionization mass spectrometry (ESI-MS). For this, small aliquots of completely deprotected linear precursors, were obtained as follows. A small amount of each resin was treated with 2 mL of the deprotection solution: TFA/ thioanisole/ 1,2-ethanedithiol/ anisole (%v/v=90:5:3:2). The reaction was stirred for 2 h at RT. After, the resin beads were filtrated out and the TFA solution containing the deprotected peptide was concentrated using a stream of nitrogen gas. Cold diethyl ether was added to the residue to precipitate the peptide and then the pellet was separated by centrifugation (80000xg, 5 min, 4°C). This process was repeated several times.

The peptides were characterized by ESI-MS in positive mode and by direct infusion at a flow rate of 5 mL $min^{-1}$. The following peaks were detected (m/z): linear M0 (MW=1028.4 Da) $[M+H]^{+}$-1029.4 (calculated)/1029.4 (measured), $[M+H+K]^{2+}$-534.2/534.2, $[M+H+Na]^{2+}$-526.2/526.2 and $[M+2H]^{2+}$-515.2/515.2; linear M9 (MW = 1275.6 Da) $[M+2H]^{2+}$-638.8/638.8 and $[M+3H]^{3+}$-426.2/426.2; linear M3 (MW = 1329.7 Da) $[M+H]^{+}$-1330.7/1330.7, $[M+2H]^{2+}$-665.8/665.9, $[M+H+K]^{2+}$-684.8/684.8, $[M+2K+K]^{3+}$-456.9/456.9. For the latest two extra peaks were observed: $[M'+H]^{+}$-1217.6/1217.6 and $[M'+2H]^{2+}$-609.3/609.2 that correspond to a compound with MW'= 1216.6 Da.

The protected peptides were cleaved manually, from the resin by using mild acidic conditions. First, resin containing peptide was pre-swelled in DCM and afterwards short treatments (8 x 2 min), with TFA in DCM (10 mL, 1:99 % v/v) were performed. After each treatment the solution was collected by vacuum filtration in a glass tube and neutralized by adding 8-9 drops of DIEA. The presence of the peptide was verified by TLC using the system chloroform :methanol (90:10 %v/v). Fractions containing the peptide were combined and the DCM was removed in a

rotary evaporator. The oily residue was precipitated and washed several times with cold diethyl ether to obtain final white precipitate.

Cyclization was performed in DCM (25 mL were used for 0.25 mmol scale and 65 mL for 0.63 mmol scale), at RT and in the presence of 0.95 equiv. of HATU (DMF solution) and 12 equiv. of DIEA (NMP solution). All compounds were added to DCM in 4 steps with 30 min intervals to ensure high dilution conditions. The cyclization reaction was followed by analytical RP-HPLC and it was complete after 2-3 h. The reaction mixture was concentrated using a rotary evaporator. Next, liquid-liquid extraction with water (2x10 mL) was performed and the resultant organic phase was collected. Finally, DCM was removed in a rotary evaporator.

### 2.3.4 Deprotection

After cyclization, removal of the side chain protecting groups was performed by solubilizing the cyclic peptides in the deprotecting solution: TFA/thioanisole/1,2-ethanedithiol/anisole (90:5:3:2 %v/v). The reaction was stirred for 2 h at RT. Afterwards, TFA solution was concentrated using a stream of nitrogen gas and cold diethyl was added to precipitate the peptide. The peptide was separated from solution by centrifugation (80000 xg, 5min, 4°C). This process was repeated several times. The final crudes cyclic peptides were dissolved in MilliQ water, lyophilized and stored at -20°C until purification.

The cyclic crude peptides were characterized by ESI-MS. The following peaks were detected (m/z): cyclic-M0 (MW=1010.4) $[M+H+K]^{2+}$ 525.2 (calculated)/525.2 (measured), $[M+H]^{+}$ 1011.4/1011.5, $[M+Na]^{+}$ 1033.4/1033.4; cyclic-M9 (MW=1257.6) $[M+3H]^{3+}$ 420.2/420.2, $[M+2H]^{2+}$ 629.8/629.8, $[M+H]^{+}$ 1258.6/1258.6; cyclic-M3 (MW=1311.7) $[M+H]^{+}$ 1312.7/1312.7, $[M+2H]^{2+}$ 656.9/656.8, $[M+H+Na]^{2+}$ 667.8/667.9, $[M+H+K]^{2+}$ 675.8/675.8 and $[M+2H+K]^{3+}$ 450.9/450.9. As expected, peaks corresponded to cyclic 13-mer peptide (MW=1198.6 Da) were detected in the crude of cyclic-M3 (m/z): $[M'+H]^{+}$ 1199.6/1199.6, $[M'+2H]^{2+}$ 600.3/600.3 and $[M'+H+Na]^{2+}$ 611.3/611.3.

### 2.3.5 Purification of cyclic peptidomimetics

Purification was carried out by preparative RP-HPLC using a C12 column (Phenomenex Jupiter Proteo column, 250 nm x 4.6 mm, 4 µm, 90 Å) using solvent A (water/TFA, 99.9:0.1% v/v) and solvent B (ACN/water/TFA, 90:9.9:0.1 % v/v). The HPLC chromatograms were obtained by monitoring absorbance at 220 nm and 280 nm. All peptides were eluted from the column with the flow rate of 10 mL $min^{-1}$. Cyclic-M0 was eluted from the column with the linear gradient 15-30% of solvent B in 20 min (Retention time (Rt) = 16.6 min); Cyclic-M3 was eluted from the column with a linear gradient 30- 42% of B in 24 min (Rt = 23.1 min); Cyclic-M9 was eluted from the

column with a linear gradient 15-30% of B in 20 min (Rt = 16.5 min). All peptides were lyophilized and stored at -20°C. For cyclic-M0 and cyclic-M9 ~10 mg and ~60 mg of purified peptides were recovered respectively. For cyclic-M3 ~85 mg of pure peptide was obtained. The overall yields obtained were 3.9%, 10.5% and 20.4% for cyclic-M0, cyclic-M3 and cyclic-M9 respectively.

The purity of the peptides was checked by analytical RP-HPLC with a C18 column (Zorbax SB-C18 5.0 µm, 4.6 x150mm, 80 Å) using the same solvent A and solvent B system and gradient method optimized for this system.

The pure cyclic peptides were characterized by ESI-MS. The following peaks were detected: cyclic-M0 (MW=1010.4 Da) $[M+H]^+$ 1011.4 (calculated)/1011.5 (measured), $[M+2H]^{2+}$ 506.2/506.3, $[M+H+Na]^{2+}$ 517.2/517.2, $[M+H+K]^{2+}$ 525.2/525.2, $[M+NH4]^+$ 1028.4/1028.5 and $[M+Na]^+$ 1033.4/1033.5; cyclic-M3 (MW=1311.7 Da) $[M+H]^+$ 1312.7/1312.7, $[M+H+Na]^{2+}$ 667.8/667.9, $[M+2H]^{2+}$ 656.9/656.9, $[M+2H+K]^{3+}$ 450.9/450.9; cyclic-M9 (MW=1257.6 Da) $[M+3H]^{3+}$ 420.2/420.2 and $[M+2H]^{2+}$ 629.8/629.8.

### 2.3.6 Circular Dichroism

To analyze folding of the cyclic peptides and their stability in previously stablished binding (50 mM HEPES, pH 8.0) and elution conditions (50 mM HEPES, 250 mM NaCl, pH 8.0), far ultraviolet (UV) CD spectra of the cyclic peptides were recorded using a Chirascan qCD spectropolarimeter. The data were acquired at 23 °C, in the wavelength scan mode, using 1 nm bandwidth with a step size of 1 nm. Scans were acquired from 195 to 300 nm using a 20 points $min^{-1}$ scan speed. Three accumulations were acquired for each sample and averaged. The resulting data were baseline corrected for buffer contributions, converted to molar ellipticity units (θ, deg $cm^2$ $dmol^{-1}$) calculated by means of Equation 2.1 (where: $\theta_{obs}$ is the observed ellipticity in m°; C is the concentration in g $L^{-1}$; M is an average molecular weight in g $mol^{-1}$; L is path of cell in cm (0.1)) and processed using OriginPro 8 software (Northampton, Massachusetts, USA).

$$\theta = \frac{\theta obs * M}{C * L * 10} \quad \textbf{Equation 2.1}$$

Assays for the CD spectra of cyclic peptides in binding conditions, were carried out using 50 µM peptides in 3.3 mM HEPES buffer, pH 8.0 (mimicking binding conditions). Additionally, CD spectra of cyclic peptides in elution conditions were carried out with 100 µM peptides in 10 mM HEPES, 12.5 mM NaCl pH 8.0 buffer (mimicking elution conditions). All solutions were purged with nitrogen.

### 2.3.7 MicroScale Thermophoresis (MST)

MST method allows to analyze the interactions between different biomolecules in solution. It detects the thermophoretic movement of fluorescent molecules, which is induced by a temperature gradient created inside the equipment. During MST assay just one of the binding partners is fluorescently labeled or has intrinsic fluorescence that can give a significant signal, the concentration of this molecule is held constant in all reaction mixtures of each assay.

#### 2.3.7.1 Cyclic peptides labeling

In this study the fluorophore Cyanine5 maleimide (Cy5) was coupled to the cyclic β-hairpin peptides by reaction with thiol group of cysteine residues. Labeling was done according to the standard protocol recommended by the supplier: (i) peptides were dissolved in 900 µL of, previously degassed, labeling buffer (50 mM Tris, 150 mM NaCl and 5 mM EDTA, pH 7.2) with 10 µL of TCEP; (ii) 1 mg of fluorophore dye (which corresponded to 6x fold excess of dye) was dissolved in 100 µL of DMF; (iii) the mixture of peptide and Cy5 was incubated at 4°C, overnight in darkness. Unreacted Cy5 and the Cy5-labeled peptides were separated by dialysis using CE membranes from Spectra/Por® Float-A-Lyze® (MWCO 0.5 - 1.0 kD) against binding buffer (50 mM HEPES, pH 8). Dialysis was performed for 48 h and the concentration of Cy5-labeled peptide was determined by measuring the absorbance at 646 nm (extinction coefficient for Cy5: 25000 $L \cdot mol^{-1} \cdot cm^{-1}$).

#### 2.3.7.2 MST assay

Labeled peptides were diluted in 50 mM HEPES, pH 8.0 with 0.1% Tween 20 to final concentration of 20 nM. All stock solutions of unlabeled target peptides (Table 2.2) were dissolved in assay buffer (50 mM HEPES, pH 8.0) to concentrations 5 mM (GK14 and GK14P) and 0.250 mM (for SW6P, SW6, YW6P, YW6, YW13and YW13P). For the MST assay 16 step (1:1) serial dilutions of target peptides were prepared with final volume 10 µL. Next, 10µL of a stock solution of the labeled peptide were added into each target dilution and mixed carefully (Figure 2.4 - A). With this, the highest concentration for target were 2.5 mM (GK14 and GK14P) and 0.125 mM (for SW6P, SW6, YW6P, YW6, YW13and YW13P), for labeled peptide being 10 nM and 0.05% for Tween 20. All the solutions were filled in the standard glass capillaries (Monolith NT.115 Standard Treated Capillary K002) after the sample mixing and the signal was measured by Monolith NT.115 (Nanotemper). All measurements were performed at 22°C with the red LED color, which correspond to optimal excitation of Cy5 fluorophore (excitation maximum at 646nm). MST-power and excitation power were held at 40% and 20% respectively. The data was recorded using NT Control 1.6 (NanoTemper® Technologies GmbH), analyses and curve fitting were carried out using NT Analysis 2.3 (NanoTemper® Technologies GmbH).

For the competition assays (Figure 2.4 - B) the mixture of 20 nM of Cy5-labeled peptide, 20 mM of sodium phosphate solution and 0.1% Tween 20 was incubated for 30 min in darkness. Next, 10 µL of this mixture was added to the 16-step serial dilution of GK14P (being the maximum concentration tested 2.5 mM) to get the final concentrations: 10 nM of Cy5-labeled peptides, 10 mM sodium phosphate pH 8.0 and 0.05% Tween 20.

**Table 2.2**: Sequences of the target phosphorylated peptides and their non-phosphorylated counterpart and, the highest concentrations tested during preparation of serial dilutions.

| Target ID | Sequence | [max] tested (mM) |
|---|---|---|
| GK14 | G-A-A-Y-D-I-S-Q-V-F-P-F-A-K | 2.5 |
| GK14P | G-A-A-Y-D-I-**p**S-Q-V-F-P-F-A-K | 2.5 |
| SW6P | **p**S-Q-V-F-P-W | 0.125 |
| SW6 | S-Q-V-F-P-W | 0.125 |
| YW6P | **p**Y-G-G-I-P-W | 0.125 |
| YW6 | Y-G-G-I-P-W | 0.125 |
| YW13P | Y-A-G-**p**S-**p**T-D-E-N-**p**T-D-S-E-W | 0.125 |
| YW13 | Y-A-G-S-T-D-E-N-T-D-S-E-W | 0.125 |

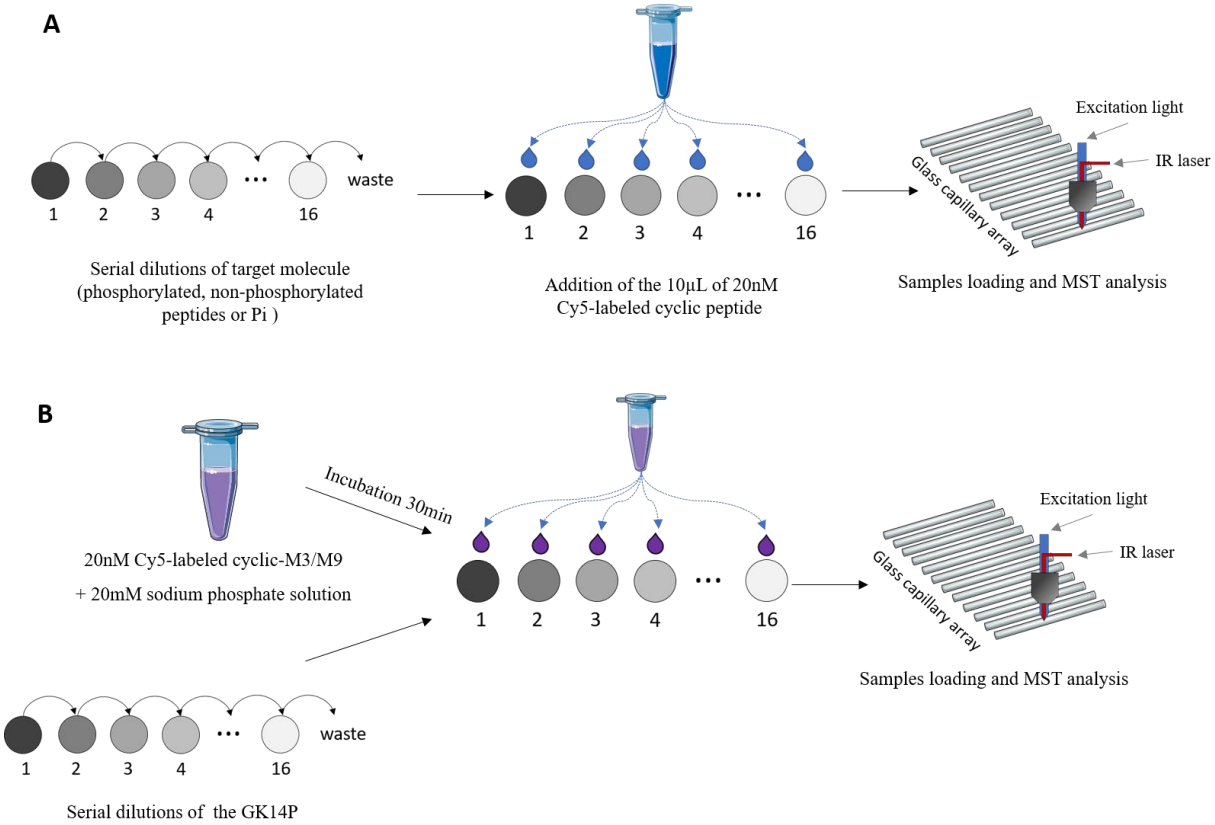


**Figure 2.4:** Schematic overview of the samples preparation for standard MST experiment (A) and competition experiment (B).

## 2.4 Results and discussion

### 2.4.1 Synthesis and purification of cyclic peptides

The three cyclic 14-mer peptides, M0, M3 and M9 (Table 2.1), were synthetized by Fmoc-based Solid Phase Peptide Synthesis (SPPS) methods, purified by RP-HPLC and characterized by ESI-MS. Cyclic-M0 and cyclic-M9 were synthetized in an automated microwave Biotage equipment which uses microwave technology to improve peptide synthesis due to the kinetic power input [101]. Cyclic-M3 was manually produced using similar SPPS methods in a glass reactor and at RT. Chlorotrityl resin was used in all three cases, and peptides were cleaved from resin under mild acidic conditions to obtain fully protected peptides with free N- and C-terminals.

The HPLC chromatogram of cyclic-M0 crude peptide (Figure 2.5-A) shows a peak with retention time (Rt) = 16.6 min (eluted at ~22% solvent B), which was collected and lyophilized. Figure 2.5-B shows the corresponding analytical HPLC, with final purity of 92% and ESI-MS analysis (Figure 2.5-C) confirmed the presence of the target, cyclic-M0, peptide (1010.4 Da).

The ESI-MS analysis of the crude cyclic-M3 peptide (Figure 2.6-A), confirmed the presence of two peptides: the target sequence (1311.7 Da) and another peptide with lower molecular weight (1198.6 Da). The last one, corresponds to the target sequence minus one isoleucine residue (Ile5 or Ile7). During SPPS, the synthesis is performed from C-terminal to N-terminal which implies that, there was a failure during the coupling Ile to either Gly6 or Pro8. Considering that Pro has a secondary amine, most likely the source of error was the coupling of Ile7 to Pro8. Additionally, the steric effect between a large side chain group of Ile and side chain group of Pro may decrease coupling efficiency. Therefore, double-coupling for the incoming residues after proline could be performed to increase synthesis yield.

During the purification of cyclic-M3, (figure 2.6-B) two different fractions were collected: first one corresponded to the peak with Rt = 14.0 min (eluted at ~33%solvent B), and the second corresponded to the peak at Rt = 23.1 min (eluted at ~37.5% solvent B). The ESI-MS analysis of purified peptides revealed that the second fraction (with Rt = 23.1 min) corresponded to the target sequence (Figure 2.6-D). The purity of this fraction was analyzed by analytical HPLC (Figure 2.6-C) as 95%.

For cyclic-M9, the HPLC chromatogram of crude peptide (Figure 2.7-A) identifies a peak with Rt = 16.5 min (eluted at ~21.5%) that was collected and lyophilized. It was confirmed as a target sequence (Figure 2.7-C) with expected molecular weight of 1257.6 Da and the final purity of 97% (Figure 2.7-B).

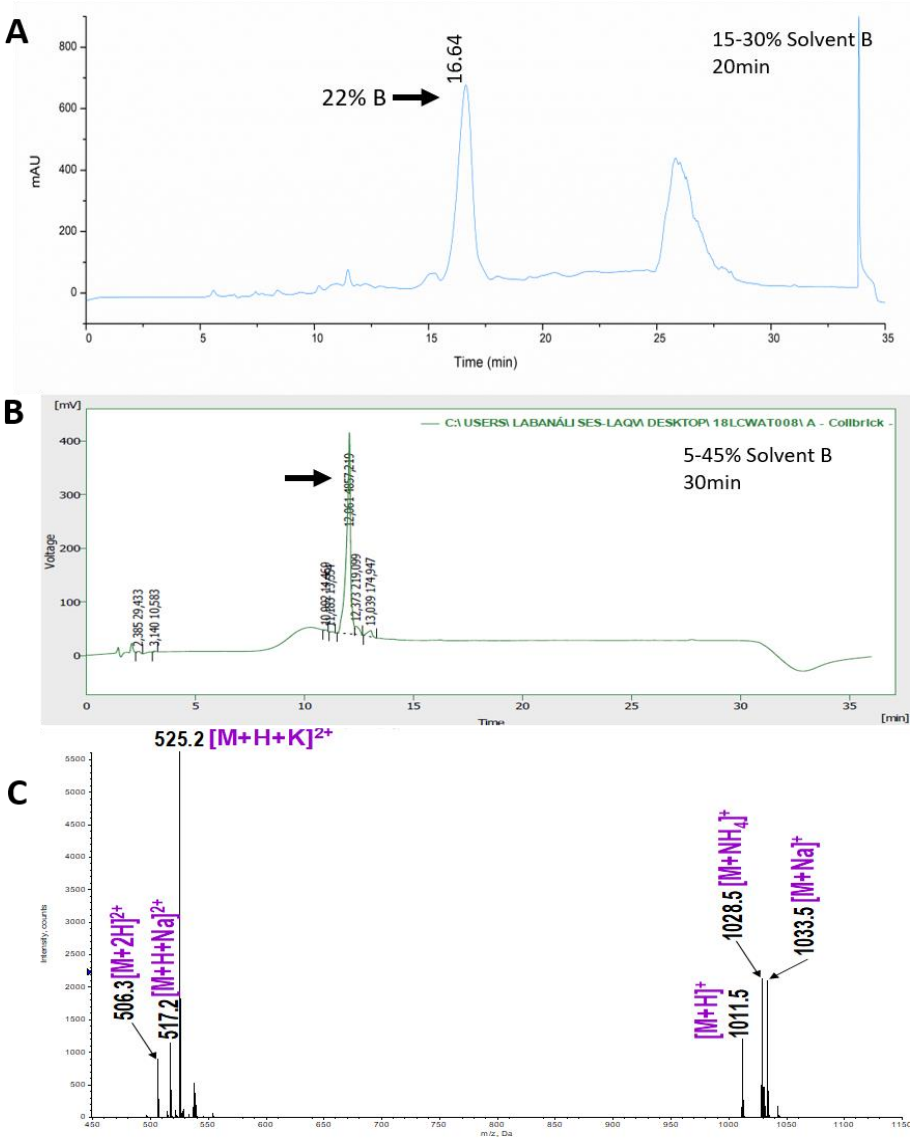


**Figure 2.5:** Purification and identification of cyclic-M0 (A): Preparative HPLC chromatogram of crude peptide. Collected peak and eluting conditions are identified by arrows, details of gradient methods used in top right of chromatograms. (B): Corresponding analytical HPLC chromatogram of collected peak. (C): ESI Mass spectrum of collected peak m/z: $[M+H]^+$ 1011.4(calculated)/1011.5(measured); $[M+2H]^{2+}$ 506.2/506.3; $[M+H+Na]^{2+}$ 517.2/517.2; $[M+H+K]^{2+}$ 525.2/525.2; $[M+NH4]^+$ 1028.4/1028.5 and $[M+Na]^+$ 1033.4/1033.5. Absorbance signals monitored at 220 nm. M – cyclic-M0

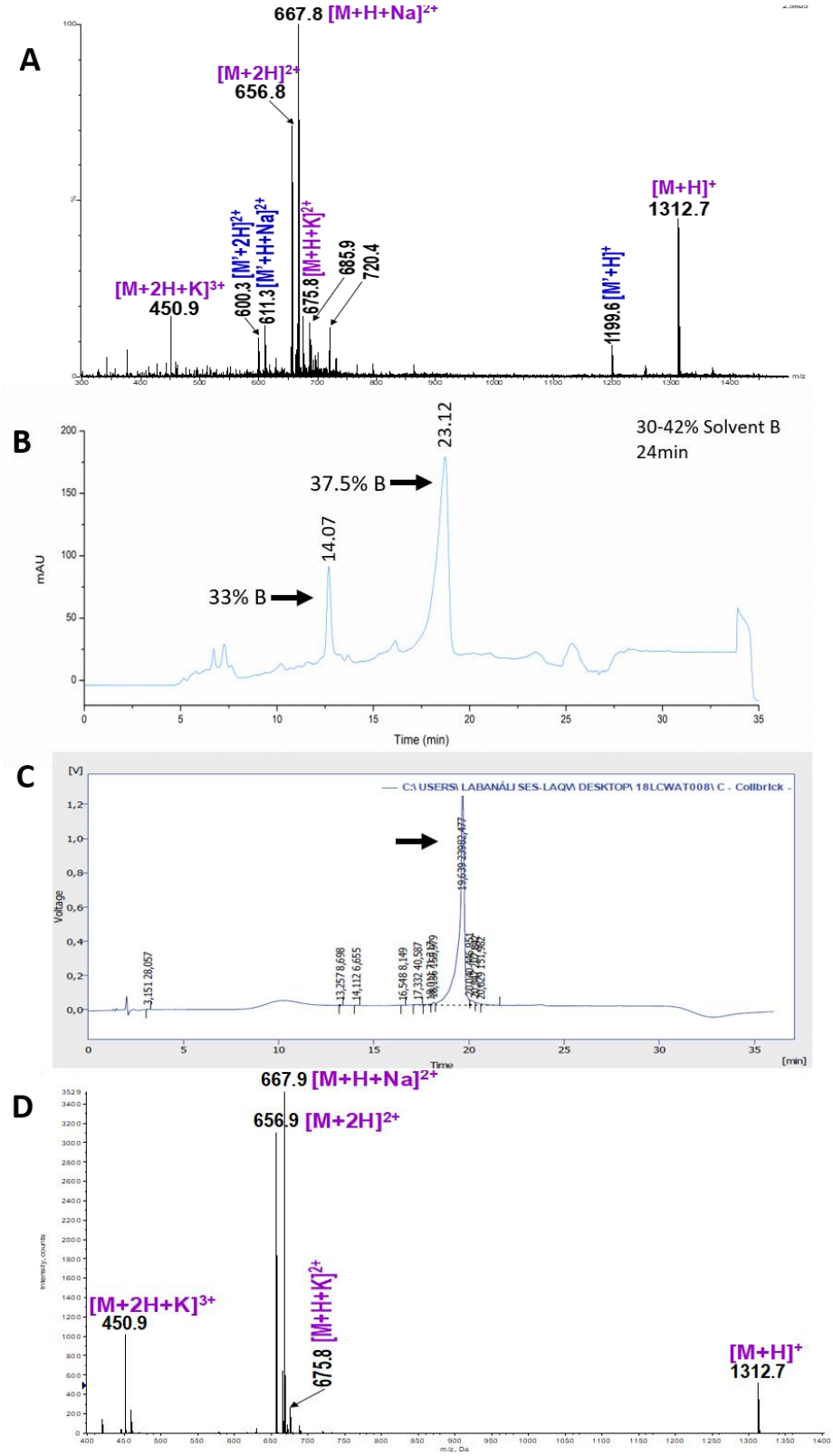


**Figure 2.6:** Purification and identification of cyclic-M3. (A): ESI Mass spectrum of crude peptide m/z: [M+H]+ 1312.7(calculated)/1312.7(measured); [M+2H]2+ 656.9/656.8; [M+H+Na]2+ 667.8/667.9; [M+H+K]2+ 675.8/675.8; [M+2H+K]3+ 450.9/450.9; and [M'+H]+ 1199.6/1199.6; [M'+2H]2+ 600.3/600.3; [M'+H+Na]2+ 611.3/611.3. (B): Preparative HPLC chromatogram of crude peptide. Collected peak and eluting conditions are identified by arrows, details of gradient methods used in top right of chromatograms. (C): Analytical HPLC chromatogram of peak collected at Rt =23.12min. (D): ESI Mass spectrum of collected peak (Rt = 23.12min) m/z: [M+H]+ 1312.7(calculated)/1312.7(measured); [M+H+Na]2+ 667.8/667.9; [M+2H]2+ 656.9/656.9; [M+2H+K]3+ 450.5/ 450.9. Absorbance signals monitored at 220 nm. M- cyclic-M3. M´- 13-mer peptide.

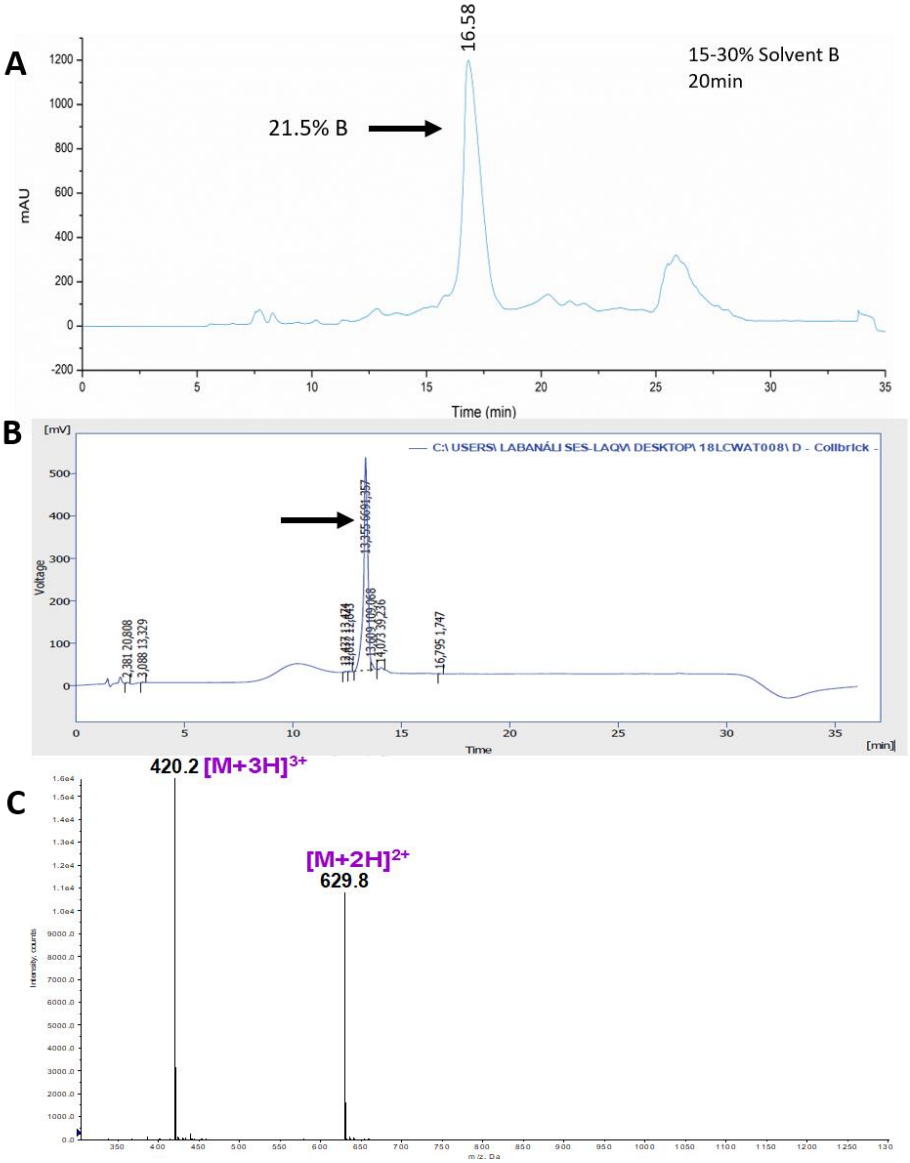


**Figure 2.7**: Purification and identification of cyclic-M9. (A): Preparative HPLC chromatogram of crude peptide. Collected peak and eluting conditions are identified by arrows, details of gradient methods used in top right of chromatograms. (B): Corresponding analytical HPLC chromatogram of collected peak. (C): ESI Mass spectrum of collected peak m/z: [M+3H]3+ 420.2(calculated)/420.2(measured) and [M+2H]2+ 629.8/629.8. M- cyclic-M9.

### 2.4.2 Characterization of peptides folding

CD spectroscopy was carried out to determine secondary structure of the cyclic peptides and their stability in previously stablished binding and elution conditions. The β-hairpins have a characteristic CD signal similar to β-sheet structure showing a positive ellipticity around 200 nm and a maximum negative ellipticity close to 216 nm [102,103].

The far-UV CD spectra of cyclic peptides were measured using optimized buffer conditions: 3.3 mM HEPES buffer pH 8.0 (mimicking binding condition - Figure 2.8-A) and 12.5 mM NaCl

10 mM HEPES buffer pH 8.0 (mimicking elution condition- Figure 2.8-B). All three cyclic peptides have a similar folding. For cyclic-M3 and cyclic-M9 the spectra present typical maximum negative bands (around 214-216 nm) and positive bands (around 195-200 nm) in both tested conditions. In the case of cyclic-M0, the positive ellipticity was observed around 196-200 nm and the maximum negative ellipticity at 212 nm in both conditions. These results are consistent with the typically observed bands for β-hairpin motifs and are similar to the results reported for peptides containing the $^{D}$Pro-$^{L}$Pro template which induce the β-turn formation [104–106].

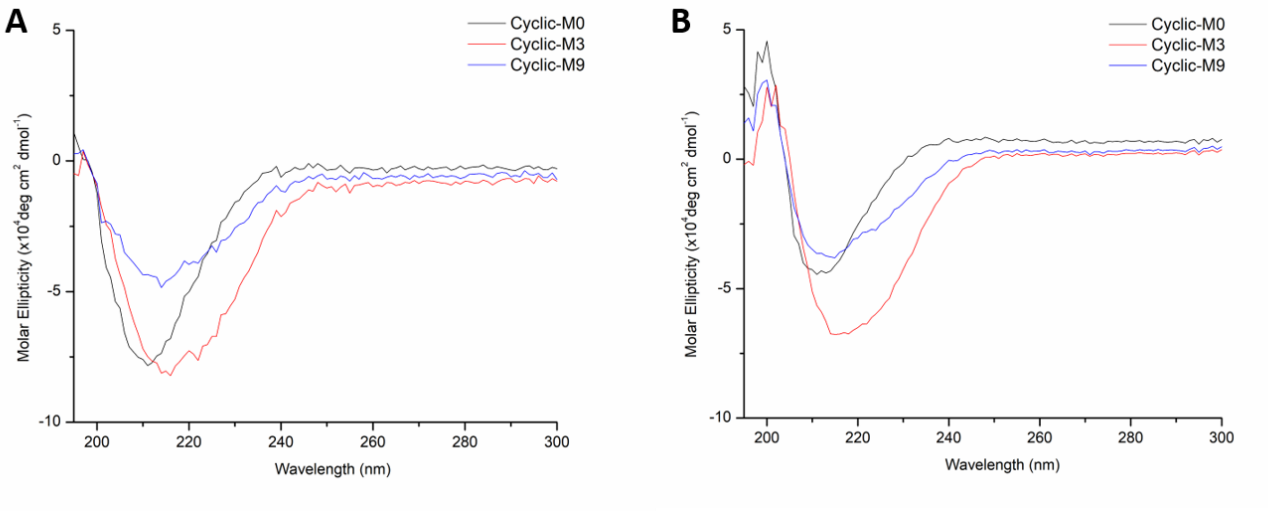


**Figure 2.8:** Far-UV CD spectra of the cyclic β-hairpin peptides. (A): 50 µM of peptides in 3,3 mM HEPES buffer pH 8.0; (B): 100 µM of peptides in 12.5 mM NaCl 10 mM HEPES buffer pH 8.0

Considering the intensity of molar ellipticity, cyclic-M3 and cyclic-M9 demonstrated to be more stable in tested conditions (binding and elution) in comparison to cyclic-M0, whose signal intensity of molar ellipticity decreased with increasing of ionic strength of the solution. Additionally, cyclic-M3 has higher signal, in comparison to the signal intensity observed for cyclic-M9. This can indicate that cyclic-M3 has a more constrained system then cyclic-M9.

### 2.4.3 Detection of binding and determination of affinity parameters

The interaction between designed peptides (cyclic-M3 and cyclic-M9) and phosphorylated target peptides was analyzed using MicroScale Thermophoresis (MST). The non-phosphorylated counterparts of targets were also tested. Cyclic-M0 was used as a negative control ligand.

The MST is used to determine dissociation constant ($K_D$) and to study interaction between molecules. It relies on the difference in thermophoretic movement of the unbound fluorescent molecule relative to the complex. This movement is induced by the temperature gradient created during the measurements inside the equipment. MST technique requires small sample quantities and, due to high sensitivity, can detect small changes in size, charge, hydration shell or

conformation of molecules that influence the thermophoretic movement. It is a relatively high throughput method (each MST assay take in average about 15min) therefore can test many different conditions in a small period.

All three peptides were labelled on Cys1 residue, which was introduced into the sequence as an anchoring point for immobilization. Cyanine5 Maleimide (Cy5) dye was used for labeling the cyclic peptides as it contains maleimide group which is known to have high selectivity towards free thiols. During the measurements, the concentration of labeled peptides was maintained constant whereas the targets were prepared in serial dilutions (Figure 2.4). All MST measurements were performed in 50 mM HEPES buffer pH 8.0 and 0.05% Tween 20. Tween 20 (Polysorbate 20) was added to the reaction mixtures, at a low concentration, to prevent peptides aggregation.

As synthetic peptides cyclic-M3 and cyclic-M9 were designed to mimic the binding site of BRCT domain from BRCA1 protein, GK14P was chosen as target. GK14P is an optimized peptide for binding the BRCT domain of BRCA1 previously identified from an oriented peptide library with high affinity ($K_D$ 400nM) [107] comprising the consensus sequence, pSer-X-X-Phe.

The phosphopeptides SW6P, YW6P, YW13P were selected for binding tests with cyclic peptides to assess specificity. When comparing with GK14P these peptides present different sequences, phosphorylated pattern and length. SW6P peptide has six amino acid residues and a consensus sequence (pS-X-X-F); YW6P has a pTyr residue; and YW13P has pSer and two pTyr residues. The sequences of all peptides are shown in Table 2.2.

The results of interaction of cyclic-M3 with GK14P and the control GK14 are shown in Figure 2.9-A. The upper plateau is clearly visible for GK14P, whereas the lower plateau was not distinguishable with the concentrations used. In order, to achieve lower plateau, higher concentrations of target peptides were tested, however it was not possible to solubilize the peptides completely even with the addition of 5% DMF (higher concentration allowed during MST assays). Without a visibly defined lower plateau, the determination of the $K_D$ is not as accurate, but still possible yielding 1.0±0.46 mM for GK14P. In contrast, no binding was observed between cyclic-M3 and the non-phosphorylated peptide, GK14. This indicates that the presence of the phosphate group is critical for the molecular recognition to occur. For cyclic-M9 (Figure 2.9-B), the absence of the lower plateau was also observed for MST assays. Unlike cyclic-M3, it was able to bind both GK14P and GK14 and the obtained $K_D$ values were 0.73±0.32 mM and 1.93±0.11 mM, respectively. Figure 2.11 show an overview of the affinity constants ($K_A$) obtained for interactions between cyclic peptides and GK14P, and GK14. Both, cyclic-M3 and cyclic-M9 were able to bind GK14P with similar affinity, however cyclic-M3 shows higher selectivity towards phosphorylated target. This specificity of the cyclic-M3 probably is a result

of the higher constrained structure (basing on CD results), that produce the rigid conformation required for binding of the phosphate moiety.

Cy5 dye used for peptides labeling has a positive charge, which probably may interact with a negative charged phosphate moiety of target ligands. However, cyclic-M0 (Figure 2.9-C) does not show any binding for GK14P neither for GK14. This indicates that the affinity observed between GK14P and cyclic-M3, and cyclic-M9 is due to interactions between amino acid residues and that Cy5 does not contribute significantly for molecules binding.

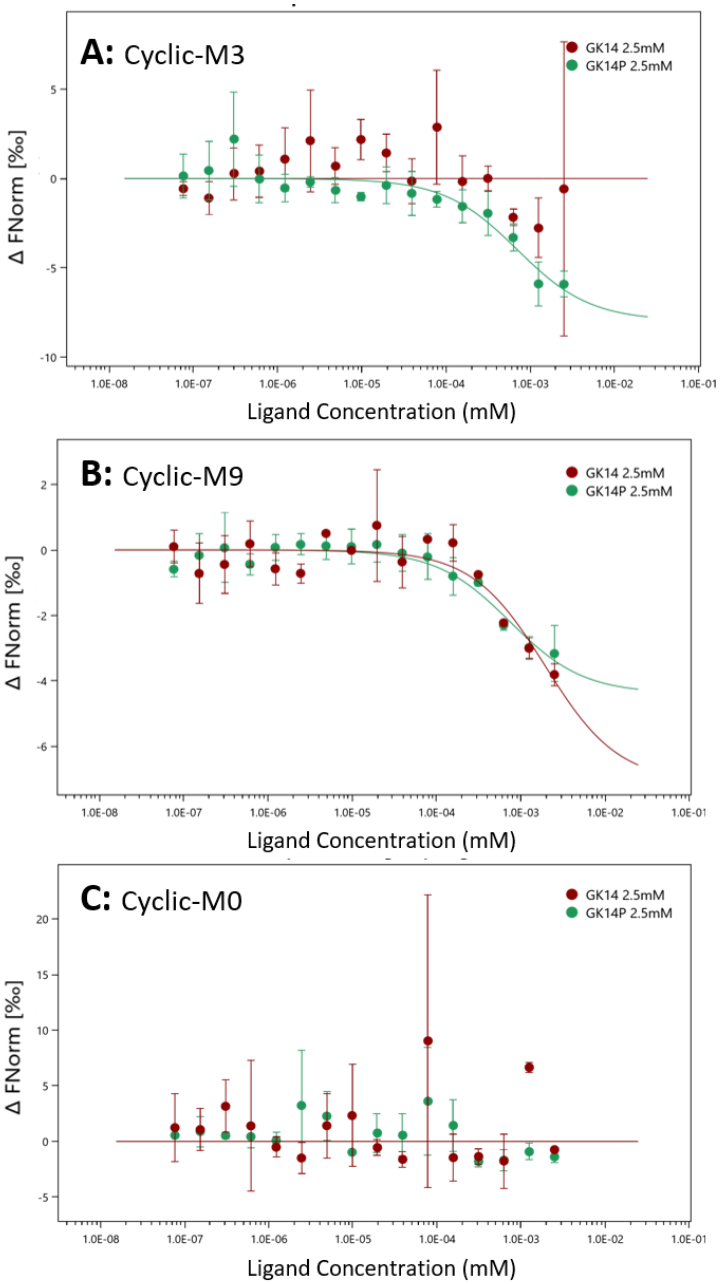


**Figure 2.9:** Dose-response curve fitted from MST assay. 10 nM labelled peptide was used in each assay, with a variation of GK14P and GK14 (maximum fixed at 2.5 mM). (A): cyclic-M3; (B): cyclic-M9; (C): cyclic-M0.

Inorganic phosphate (Pi) is frequently present in buffers and biological samples. To analyze its effect on the binding interaction of designed peptides with GK14P, competitive MST assays were performed with addition of Pi to reaction mixture.

Initially the interaction between cyclic peptides and Pi (presented as sodium phosphate solution) was analyzed. Figure 2.10-A shows the results for these assays, where cyclic-M3 and cyclic-M9 were added to the serial dilution of the sodium phosphate solution (maximum fixed at 10mM). No binding was observed for cyclic-M3 while cyclic-M9 was able to form a complex with Pi (the $K_D$ determined was 2.67±0.89 mM).

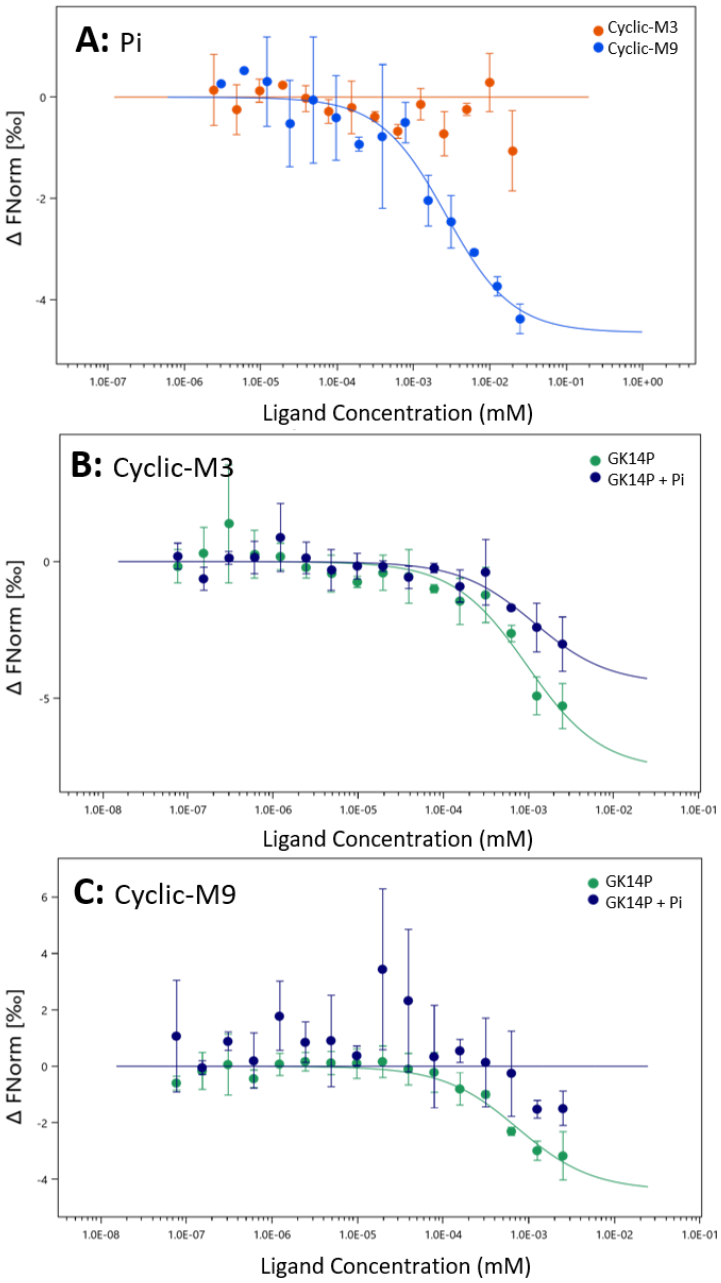


**Figure 2.10**: Evaluation of the Pi influence during molecular recognition. (A): Interactions of cyclic-M3 and cyclic-M9 with Pi. No binding was observed for cyclic-M3 in contrast to cyclic-M9 ($K_D$ 2.67±0.89 mM). Molecular recognition of GK-14P by cyclic-M3 (B) and cyclic-M9 (C) in the absence (green-GK14P) and in the presence of Pi (blue- GK14P+Pi) in reaction mixture;

Then, competition experiments were performed. For this, 20 nM of labeled peptides were incubated with 20 mM sodium phosphate solution and then added to the serial dilutions of GK14P (maximum fixed at 2.5mM) (sample preparation is shown in Figure 2.4-B). Considering, cyclic-M3 (Figure 2.10-B), the $K_D$ value obtained, for its interaction with GK14P, was 1.2±0.68 mM. On the other hand, cyclic-M9 (Figure 2.10-C) was not able to bind GK14P in the presence of Pi.

Comparing both tested conditions (with and without Pi in solution) the $K_A$ value of cyclic-M3 towards GK14P was not significantly affected (Figure 2.11) in comparison to cyclic-M9. This indicates that for cyclic-M9 both molecules (Pi and GK14P) may compete for the same binding site.

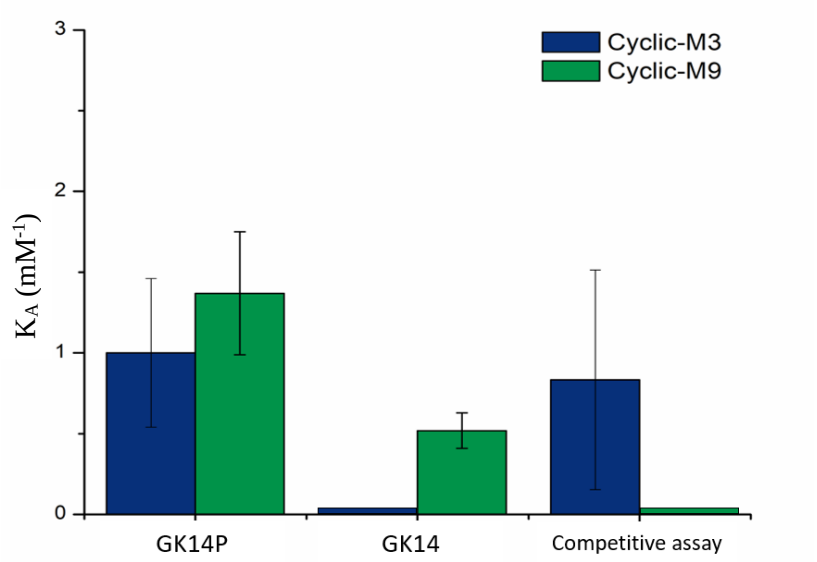


**Figure 2.11:** Overview of the $K_A$ ($mM^{-1}$) calculated from the deduced $K_D$ values from the binding assays with GK14P, GK14 and competitive assay (with Pi in reaction mixture). Cyclic-M3 was selective for GK14P ($K_A$ = 1.0 $mM^{-1}$) and, unlike cyclic-M9, was able to bind it in the presence of Pi in reaction mixture ($K_A$ = 0.83 $mM^{-1}$). Cyclic-M9 was not selective for phosphorylated target (GK14P), however the affinity for non-phosphorylated counterpart (GK14) was ±2.5 fold lower ($K_A$ = 1.37 $mM^{-1}$ for GK14P and $K_A$ = 0.51 $mM^{-1}$ for GK14).

Considering phosphorylated peptides SW6P, YW6P YW13P and their non-phosphorylated counterpart (SW6, YW6 and YW13), binding studies were performed with all three cyclic peptides. Samples were prepared as previously described for GK14P and GK14: serial dilutions were made with the highest concentration of 0.125 mM and 10 nM of cyclic labeled peptide in each sample Since the $K_A$ values for these interactions (cyclic peptides vs target peptides) were unknown the concentration of target peptides (0.125 mM) was chosen in order to ensure that the minimal concentration of titrated peptide is sufficiently low to measure the thermophoretic movement of the unbound cyclic peptide. Results are shown in Figure 2.12 and it was observed that none of the three cyclic peptides showed binding to SW6P or YW6P or YW13P (Figure 2.12-A, B and C respectively) or their non-phosphorylated counterpart (data not shown) in the tested conditions. As it was mentioned in Chapter 1, the mimicking of the proteins binding site in a smaller ligand structure may lead to decreasing in affinity to the target molecules, as not all

interactions that are stabilized naturally can be transplanted into a scaffold. During the design of cyclic β-hairpins peptides just amino acids residues that are known to interact directly with pSer (at position 0) and Phe (at position +3 from pSer) residues were considered. However, some recent studies have demonstrated that interactions which BRCA1-BRCT domain establish with adjacent peptide residues (pSer and Phe) are also important for binding affinity [108,109]. This can explain the observed results like no binding to SW6P, and the lower determined affinity of cyclic peptide ligands for GK14P in the tested conditions.

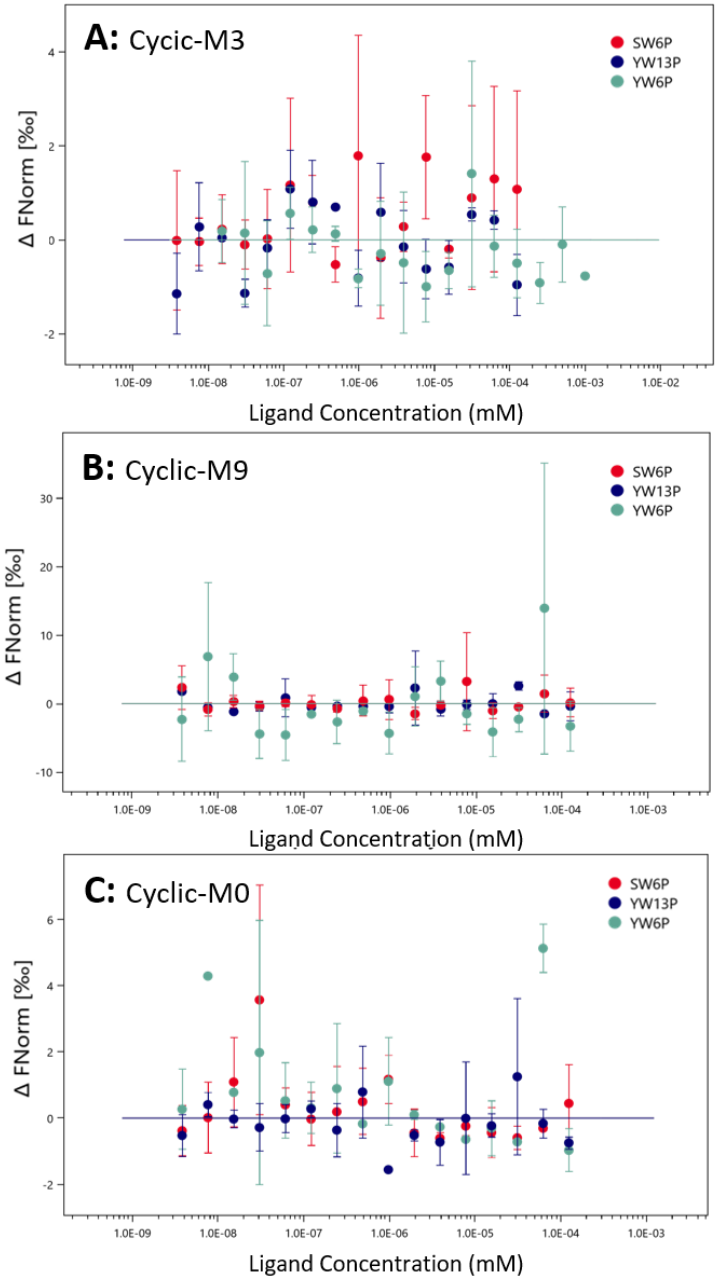


**Figure 2.12**: Dose-response curve fitted from MST assays. 10 nM of labelled peptide was used in each assay, with a variation of SW6P, YW13P and YW6P concentrations (maximum fixed at 0.125 mM). (A): cyclic-M3; (B): cyclic-M9; (C): cyclic-M0.

## 2.5 Concluding remarks

This study focused on the chemical synthesis and characterization of cyclic β-hairpin peptides (two previously designed peptides: cyclic-M3, cyclic-M9 and one control cyclic-M0) and evaluation of their potential as affinity ligands towards specific phosphorylated peptides. All cyclic β-hairpins were successfully produced by solid phase peptide synthesis. The production of peptides using microwave heating showed to be more efficient, as synthesis were faster, with higher yields and higher crudes purity.

During the design of these cyclic peptides, a stable β-hairpin template was used where novel chemical functions were introduced. As expected, the purified cyclic peptides presented the β-hairpin folding as measured by CD.

The affinity of ligands towards phosphorylated target peptides was measured by MST. This technique provides a fast analysis of interactions with a low sample consumption. It allows the use of the binding conditions (50 mM HEPES pH 8.0) and to determinate the $K_D$ values.

The analysis of the binding against GK14P e GK14 shows that both cyclic-M3 and cyclic-M9 have ability to interact with the GK14P with similar affinity. Nevertheless, in the tested conditions only cyclic-M3 shows higher capacity to discriminate between phosphorylated (GK14P) and non-phosphorylated (GK14) peptides. Such higher selectivity may be related to the more constrained structure of cyclic-M3 (according to CD results), which allows to display the amino acid residues essential for recognition of the phosphate moiety in a more appropriate way.

The results obtained indicate that cyclic β-hairpin is a promising scaffold for generation of affinity ligands. Still, additional studies need to be performed. For example, to test different experimental conditions to determine more accurate binding constants.

# CHAPTER 3

# Biological production of Odorant-Binding Proteins for application in biosensing

## 3.1 Odorant-binding proteins as affinity reagents

Volatile organic compounds (VOCs) are small organic molecules with a high vapor pressure at room temperature conditions and low boiling point. The identification and quantification of VOCs attract attention due to the extensive abundance of these compounds in nature. VOCs are produced by human body and different microorganisms, thus VOCs can act as markers in certain diseases and identified from biological samples (blood, breath, faeces) [110–113]. The control of VOCs is also important in food quality, to monitor food decay processes in which different VOCs are produced and released [114], and to analyze spoilage bacteria [115,116]. Due to adverse environmental effects of many natural and anthropogenic VOCs, there is also a need to control their emission through identification and quantification [117].

Electronic nose (e-nose) devices have been successfully used for the detection of a wide range of VOCs for different purposes [118]. E-noses are artificial olfactory systems that use arrays of sensors with the ability to identify different odors which are a set of VOCs. In some cases, sensors use affinity ligands, typically peptides or proteins to recognize a specific volatile molecule.

Odorant-binding proteins (OBPs) have received attention as potential affinity reagents for VOCs, as they are naturally optimized for a specific olfactory function. These proteins are known to transport small hydrophobic molecules, like odorant and pheromones from air towards olfactory receptors. OBPs are relatively small (15-20kDa) soluble proteins, found in the mucus of insects and vertebrates. Vertebrate OBPs are a sub-class of the lipocalin superfamily [78] and have the conserved eight-strand β-barrel scaffold with small α-helix on C-terminal. As all lipocalins, vertebrate OBPs are known to have low sequence identity among different OBPs (generally 21-26%) [119], however few completely conserved residues allow their identification (Figure 3.1).


**Figure 3.1**: Amino acid sequence alignment of vertebrates OBPs from different species: human OBP (UniProt ID: Q9NY56), rat OBP1 (UniProt ID: P08937), rat OBP2 (UniProt ID: Q63613), rat OBP3 (UniProt ID: Q78E14), giant panda OBP3 (PDB: 5NGH [81]), porcine OBP (UniProt ID: P81245), bovine OBP (UniProt ID: P07435). Conserved residues are highlighted.

OBPs have several specific features making them good candidates as molecular recognition ligands for application in protein-based biosensors. These proteins can be expressed with high yield in bacterial systems and are relatively easy to purify when produced fused to a His-tag [120], making their production cost effective. Another important property, is their ability to bind reversibly a broad range of VOCs with a micromolar affinity [121]. OBPs tolerate well point mutations, and as a result their binding properties can be tailored through mutagenesis [122,123]. The proteins of this family have also demonstrated a high chemical and thermal stability, and to reversibly fold when the initial conditions are reestablished [124].

Several studies describe the potential of vertebrate OBPs as sensing elements. Ramoni et al. [122], used different forms of pOBPs and bOBPs (mutated and wild-type) as models to create affinity reagents for the detection of hazard explosive molecules. Di Pietrantonio and co-workers [125] have developed an array of biosensors using pig OBP (pOBP), bovine OBP (bOBP) and a double mutant of bOBP. These biosensors were able to distinguish between two tested compounds (octenol and carvone) and have a high sensitivity in the range of ppm (13 ppm for octenol and 9 ppm for carvone). By changing some specific amino acid residues in the pOBP biding site, Yin Wei et al. [126] improved the specificity of this protein towards polyaromatic hydrocarbons for monitoring environmental pollution. This demonstrates the potential of site-directed mutagenesis to improve binding properties of OBPs.

In this work, after performing an extensive study of structural and biochemical data about mammalian OBPs and their interaction with VOCs, the rat OBP2 and rat OBP3 were selected as

models for application in biosensing through rational design. The selection of these proteins was based on their higher specificity as well as the possibility of recombinant expression and relatively easy purification, previously reported [127–129130].

## 3.2 Materials

### 3.2.1 Reagents

The reagents utilized were of the highest grade available. Ethanol, glycerol and methanol were obtained from PanReac. Bacterial growths were performed using Luria Broth (LB), ampicillin (Amp) and agar powder from NzyTech. For agarose gels, tris base, agarose (ultrapure grade), loading dye, GreenSafe Premium staining, NzyDNA Ladder II and NZYDNA Ladder V DNA markers were purchased from NzyTech. Ethylenediaminetetraacetic acid (EDTA), sodium phosphate dibasic heptahydrate and sodium phosphate monobasic monohydrate were obtained from Sigma-Aldrich (now Merck). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from NzyTech. FastDigest buffer (10x) was purchased from Thermo Scientific. For pDNA gel extraction and purification the NZYMiniprep kit from NzyTech was used. The sodium dodecyl sulfate (SDS) 10% solution and 30% Acrylamide/Bis Solution (ratio 37.5:1) were purchased from Bio-Rad. Ammonium persulphate (PSA), N-N-N'-N'-tetramethylethylenediamine (TEMED), Blue bromophenol sodium salt were obtained from Carl Roth. Coomassie Blue R-250 dye was obtained from Sigma-Aldrich. The protein marker used in SDS-PAGE gel, was the Low Molecular Weight protein marker (LMW) purchased from Nzytech. Restriction enzymes (NheI and HindIII) and Fast-AP (Thermosensitive Alkaline Phosphatase) were purchased from Thermo Scientific. T4 DNA ligase (5U $\mu L^{-1}$) and DNasel were obtained from NzyTech and Roche respectively.

The plasmid pAP006 was previous constructed in the laboratory (UCIBIO, Portugal; Figure S1) and it contains: a bacteriophage T7 promoter, an Amp resistance, a pBR322 origin, and a lacI coding sequence. Plasmids pEX-A128-rOBP2 (pEX-OBP2) and pEX-A128-rOBP3 (pEX-OBP3) were ordered from Eurofins (Genomics) and they contain designed OBP2 and OBP3 gene fragments respectively, an Amp resistance and a pUC origin. The competent cells *Escherichia coli* BL21(DE3), Nzy5α and NzyStar were purchased from NzyTech and their genotype is described in Table 3.1

**Table 3.1**: Genotype of competent cells used for cloning and expression of OBPs

| Name | Genotype | Supplier |
|---|---|---|
| Nzy5α | *fhu*A2Δ(*argF-lacZ) U169 pho*A *gln V44 Φ80* Δ(*lac*Z)M15 *gyr*A96 *rec*A1 *rel*A1 *end*A1 *thi*-1 *hsd*R17 | NzyTech |
| NzyStar | *endA1 hsdR17($r_k$-,$m_k$+) supE44thi1 recA1 gyrA96 relA1 lac[F´proA$^+$B$^+$ lacI$^q$ ZDM15 :Tn10(Tc$^R$)]* | |
| BL21(DE3) | *F$^-$ ompT gal dcm lon hsdSB(rB- mB-) λ(DE3 [*lacI lacUV5-T7 gene 1 ind1 sam7 nin5]* | |

### 3.2.2 Equipments

The temperature was controlled during microbial growth and proteins were produced using the Incubator KS4000ic from IKA. The T100 Thermal Cycler from Bio-Rad was used for the reactions involving DNA. E-gel Precast Agarose Electrophoresis System from Thermo Scientific for DNA extraction and purification. For agarose and SDS-page gels the systems PowerPac Basic Power supply with MiniSub cell GT and Mini-Protean Tetra System from Bio-Rad were used respectively. For gel image and analysis UV Transiluminator (Bio-Rad Gel Doc XR+) was used. The DNA quantification and purity evaluation were performed with a Nanodrop ND-1000 spectrophotometer (Thermo Scientific). Cells disruption was performed with French Press (Thermo Scientific). Following centrifuges were used: ScanSpeed mini (Rotor: GAM-1.5-12) from Labogene; refrigerated centrifuge HERAEUS Multifuge X3R (rotor: Fibralite F14-6x250LE) from ThermoScientifc and Ultracentrifuge Optima LE-80K from Beckman Coulter.

## 3.3 Methods

### 3.3.1 Literature review and data collection

To evaluate the availability of structural information on OBPs, a PDB search [131] was performed using key words: "*odorant-binding protein", "OBP", "mammalian", "lipocalin"*. The crystal structures of six OBPs from different species were identified: one bovine OBP (bOBP), one porcine (pOBP), one human (hOBP-2A), one pandas' (AimelOBP3), and two rats' (OBP1 and OBP3). Some of the proteins were co-crystalized with volatile molecules inside their binding site. In total 22 PDB files were obtained from this search. (Table S1).

To analyze the binding specificity of mammalian OBPs, scientific data, available until May 2018 were obtained from searches performed in the online databases Pubmed, Web of Science and Google Scholar. Review articles were not considered, only scientific articles containing experimental data were considered. Articles selected for the analysis contained: (i) description of proteins used (wild type or mutated and mutation performed); (ii) binding studies performed through fluorescence competition assays with a wide range of different VOCs and obtained values of IC50 and/or $K_{D.}$ Ligand data were collected and organized in a database to which SMILE (Simplified Molecular-Input Line-Entry system) and Isomeric SMILE were added.

In the case of OBP2 and OBP3, the collected data was analysed with the Software DataWarrior v 4.7.3 [132] used to calculate chemical descriptors (total MW, cLogP, molecular flexibility, total surface area, shape index) and analyze correlations between compounds features and their ability to interact with rat OBPs. To each tested compound, values "1" and "0" were attributed to analyze if protein was able or not to bind compound, respectively.

### 3.3.2 Homology modeling techniques

Since there was no three-dimensional structure of rat OBP2 available, homology modeling methods were used to construct a model for this protein. Firstly, the amino acid sequence of OBP2 (UniProtKB entry: Q63613), from *Rattus norvegicus*, was extracted from UniProtKB/TrEMBL [132] protein sequence database. Sequence similarity search was performed using NCBI-BLAST (Basic Local Alignment Search Tool [133]) against the Protein Data Bank, and the hOBP-2A (PDB: 4RUN [134]) reported the highest [135]sequence similarity and identity. The alignment of OBP2 sequence against hOBP-2A sequence was performed by Clustal Omega program [136] using default parameters, from which the FASTA file was created and used to build a model. Modeling was performed using Chimera software v.1.12 [137] interfaced to Modeller v.9.19 [138]. Initially, the template structure was loaded in Chimera and prepared by removing water molecules, salts and

ligands, and energy minimization was also performed (using "Minimize Structure" tool and default parameters). The prepared three-dimensional structure of the template and sequence alignment were next submitted to Modeller. Ten model structures of OBP2 were created and the top five were selected according to zDOPE score. Next, selected model structures were evaluated using PROCHECK software [139] and ProSA-web. The model with a better global score was selected to proceed with structural analysis.

## 3.3.3 Experimental methods

### 3.3.3.1 Preparation of Luria Broth (LB) liquid and LB agar medium

LB medium (25 g LB $L^{-1}$ MilliQ water) and LB agar medium (25 g LB $L^{-1}$ and 15g agar $L^{-1}$) were autoclaved at 120°C for 2 h. For LB agar containing Amp, the liquid was cooled down at room temperature and a final concentration of 100 μg $ml^{-1}$ of Amp was added under sterile conditions.

### 3.3.3.2 Transformation of the plasmids pAP006, pEX-rOBP2 and pEX-OBP3 into E. coli NZY5α competent cells

NZY5α competent cells were used to amplify the plasmids amount. pAP006 is a pET21c plasmid used for molecular cloning. pEX-OBP2 (pEX-A128-OBP2) and pEX-OBP3 (pEX-A128-OBP3) were designed and purchased with the sequence of interest. All plasmids were introduced into competent cells using thermal shock. 1 μL of each plasmid and 50 μL of Nzy5α competent cells were mixed and incubated in ice for 30 min. Afterwards the solution was subjected to a heat shock during 40 s at 42°C in a water bath and then incubated in ice for 2 min. LB medium was added to the mixture of each preparation, up to 1 mL of total volume, and it was incubated at 37°C for 1 h with a constant orbital agitation (225 rpm). Afterwards, 50 μL and 150 μL from each sample were spread on previously prepared LB/Agar/Amp plates (Amp 100 μg $mL^{-1}$). The remaining volume of each sample was centrifuged (2 min, 5000 rpm), most of the supernatant was discarded and the pellet re-suspended in a small volume of LB media. This was also spread in LB/Agar/Amp plates. Plates were incubated overnight, at 37°C and stored at 4°C.

The negative control was performed without any plasmid, just 20 μL of competent cells whereas the positive control was performed by adding 1 μl of plasmid control provided by NzyTech (pCCC) to 20 μL of competent cells.

### 3.3.3.3 Isolation and quantification of plasmid DNA (pDNA) from bacterial cells

A single colony of each plasmid was picked and added to LB media containing 100 μg $mL^{-1}$ Amp, for pre-inoculum transformation. In high-copy vectors pEX-OBP2 and pEX-OBP3 6 mL of LB

media was used, while for low-copy vector pET-21c (pAP006) 12 mL were used. Pre-inocula were prepared in duplicate and incubated overnight at 37°C with constant orbital shaking (225 rpm).

The resulting cells, from pre-inocula, were harvested by centrifuging for 2 min at 11000 rpm, and the supernatants were removed. The isolation of the pDNA was carried out with the NZYMiniprep kit, according to the supplier indications. The following volumes are referred to high copy vectors. For low-copy, the volumes of A1, A2 and A3 buffers were the double of high-copy. For cell lysis, the cell pellets were resuspended in 250 µL of Buffer A1/RNAse by vortexing. 250 µL of Buffer A2 was added, solutions were mixed by inverting and incubated for 4 min at RT. Next, 300 µL of Buffer A3 were added and gently mixed. Afterwards, the resulting solutions were centrifuged for 10 min at 11000 rpm. The resulting supernatants were loaded into NZYTech spin columns, previously placed in 2 mL collecting tubes. Then, were centrifuged for 2 min at 11000 rpm and the flow-through sample were discarded. The spin columns were washed by adding 500 µL of Buffer AY (previously heated), centrifuged 2 min at 11000 rpm and next 600 µL of Buffer A4 (containing ethanol) was added and centrifuged again. The flow-throughs were discarded, and the columns were centrifuged again for 3 min at 11000 rpm to dry the columns. For the pDNA elution, the columns were placed inside a new collecting tubes and the first elution was performed by adding 30 µL of milliQ water (previously warmed) followed by 1 min incubation in water bath at 42°C and 2 min centrifugation at 11000 rpm. The flow-through was collected and the second elution was performed by adding 50 µL of pre-heated milliQ. Finally, the pDNA concentration was quantified in Nanodrop and elution fractions were stored at -20°C.

#### 3.3.3.4 Evaluation of pDNA integrity through agarose gel electrophoresis

Agarose gel (0.8%) was prepared by dissolving 0.64 g of the agarose in 80 mL of 1xTAE pH 8.5 (40 mM Tris-base, 20 mM glacial acetic acid, 1 mM EDTA) in order to evaluate the integrity of the extracted pDNA. The solution was poured into the casting frame with the comb inserted. The gel dye used was Green Safe Premium, about 3 µL/100 mL agarose solution. The agarose gel was polymerized for 1 h and transferred to the running module. The tank was filled with TAE and the samples were applied in each well. The running conditions were set as 100 V for 1 h. The loaded samples were: NzyDNA Ladder III DNA Market, and two elutions of the extracted pEX-OBP2, pEX-OBP3 and pAP006 plasmids. The solutions for loading were prepared as followed: 2 µl of plasmid sample, diluted with 2 µL of milliQ water and 1 µL of Loading Dye. 5 µL of molecular weight marker was also loaded onto the gel. For gel revelation, it was used a UV Transiluminator.

#### 3.3.3.5 Plasmids digestion through NheI and HindIII restriction enzymes

The purified pEX-OBP2, pEX-OBP3 and pAP006 pDNA were digested with NheI and HindIII restriction enzymes according to the protocol provided by supplier. The reaction mixtures for pEX OBP2 and pEX OBP3, were prepared with a final volume of 40 µL to digest 4 µg of pDNA. Each sample was prepared by adding next reagents: purified pDNA (to final concentration 0.1 µg $\mu L^{-1}$), 4 µL of Fast Digest Buffer 10x, 2 µL of NheI and 2 µL of HindIII restriction enzymes, and sterile MilliQ water up to 40 µL final volume. In the case of pAP006, the digestion mixture was performed with 2 µg of pDNA and final volume of 20 µL. Next volumes were used for mixture preparation: 3.9 µL of pDNA, 2 µL of Fast Digest Buffer 10x, 1 µL of NheI and 1 µL of HindIII restriction enzymes, and sterile MilliQ water up to 20 µL final volume. All samples were incubated at digestion temperature 37°C for 2 h on T-100 Thermal Cycler. Then, heat inactivation of restriction enzymes was performed by incubation of the reaction mixtures for 5 min at 65°C (NheI inactivation) and for 5 min at 80°C (HindIII inactivation). To prevent recircularization of double-digested pAP006 (ddpAP006) FastAP (Thermosensitive Alkaline Phosphatase) was added to the reaction mixture in a proportion 1 µL FastAP/10 µL reaction mixture. Consequently, the mixture was incubated for 30 min at 37°C, and heat inactivation was performed by incubation at 80°C for 20 min.

Agarose gel (0.8%) was prepared (as described in 3.3.3.4) to visualize the release of inserts corresponded to genes of interest and to evaluate the digestion of transformation plasmid (pAP006). Subsequently, samples were loaded onto the gel: 5µL of NzyDNALadder III DNA Marker, undigested pEX-OBP2 and pEX-OBP3 and respective double-digested plasmids (ddpEX-OBP2 and ddpEX-OBP3). All plasmid samples were prepared as described next: 2 µL of purified plasmid was mixed with 2 µL of MilliQ water and 1 µL of Loading Dye. The gel running conditions, and revelation was performed as described in section 3.2.3.4.

#### 3.3.3.6 Purification of the inserts and double-digested plasmids by e-gel

Before the cloning, the inserts resulted from the double digestion of pEX-OBP2 and pEX-OBP3, and the ddpAP006 plasmid were purified using the E-gel equipment and purifications were performed according to the supplier instruction. 700 ng of digested pDNA in 20 µL MilliQ water were loaded in each upper-well. 5 µL of provided DNA marker was used in each e-gel and "0.8% gel run" method was chosen. Bands with 537 bp and 552 bp were collected in the collection wells for the ddpEX OBP2 and ddpEX OBP3, respectively. These bands corresponded to the released inserts which contained the genes of interest. In the case of ddpAP006 the band with approximately 5433 bp was collected. The DNA in all collected samples were quantified by Nanodrop and kept in -20°C.

#### 3.3.3.7 Cloning of OBP2 and OBP3 sequence inserts in ddpAP006 expression vector

The ligation reactions of the inserts in the double digested expression vector (ddpAP006) were performed using T4 DNA Ligase according to the supplier instruction for a sticky-end ligation. The amount of insert DNA in ligation reaction was calculated by the following Equation 3.1 considering that 25 ng of vector are needed:

$$\frac{ng(vector)\ \text{x}\ kb\ (insert\ size)}{kb\ (vector\ size)}\ \text{x}\ \frac{insert}{vector}\ (molar\ ratio) = ng\ (insert) \qquad \textbf{Equation 3.1}$$

During cloning of the OBP2 and OBP3 inserts into a vector, two molar ratios insert:vector were tested 3:1 and 10:1. All samples were prepared in 1.5 mL PCR tubes for total volume 20 µL and the following components were added by order: sterile MilliQ water up to 20 µL, 25 ng of vector DNA, calculated amount of insert (according to molar ratio), 2 µL of 10xReaction Buffer and 0.4 µL of T4 DNA ligase. The reaction mixtures were incubated in Thermal Cycler for 2 h at 22°C followed by 5 min at 70°C for T4 DNA Ligase inactivation. All samples were stored at -20°C.

#### 3.3.3.8 Transformation of the ligation reaction products in NzyStar competent cells

The ligation products obtained in 3.3.3.7 were transformed in NzyStar competent cells. 10 µL of each sample were added to the 100 µL of NzyStar competent cells. For positive control 50 µL of cells were mixed with 2 µL of control plasmid (pCCC) provided by NzyTech. In the case of negative control just 50 µL of NzyStar cells were used. The methodology used for the transformation was the same as described in section 3.3.3.2.

#### 3.3.3.9 Restriction analysis – Insert release confirmation

To confirm the presence of inserts (containing genes of interest) in plasmid vector, the restriction analysis was performed. 20 pre-inocula of selected colonies were performed (as described in 3.3.3.3): 10 corresponded to the plasmid with OBP2 insert (5 for each ration insert:vector 3:1 and 10:1) and 10 corresponded to the plasmid with OBP3 insert (5 for each ration insert:vector 3:1 and 10:1). The pDNA from the pre-inocula was extracted and purified as described in 3.3.3.3. Afterwards, 2 µg of each pDNA was digested with restriction enzymes NheI and HindIII in a final volume of 20 µL, as described in section 3.3.3.5.

To analyze the successful insertion of OBP2 insert 1.5% agarose gel was prepared and the samples were loaded with a total volume of 5 µL (2.5 µL digestion product, 1 µL of loading dye, 1.5 µL of MilliQ water). In the case of insert corresponded to OBP3, 0.8% agarose gel was prepared and 10 µL of each sample were loaded (7.5 µL of digestion product and 1.5 µL of loading dye). NzyDNA Ladder III was used as marker in both cases.

#### 3.3.3.10 Transformation of positive clones in *E. coli* BL21 (DE3) competent cells

Two positive clones were selected: one of pAP006-OBP2 and one of pAP006-OBP3. Each clone was transformed in 100 µL of *E. coli* BL21 (DE3) as described in section 3.3.3.2. For positive control 2 µL of control plasmid were transformed in 50 µL of competent cells and for negative control just 50 µL of competent cells were used. LB/Agar/Amp plates were used.

#### 3.3.3.11 Expression of the OBP2 and OBP3 proteins

6 mL of LB medium with 100 µg $mL^{-1}$ Amp was pre-inoculated with a single colony of each protein (OBP2 and OBP3) and grown for ± 9 h at 37°C and 225 rpm. Afterwards, 1 mL from each pre-inoculum was added to 50 mL of LB medium containing 100 µg $mL^{-1}$ Amp and incubated overnight in the same conditions (37°C, 225 rpm). Finally, 1 L of LB medium (with 100 µg $mL^{-1}$ Amp) was inoculated with 10 mL of the overnight culture in 5 L Erlenmeyer flasks and grown at 37°C with shaking at 225 rpm until an optical density of 0.6-0.8 at 600 nm (OD600) was reached. Next, IPTG (1 mM final concentration) was added to each flask to induce expression and small aliquots were collected every 2 h, for 6 h, for each protein. Cultures were kept overnight, and 2 more aliquots were then collected. The flasks were kept at 30°C with 225 rpm of orbital shaking during the whole expression. The protein expression overtime was monitored by OD600nm and analyzed by SDS-PAGE analysis (described in section 3.3.3.13).

#### 3.3.3.12 Cellular Fractionation

Cells were harvested by centrifugation at 6500 rpm for 15 min at 4°C. Then, were resuspended in 35 mL of 20 mM PBS, pH 7.4 and subjected to three freeze/thaw cycles, to promote the fragility of the cellular membrane. Next, three passages through French Press, under high pressure (4000 psi), were performed to promote cell lysis. The obtained lysates were incubated with 10% DNAseI for 30 min in ice and centrifuged 30 min at 10000 xg at 4°C. The pellets were resuspended with 25 mL of 20 mM PBS and stored at -20°C. Supernatant fractions were then subjected to ultracentrifugation (204710 xg, 90 min at 4°C). The supernatants of ultracentrifugation were collected, and pellets resuspended in 25 mL of 20 mM PBS. The cellular fractionation was evaluated using SDS-PAGE analysis.

#### 3.3.3.13 Sample preparation and SDS-PAGE electrophoresis Analysis

12.5% acrylamide/bisacrylamide SDS-PAGE running gels were prepared according to the standard protocol and next solutions were mixed: 1.5 mL of solution I (3 M Tris Base pH 8.8); 4.16 mL of solution III (30% acrylamide/bis solution 37:5:1); 0.1 mL of 10% SDS; 4.2 mL of distilled water; 0.076 mL of 10% PSA; 0.005 mL of TEMED.

The 5% acrylamide stacking gels were prepared by mixing the following solutions: 0.9 mL Solution II (0.5M Tris Base pH 7.0); 0.6 mL of solution III; 0.036 mL of 10% SDS; 1.88 mL of distilled water; 0.027 mL of 10% APS and 0.004 mL of TEMED.

For the time course evaluation, samples volumes were normalized according to the Equation 3.2. The corresponding volumes were centrifuged for 5 min, 5000 rpm, supernatants were discarded, and pellets resuspended in 50 µL of sample buffer (5 mL of 0.5 M Tris Base, pH 6.6-6.8; 2 mL of 100% glycerol, 4 mg of Blue Bromophenol; 8 mL of 10% SDS). For the fractionation samples, 5 µL of each sample were mixed with 5 µL of sample buffer. All sample mixtures were boiled for 10 min. 15 µL of time course samples and 10 µL of fractionation samples were loaded into the gels. Low Molecular Weight (LMW) Protein Marker was used.

$$V(mL) = \frac{1.2}{Sample\ OD_{600nm}}$$ **Equation 3.2**

The SDS-PAGE gels were run for 90 min at 100 V and next stained using blue Coomassie staining method. For this, gels were transferred for 30 min, into the staining solution (1 g Coomassie Blue R-250, 15 mL glacial acetic acid, 90 mL methanol and 95 mL distilled water) and distained, overnight, with the distaining solution (75 mL glacial acetic acid, 450 mL methanol and 475 mL distilled water). The gels were visualized, with a UV Transiluminator.

## 3.4 Results and discussion

### 3.4.1 Evaluation of structural features on OBPs-VOC recognition

The proteins pOBP and bOBP were the first to be isolated and characterized and are the most extensively studied proteins of this family. Several binding studies have been performed to find features of the "best ligand". In these reports, compounds presenting diversity in size, shape, polarity and chemical class were used. However, pOBP and bOBP have displayed no high specificity to any of the tested compounds and were capable to recognize molecules belonging to different chemical classes as terpenoids, aromatic compounds, alcohols and aldehydes [126,140–142]. For example, in the case of bOBP, among 81 tested organic compounds, this protein was able to bind 64 with different chemical and structural features (Table S2). For pOBP the similar result was observed: among 23 tested compounds it was able to bind 21 (Table S3). The resolved three-dimensional structures of bOBP and pOBP in complex with different ligands compounds showed that odorant molecules stabilize mostly hydrophobic interactions with the amino acid residues forming the cavity wall. Additionally, the binding mode was defined as opportunistic given that no significant correlation was found between chemical class of the odorant and its orientation inside the binding pocket [121] (Figure 3.2).

Some animal species can express more than one subtype of OBP. For example, in rat three different OBPs were detected [127], at least eight were identified in porcupines [143], four in mouse [144], six in panda [81] and two OBPs were identified in the human genome [145]. To understand the binding properties of different OBP subtypes, several studies had been performed with different VOCs, using mostly fluorescent probe competitive assays and isothermal titration calorimetry. Relatively to pandas, among the tested compounds, AimelOBP3 was more specific for natural terpenoids and long-chain unsaturated aldehydes, while AimelOBP5 was able to bind fatty acids but not their corresponding aldehydes [81] (Table S3). Human OBP-2A bound to a broad range of different compounds through hydrophobic interactions. Nevertheless, it has a higher affinity for aldehydes and carboxylic acids [146] (Table S4). In the case of rat proteins, all three are monomeric molecules with a lipocalin folding and low sequence homology between them. It has been demonstrated that each rat OBP is naturally tuned towards a different chemical class of organic compounds [127] and that OBP2 and OBP3 are more selective.

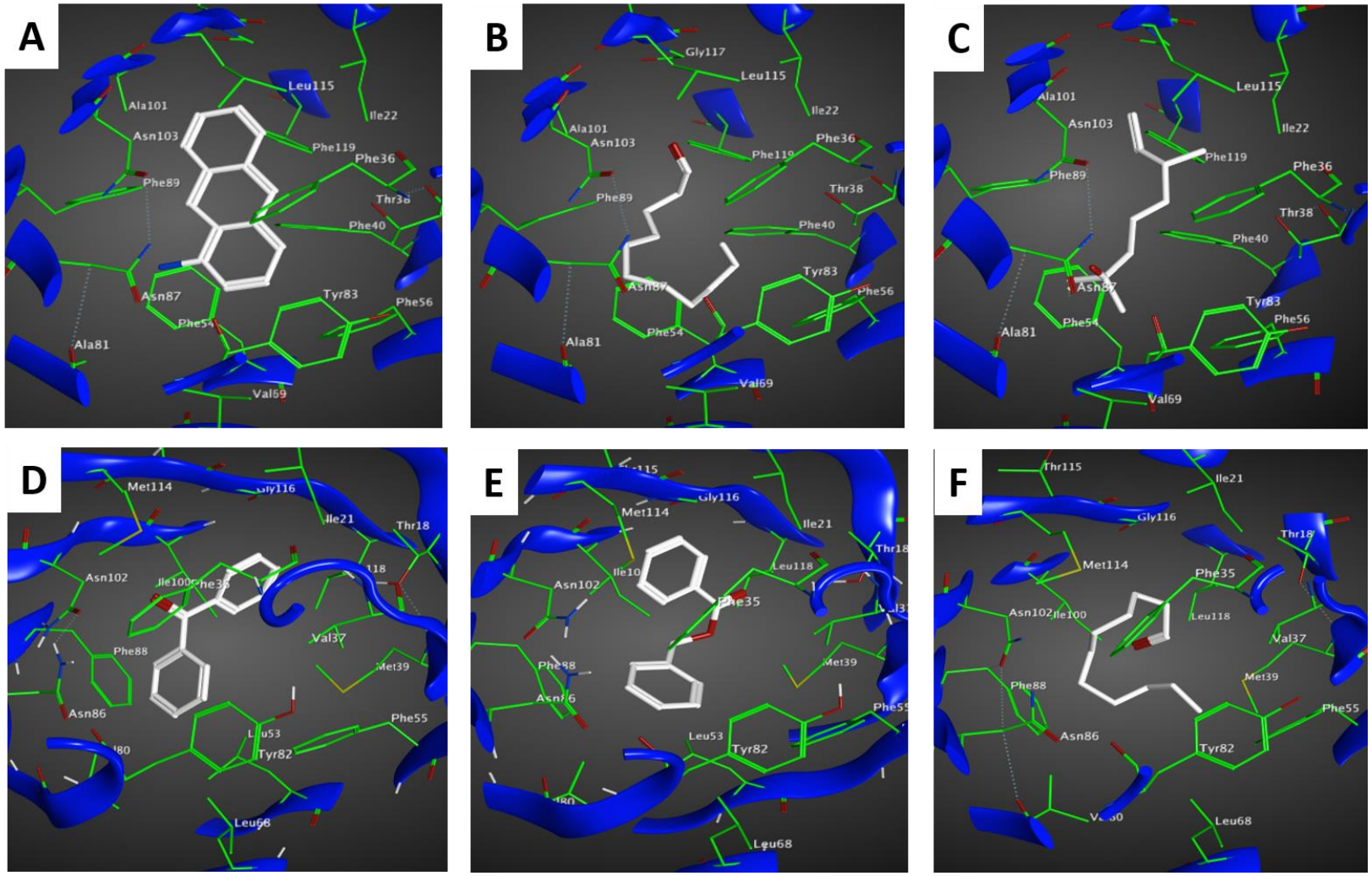


**Figure 3.2**: Representation of the binding site of porcine (pOBP) and bovine OBP (bOBP) in comlex with different ligands. Residues involved in binding pocket are colored green and ligands colored white The orientation of the ligand inside the cavity is opportunistic, as there are no specific side-chains on which a directional binding could depend. (A): bOBP with aminoanthracene (PDB: 1GT1[140]); (B) bOBP with undecanal (PDB: 1GT4[140]); (C) bOBP with dihydromyrcenol (PDB: 1GT3[140]); (D): pOBP with diphenylmethanone (PDB: 1DZP[121]); (E): pOBP with benzoic acid phenylmethylester (PDB: 1DZM[121]); (F): pOBP with undecanal (PDB: 1E02[121]).

### 3.4.2 Homology modeling

To better understand the binding properties of rat OBPs, the structure of their binding pocket should be analyzed. The only available was structure of the rat OBP3 solved by X-ray crystallography (PDB: 3ZQ3[147]). For OBP2 there was no structure available, so homology modeling techniques were applied to obtain a 3D model of the protein. Homology modeling is based on sequence similarity between a target and template and predict the conformation of the target protein, using an experimentally determined structure of a template. It relies on the fact that during evolution the structure of the family-related proteins has been more preserved than their amino acid sequence [148]. This method includes four main steps: (i) identification of a proper template; (ii) alignment of amino acid sequences of both proteins (target and template); (iii) model construction; (iv) quality check of the obtained structures [149].

The model is created based on the structural information from the template, so its selection constitutes a critical step that will influence the accuracy of the final model. Thus, proteins with percentage of sequence identity >40% [150], high structural resolution and without genetic modifications that may cause significant structural changes, should be selected as template [148].

The amino acid sequence corresponding to the OBP2 protein that was used in this work is shown in Figure 3.3. Initially, a BLAST search against the PDB was performed and the list of top five structures is shown in Table 3.2.

**Table 3.2:** Top ranking sequence alignment, using the rat OBP2 sequence of this work as query in the search. The E-value (or Expect value) is a parameter that describes the number of alignments can "expect" to see by chance when searching a database of a particular size. The lower E-value more significant the alignment is.

| Structure | PDB ID | Sequence Identity (%) | Sequence similarity (%) | Resolution (Å) | E-value |
|---|---|---|---|---|---|
| X-ray structure of human OBP-2A | 4RUN [134] | 46.0% | 63.0% | 2.6 | 7e-43 |
| NMR structure of human tear lipocalin | 5T43 | 45.0% | 58.0% | - | 5e-41 |
| X-ray structure of tear lipocalin | 1XKI [151] | 45.0% | 58.0% | 1.8 | 5e-40 |
| X-ray structure of Anticalin | 4QAF | 39.0% | 58.0% | 1.8 | 4e-29 |
| X-ray structure of murine siderocalin | 3S26 [152] | 23.0% | 51.0% | 1.8 | 6e-05 |

Human odorant binding protein, hOBP-2A, (PDB: 4RUN [134]) found had the highest sequence identity around 46%. Human tear lipocalin (TL) (PDB:1XKI [151]) was identified as the second-best template with the sequential identity around 44%. The structure of TL has higher resolution (1.8 Å) in comparison to hOBP-2A (2.6 Å), however it also has higher number of missing residues, this may influence the predicted folding and its quality. In addition, hOPB-2A and rat OBP2 belong to the same sub-family (odorant binding proteins), thus it was selected as a template. Sequence alignment of OBP2 against hOBP-2A is shown in Figure 3.3.

```
RatOBP2     QEAPPDDQEDFSGKMYTKATVCDRNHTDGKRPMKVFPMTVTALEGGDLEVRITFRGKGHC   60
hOBP2-2A    -LSFTLEEEDITGTWYVKAMVVDKDFPEDRRPRKVSPVKVTALGGGNLEATFTFMREDRC   59
              :   ::**::*.**.** * *::. :.:** ** *:.**** **:**. :**  :.:*

RatOBP2     HLRRITMHKTDEPGKYTTFKGKKTFYTKEIPVKDHYIFYIKGQRHGKSYLKGKLVGRDSK   120
hOBP2-2A    IQKKILMRKTEEPGKFSAYGGRKLIYLQELPGTDDYVFYSKDQRRGGLRYMGKLVGRNPN   119
              ::* *:**:****:::: *:* :* :*:* .*.*:** *.**:*     ******: :

RatOBP2     DNPEAMEEFKKFVKSKGFREENITVPELLDECVPGSD-----        157
hOBP2-2A    TNLEALEEFKKLVQHKGLSEEDIFMPLQTGSCVLEHHHHHHH        161
             * **:*****:*: **: **:* :*   ..**   .
```

**Figure 3.3**: Sequence alignment of rat OBP2 and hOBP-2A obtained by the Clustal Omega. * (asterisk) indicates a fully conserved residue; : (colon) indicates conservation between groups of strongly similar properties; . (period) indicates conservation between groups of weakly similar properties. Gap indicates no conservation between residues. Residues involved in the binding site of ratOBP2 are highlighted with green. Conserved Lys residues (Lys112 -hOBP2A and Lys113-ratOBP2) are highlighted with red

The Chimera software interfaced to Modeller was used to predict the three-dimensional model structure of OBP2. Modeller program implements comparative proteins structure modelling by

satisfaction of spatial restraints [138]. These spatial restraints are derived from the template and then transferred to the target sequence according to alignment by assuming that the corresponding distances and angles between aligned residues are similar [153]. The models are generated by optimization and minimization of the violations of the restrains.

Initially, ten model structures of rat OBP2 were generated and, based on the z-DOPE score, five were selected. Z-DOPE (normalized Discrete Optimized Protein Energy) is an atomic distance-dependent statistical score and its negative values indicate better models. Selected models were evaluated with the following tools:

- Software PROCHECK [139]: allows to check the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall geometry structure. Provide the Ramachandran plot residue distribution and the global G-factor. Overall G-factor evaluated the geometry of the protein structure and its values above -0.5 indicate a reliable model;
- ProSA-web [154] is a diagnostic tool based on the statistical analysis of all available protein structures and checks the structure model for potential errors, calculates an overall quality score (Z-score). Z-score measures the deviation of total energy of the structure with respect to an energy distribution derived from random conformations.

The evaluation results are shown in Table 3.3.

**Table 3.3**: Summary of the comparative homology modeling evaluation for the homology models obtained for the OBP2 amino acid sequence.

| Model | Ramachandran values (%) (PROCHECK) | | | | overall G factor (PROCHECK) | Z-score (ProSA-web) | Z-Dope |
|---|---|---|---|---|---|---|---|
| | Most favored | Allowed | Generously allowed | Disallowed | | | |
| Template (PDB:4RUN[134]) | 92.7 | 6.5 | 0.8 | 0.0 | -0.15 | -7.19 | - |
| Model 1 | 90.7 | 7.0 | 0.8 | 1.6 | -0.54 | -7.15 | -1.52 |
| Model 5 | 92.2 | 6.2 | 1.6 | 0.0 | -0.32 | -7.26 | -1.65 |
| Model 10 | 89.1 | 8.5 | 2.3 | 0.0 | -0.35 | -7.09 | -1.57 |
| Model 7 | 91.6 | 6.2 | 1.6 | 0.8 | -0.37 | -7.00 | -1.61 |
| Model 6 | 92.2 | 6.2 | 0.0 | 1.6 | -0.31 | -7.06 | -1.53 |

The Ramachandran plots from PROCHECK show that the models 1, 7 and 6 have residues with disallowed conformations. Overall G-factor for models 5,10,7 and 6 are in the range for good models. Z-score of all models were within the range of scores typically found for proteins of similar size. Model 5 (Figure 3.4) was selected to proceed with analysis as its score values are quite similar to those obtained for the template.

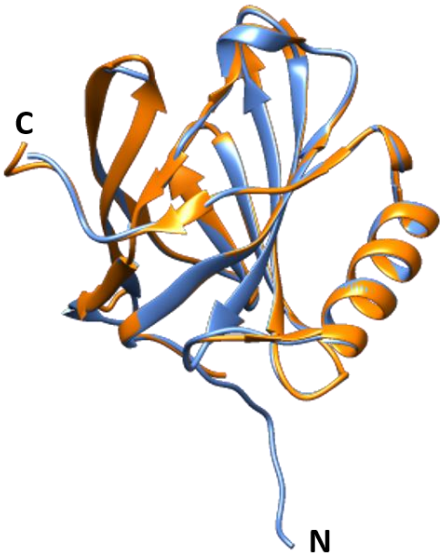


**Figure 3.4:** Homology model obtained for OBP2 (blue) with Modeller, superimposed with template hOBP-2A (orange)

### 3.4.3 OBP2 and OBP3 binding specificity

To analyze binding specificity of rat OBP2 and OBP3, experimental results of a large array of organic compounds were collected and organized from the literature. All described data ($K_D$ and IC50 values) were obtained from a competitive spectroscopic binding assay performed with a specific fluorescent chromophore as indicator. During these assays, fluorescent indicator is equilibrated with the protein then, the complex is titrated with a competitor and the displacement of probe is monitored by changes in a relative fluorescent intensity. 8-Anilinonaphthalene-1-sulfonic acid (1,8-ANS) and 1-aminoanthracene (1-AMA) have been successfully applied as a fluorescent probe for monitoring the ligand specificity in competition assay of different OBPs. This test allows to determin the concentration of competitor that caused a decay of fluoresce to half-maximal intensity (IC50) and calculate values of the $K_D$.

In total, 61 compounds were investigated (Table S5): 54 tested for OBP2 and 52 tested for OBP3. All tested compounds belong to different chemical classes, have different structure, hydrophobicity, molecular shape and flexibility.

The collected data were obtained in different experimental conditions (pH, concentrations, buffers, solutions) that may influence the final values of IC50 and $K_D$. Thus, during these analyzes

compounds were classified as binders or non-binders to the protein and attributed the score "1" and "0" respectively.

As show in Figure 3.5, OBP2 binds linear aldehydes, carboxylic acids and some hydrocarbons that display relatively high structural flexibility and molecular weight above ±150 Da. As it was already mentioned, OBP2 have a high similarity with hOBP-2A (46%). hOBP-2A also show higher specificity towards aldehydes and fatty acids. Using site-directed mutagenesis, it was demonstrated that Lys112, situated in the binding site of human proteins is mostly responsible for this specificity [146]. Figure 3.3 represents an sequence alignment of both proteins, and it possible to see that this Lys residue is also conserved in rat OBP2 (corresponds to Lys113). This suggests that, in rat OBP2, besides hydrophobic interactions ligands can stabilize hydrogen bond with Lys113 as it was observed for hOBP-2A.

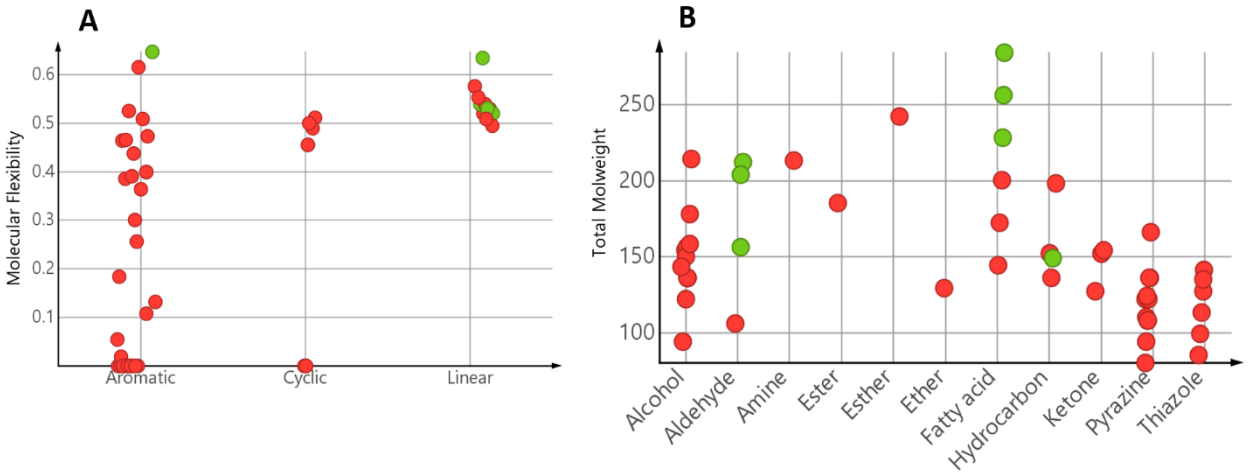


**Figure 3.5**: Distribution of compounds tested with OBP2 according to (A) their molecular flexibility and structure; (B) total molecular weight and chemical class. Green circles indicate compounds that bind and red circles, compounds that don't bind.

Figure 3.6 shows the amino acid residues of the binding pocket of OBP2 with Lys113, colored in orange, pointing into the binding pocket.

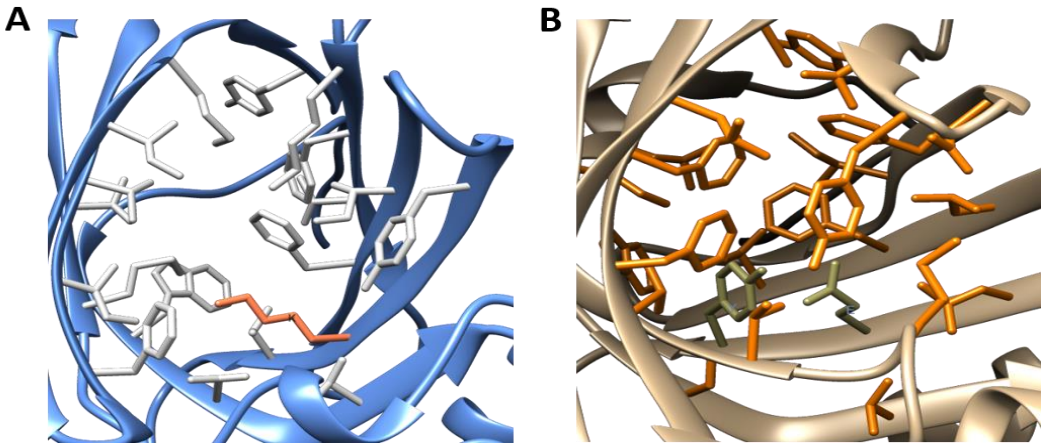


**Figure 3.6**: Representation of the amino acid residues of the binding pocket of the (A) homology model of OBP2 (Lys 113 highlighted in orange) (B) OBP3 (Glu120 and Tyr122 represented in green)

Regarding OBP3, it binds preferentially to the aromatic compounds with relatively low structural flexibility and molecular weight between 80-280 Da (Figure 3.7 - A). Most of the compounds that were able to interact derived from pyrazine and thiazole groups. In addition, alcohols, ketones and hydrocarbons stabilized interactions with the protein (Figure 3.7 – B).

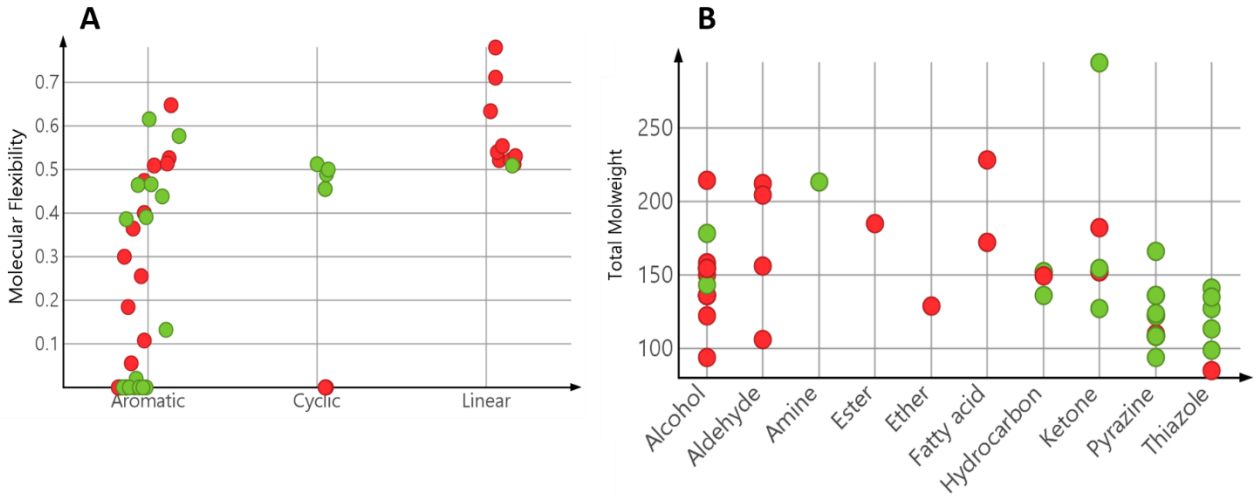


**Figure 3.7**: Distribution of compounds tested with OBP3 according to (A) their molecular flexibility and structure; (B) total molecular weight and chemical class. Green circles indicate compounds that bind and red circles, compounds that don't bind.

Using a BLAST search and OBP3 sequence (PDB:3ZQ3 [147]), mouse major urinary protein 4 (MUP4, PDB: 3KFF [155]) was identified to have a high sequential identity (~71%). The binding pocket of MUP4 contains hydrophobic residues as well as Glu118 and Tyr120 residues that can establish hydrogen bonds with ligands like alcohols, ketones pyrazine and thiazole [156,157]. Figure 3.8 shows the alignment of both amino acid sequences, and it can be observed that these residues (highlighted with red box) are also conserved in rat OBP3 (corresponded to Glu120 and Tyr122). Figure 3.6-B shows the binding pocket of OBP3 with Glu120 and Tyr122 colored in green pointing into the binding site. This observation suggests that the binding mechanism of OBP3 is similar to the one observed in MUP4.

```
RatOBP3     DPEEASFERGNLDVDKLNGDWFSIVVASDKREKIEENGSMRVFVQHIDVLENSLGFTFRI      60
MUP         --EEATSKGQNLNVEKINGEWFSILLASDKREKIEEHGSMRVFVEHIHVLENSLAFKFHT      58
              ***: :  **:*:*:**:****::**********:*******:**.******.*.*:

RatOBP3     KENGVCTEFSLVADKTAKDGEYFVEYDGENTFTILKTDYDNYVMFHLVNVNNGETFQLME      120
MUP         VIDGECSEIFLVADKTEKAGEYSVMYDGFNTFTILKTDYDNYIMFHLINEKDGKTFQLME      118
              :* *:*: ****** * *** * *** *************:****:* ::*:******

RatOBP3     LYGRTKDLSSDIKEKFAKLCVAHGITRDNIIDLTKTDRCLQA--       162
MUP         LYGRKADLNSDIKEKFVKLCEEHGIIKENIIDLTKTNRCLKARE       162
            ****. **.*******.***  *** ::********:***:*
```

**Figure 3.8:** Amino acid sequence alignment of the rat OBP3 (PDB: 3ZQ3 [147]) and MUP (PDB: 3KFF[155]). Residues involved in the binding are highlighted with green and with red.

## 3.4.4 Expression of proteins

### 3.4.4.1 Expression system and cloning strategy

OBPs are extracellular proteins and naturally they are expressed with the export signal peptide. For the expression of the recombinant OBP2 and OBP3, used in this work, the signal sequences were replaced, as explained below, to facilitate protein purification. Figure 3.9 (B and D) shows the recombinant amino acids sequences of the OBP2 and OBP3 used in this work.

**(A) Rat OBP2 nucleotide sequence**

CCAATCGCTAGCATGCATCACCATCACCATCACGACGACGACGACAAGCAGGAAGCGCCGCCG
GATGATCAGGAAGATTTTAGCGGCAAATGGTATACCAAAGCGACCGTGTGCGATCGCAACCATA
CCGATGGCAAACGCCCGATGAAAGTGTTTCCGATGACCGTGACCGCGCTGGAAGGCGGCGATC
TGGAAGTGCGCATTACCTTTCGCGGCAAAGGCCATTGCCATCTGCGCCGCATTACCATGCATAA
AACCGATGAACCGGGCAAATATACCACCTTTAAAGGCAAAAAAACCTTTTATACCAAAGAAATTCC
GGTGAAAGATCATTATATTTTTTATATTAAAGGCCAGCGCCATGGCAAAAGCTATCTGAAAGGCAA
ACTGGTGGGCCGCGATAGCAAAGATAACCCGGAAGCGATGGAAGAATTTAAAAAATTTGTGAAA
AGCAAAGGCTTTCGCGAAGAAAACATTACCGTGCCGGAACTGCTGGATGAATGCGTGCCGGGCA
GCGATGGTACCGACGACGACGACAAGTAAAAGCTTTCGAATC

**(B) Rat OBP2 amino acid sequence**

ASMHHHHHHDDDDKQEAPPDDQEDFSGKWYTKATVCDRNHTDGKRPMKVFPMTVALEGGDLEVRI
TFRGKGHCHLRRITMHKTDEPGKYTTFKGKKTFYTKEIPVKDHYIFYIKGQRHGKSYLKGKLVGRDSK
DNPEAMEEFKKFVKSKGFREENITVPELLDECVPGSD

**(C) Rat OBP3 nucleotide sequence**

CCAATCGCTAGCATGCATCACCATCACCATCACGACGACGACGACAAGGAAGAAGCGAGCTTTG
AACGCGGCAACCTGGATGTGGATAAACTGAACGGCGATTGGTTTAGCATTGTGGTGGCGAGCGA
TAAACGCGAAAAAATTGAAGAAAACGGCAGCATGCGCGTGTTTGTGCAGCATATTGATGTGCTG
GAAAACAGCCTGGGCTTTACCTTTCGCATTAAAGAAAACGGCGTGTGCACCGAATTTAGCCTGGT
GGCGGATAAAACCGCGAAAGATGGCGAATATTTTGTGGAATATGATGGCGAAAACACCTTTACCA
TTCTGAAAACCGATTATGATAACTATGTGATGTTTCATCTGGTGAACGTGAACAACGGCGAAACC
TTTCAGCTGATGGAACTGTATGGCCGCACCAAAGATCTGAGCAGCGATATTAAAGAAAAATTTGC
GAAACTGTGCGTGGCGCATGGCATTACCCGCGATAACATTATTGATCTGACCAAAACCGATCGCT
GCCTGCAGGCGCGCGGCGGTACCGACGACGACGACAAGTAAAAGCTTTCGAAT

**(D) Rat OBP3 amino acid sequence**

ASMHHHHHHDDDDKEEASFERGNLDVDKLNGDWFSIVVASDKREKIEENGSMRVFVQHIDVLENSL
GFTFRIKENGVCTEFSLVADKTAKDGEYFVEYDGENTFTILKTDYDNYVMFHLVNVNNGETFQLMELY
GRTKDLSSDIKEKFAKLCVAHGITRDNIIDLTKTDRCLQARGGT

**Figure 3.9**: (A) and (C) Designed nucleotide sequences for protein expression. (B) and (D) Amino acid sequences of the recombinant OBP2 and OBP3. Colors corresponded to: violet – restriction site of NheI; green – start codon corresponded to methionine residue; blue – His-tag; brown – EK cleavage site, red – STOP codon; gray – HindIII restriction site.

Before cloning, the DNA fragments, with the genes that encode for OBP2 and OBP3, were designed (Figure 3.9 A and C). First, the appropriate restriction enzymes sites were included in the terminals: the restriction sites for NheI in the 5'-end and for HindIII in the 3'-end to facilitate genes cloning. The DNA sequence encoding the His-tag was introduced in the N-terminal as an affinity tag, to facilitate protein purification by IMAC (Immobilized Metal Ion Affinity Chromatography) using $Ni^{2+}$ or $Co^{2+}$-loaded nitriloacetic agarose resins. An enterokinase (EK) cleavage site was incorporated to allow the removal of affinity tag. To improve the proteins' expression levels, codon optimization was also performed, by replacing the rare codons of *E. coli* strains [158]. The designed DNA fragments were synthetized by Eurofins Genomics and incorporated into pEX-A128 cloning vector. Two different plasmids were obtained pEX-OBP2 (encoding OBP2 protein) and pEX-OBP3 (encoding OBP3 protein). The pEX plasmids contain

pUC origin, which make them a high copy vector and allow to produce high number of plasmids per cell (around 500-700) and consequently a high number of inserts. The pAP006, which is modified pET21c(+) plasmid, was chosen as expression vector to be used in this work. This plasmid contains pBR322 origin which is more regulated and produce a smaller number of plasmid copies (around 30-40) per cell. The pET vectors are optimized for the cloning and expression of recombinant proteins in *E. coli* cells. In systems of this type, the expression of the target gene is placing under the control of a strong T7 promoter which is recognized, with high selectivity, by bacteriophage T7 RNAP. The cloning strategy used during this work is illustrated in Figure 3.10.

.

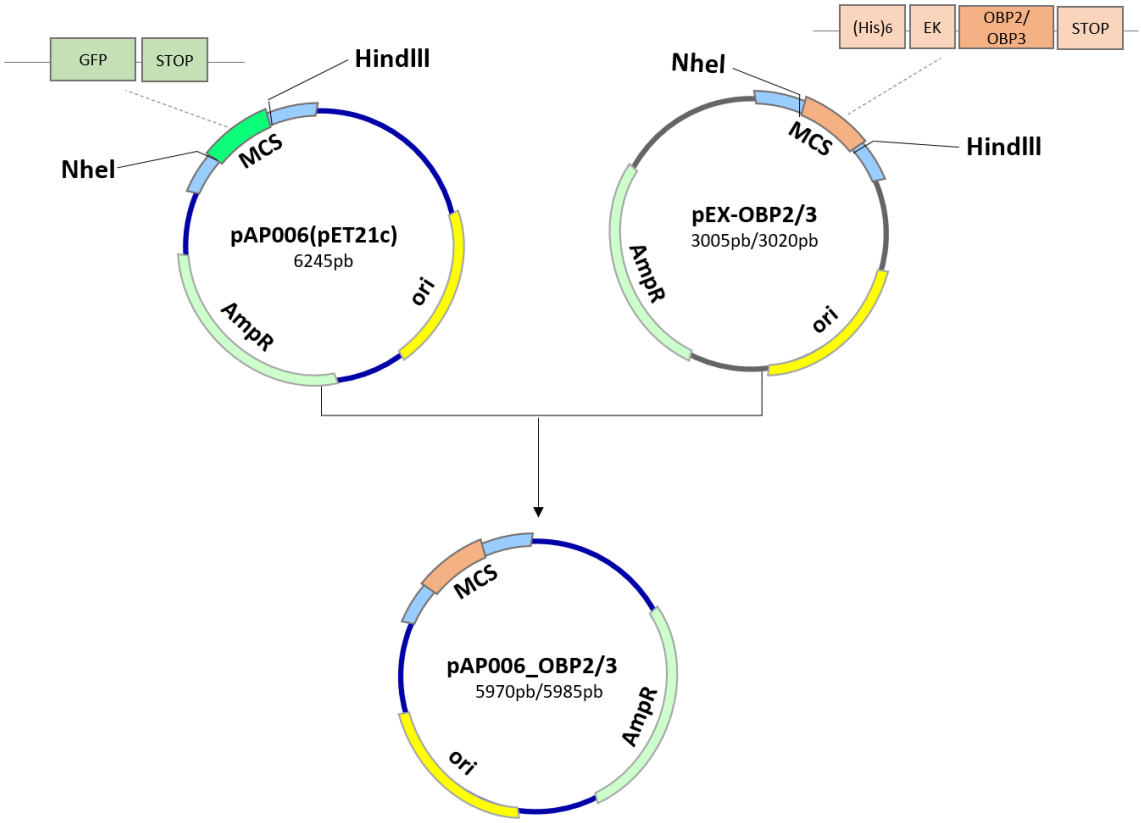


**Figure 3.10:** Representative design of the cloning strategy.

For the expression of the rat OBP proteins, *E. coli* BL21(DE3) strain was used as the host organism. Due to its genetic characteristic (Table 3.1) and the possibility to induce protein production, BL21(DE3) strain is frequently used for expression. This strain is deficient in the Lon protease and OmpT protease, which are known to degrade the abnormally folded and extracellular proteins respectively [159] the lack of these proteases resulting in a higher level of intact recombinant proteins. Plasmid degradation is prevented due to mutations in *hsdSB* and *dcm* genes. Additionally, BL21(DE3) contains λDE3 prophage that carries gene for T7 RNA polymerase (RNAP), which is expressed under control of the strong LacUV5 promoter. In the presence of

IPTG, LacUV5 drives the expression of T7 RNAP, which recognize T7 promoter and consequently start the synthesis of the protein of interest.

#### 3.4.4.2 pDNA purification and integrity evaluation

Initially, all three vectors (pEX-OBP2, pEX-OBP3 and pAP006) were successfully amplified in Nzy5α competent cells. After pDNA extraction, its concentration and purity were determined using Nanodrop spectrophotometer, by measuring the absorbance at different wavelengths (Table 3.4).

**Table 3.4**: pDNA concentration and purity after amplification. A260 corresponds to the maximum absorbance of nucleic acids; A280 the maximum absorbance of proteins; A230 corresponds to the maximum absorbance of some organic compound and salts. A260/A280 ratio of ~1.8 and A260/A230 ratio of 1.8-2.2 are generally accepted for "pure" DNA.

| Plasmid | Elution | [DNA] (ng $\mu L^{-1}$) | $A_{260}/A_{280}$ | $A_{260}/A_{230}$ |
|---|---|---|---|---|
| pAP006 | 1st | 512.0 | 1.83 | 2.20 |
| | 2nd | 170.0 | 1.89 | 2.38 |
| pEX-OBP2 | 1st | 349.7 | 1.89 | 2.23 |
| | 2nd | 62.0 | 1.90 | 2.30 |
| pEX-OBP3 | 1st | 335.5 | 1.80 | 2.33 |
| | 2nd | 131.8 | 1.87 | 2.09 |

An agarose gel analysis was performed to verify the integrity of the purified pDNA (Figure 3.11). pAP006 plasmid has 6245 bp, pEX-OBP2 and pEX-OBP3 have 3005 bp and 3020 bp respectively. In all cases a single band was observed for each sample, which indicates that there had been no hydrolysis of the pDNA. For pAP006 a band between 6000-4500 bp was observed and for pEX-OBP2 and pEX-OBP3 bands between 3000-2000 bp were detected for both. These bands corresponded to the supercoiled pDNA form, which is predominant *in vivo*, and tends to migrate faster than circular or linear forms. Samples of the second elution are less visible since they were less concentrated.

In all three plasmids, samples corresponding to the first elution step show higher plasmid concentration and acceptable purity degree, thus these samples were selected for further double-digestion.

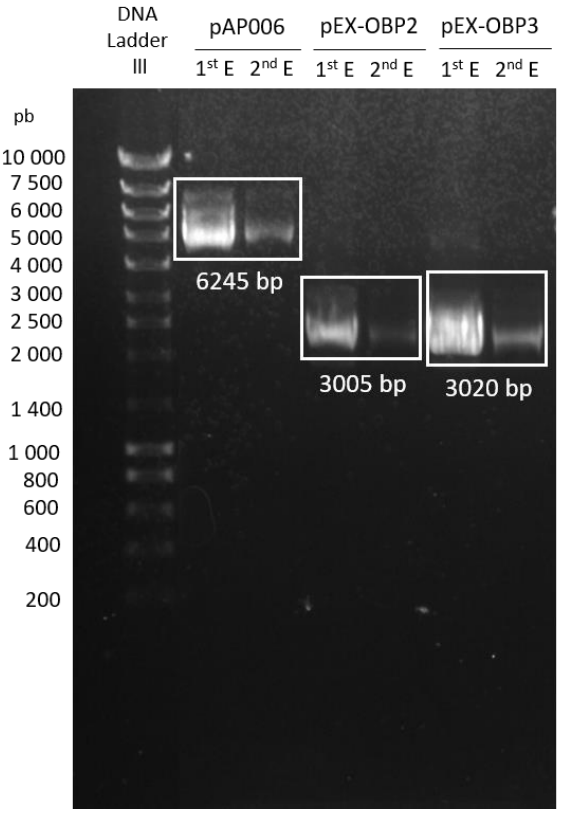


**Figure 3.11:** 0.8% (w/v) agarose gel for pDNA integrity analysis before cloning. 1st E and 2nd E correspond to first and second elution steps of the corresponded plasmid;

### 3.4.4.3 Cloning and restriction analysis

The designed DNA fragments with OBP2 and OBP3 genes were removed by double digestion with NheI and HindIII restriction enzymes. pAP006 was also double digested with the same restriction enzymes to incorporate the insert. The cloning strategy is illustrated in Figure 3.10. The success of this reaction was analyzed by a 0.8% agarose gel (Figure 3.12). In the case of pAP006 three bands were observed: 812 bp corresponded to the released fragment, 5433 bp to the double digested (dd) plasmid and 6245 bp. The last one corresponds to the single digested (sd) plasmid, which was confirmed by the performed controls (sd NheI and sd HindIII Figure 3.12 - A) in which pAP006 was digested just with one restriction enzyme. This means that one of the enzymes did not have 100% activity. The same was observed during the digestion of pEX-OBP2 (Figure 3.12 - B): sd plasmid appeared around 3000 pb, dd plasmid corresponded to the band of 2548 bp and released DNA insert fragment to the band with 537 pb. For pEX-OBP3 the band

with 552 bp corresponded to released insert DNA fragments and no sd form was observed, nevertheless it may be due to the lower concentration of this form in the sample.

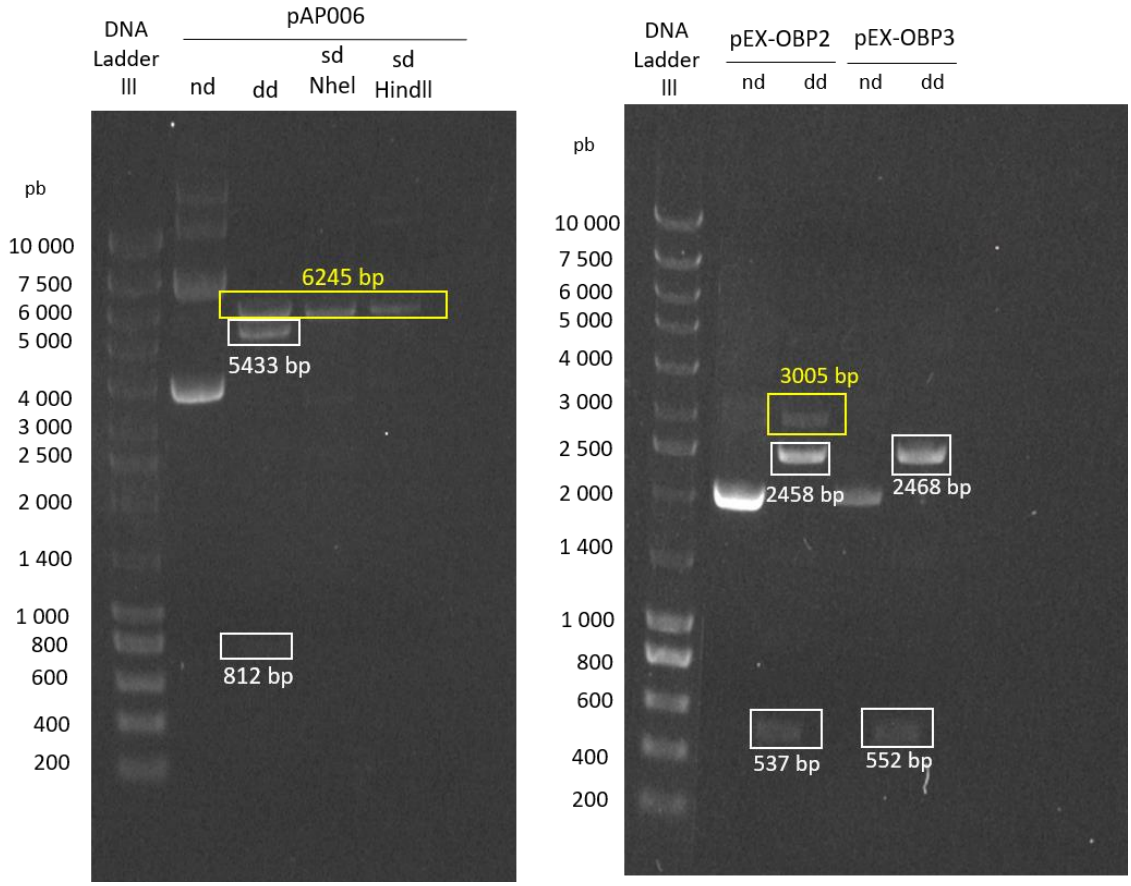


**Figure 3.12**: 0.8% agarose gels of digested plasmids. (A) - Double digested (dd) pAP006 with single digested (sd) controls and non-digested (nd) plasmid; (B) - Non-digested (nd) and double digested (dd) pEX plasmids. Band corresponded to sd plasmids are highlighted with yellow color.

dd pAP006 plasmid and insert DNA fragments were purified using e-gel. The final concentration and purity were analyzed by Nanodrop (Table 3.5).

**Table 3.5:** Quantification of the purified DNA by E-gel kit.

| Plasmid | [DNA] (ng $\mu L^{-1}$) | $A_{260}/A_{280}$ | $A_{260}/A_{230}$ |
|---|---|---|---|
| pAP006 | 6.3 | 1.19 | 0.31 |
| pEX-OBP2 | 5.4 | 2.21 | 0.47 |
| pEX-OBP3 | 6.9 | 2.22 | 0.35 |

Next, the ligation reaction was performed with T4 DNA ligases. The obtained recombinant plasmids were successfully transformed into Nzystar competent cells for propagation. Afterwards, screening by restriction digestion was performed to determine the recombinant colonies. For this, the pDNA from 20 colonies (5 for each ratio vector:insert) was isolated and

double digested with NheI and HindIII restriction enzymes. Afterwards, the digested products were analyzed in agarose gels (Figure 3.13).

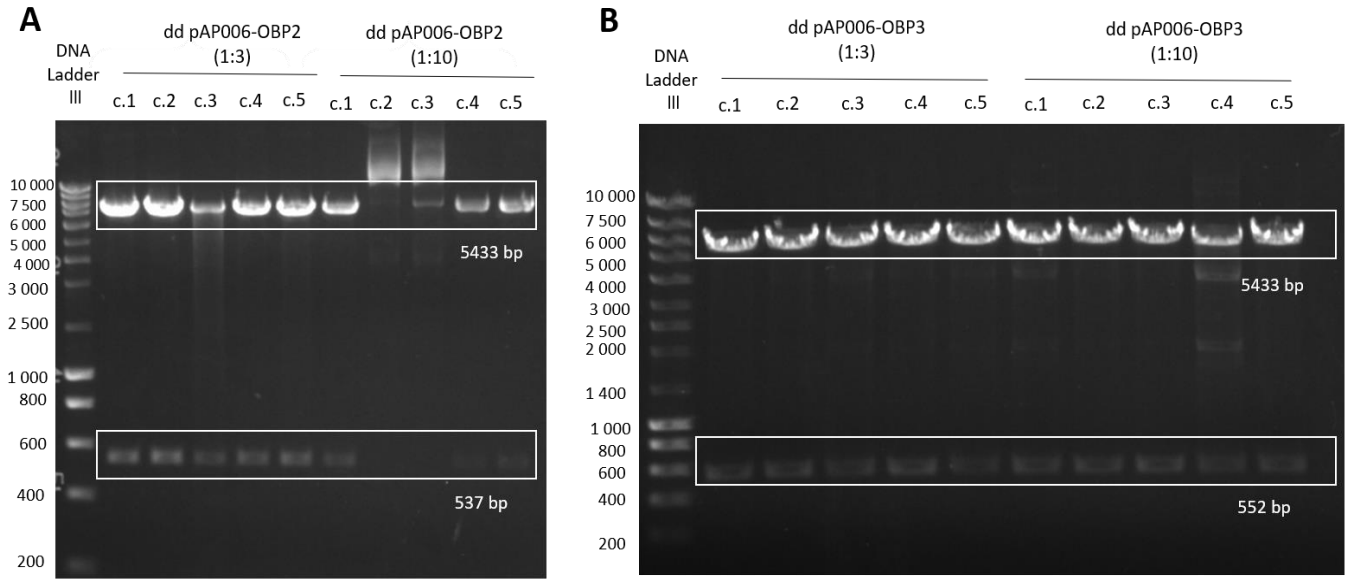


**Figure 3.13**: Hydrolysis of expression vectors with NheI and HindIII restriction enzymes for inserts release confirmation (ligation reaction confirmation). (A): release of OBP2 insert from pAP006, in a 1.5 % agarose .gel; (B): release of OBP3 insert from pAP006 in a 0.8 % agarose gel

For pAP006 with OBP2 (Figure 3.13 A), the release of the expected insert (537 bp) was observed in all colonies with 1:3 molar ratio and colonies 1, 4 and 5 with molar ratio 1:10. For colonies 2 and 3 (molar ratio 1:10) the release of the insert was not detected. This can be explained by unsuccessful ligation reaction or incomplete plasmid digestion since restriction enzymes are not totally active, as previously observed. For pAP006 with OBP3 (Figure 3.13 B), the release of the expected insert (552 bp) was detected in all screened colonies. To continue with protein expression, the recombinant plasmid, pAP006-OBP2, from colony 1 (molar ratio 1:3), and the recombinant plasmid, pAP006-OBP3, from colony 2 (molar ratio 1:3) were selected and successfully transformed into BL21(DE3) competent cells.

#### 3.4.4.4 Expression of OBP2 and OBP3

According to the literature, the expression of the recombinant OBP2 and OBP3 was mostly induced with 0.2-0.3 mM of IPTG and growth at 37°C for 3 h [128], which corresponds to *E. coli* optimal growth temperature. However, in this work the expression was induced with 1 mM of IPTG, the incubation temperature was reduced after induction to 30°C, to reduce the rates of protein production and favor proper folding [160]. The monitorization was performed for 20 h. Figure 3.14 shows that the amount of cells increase with a time until 15 h after induction (AI), which indicates that either OBP2 or OBP3 were not toxic to *E. coli*. In the case of OBP2, after 18 h, the number of cells started to decrease what can be explained by the lack of nutrients in the medium. The same can be observed for OBP3 at 15 h.

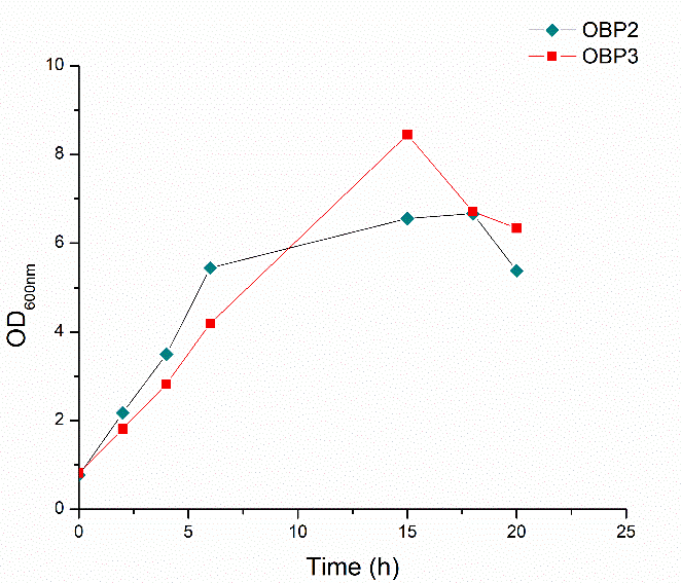


**Figure 3.14**: Optical Density measurements at 600nm after the induction (t=0h) of proteins expression.

Figure 3.15, shows time course monitoring of protein expression through SDS-PAGE. Recombinant OBP2 has a molecular weight of approximately 20.4 kDa and OBP3 around 20.8 kDa. For OBP2, protein production started, mostly, 4 h after IPTG-induction and increase over the time. In the case of OBP3, protein production was observed earlier, 2 h after induction and increase slightly over the time. There is a difference observed between the fractions before induction with IPTG and the ones after induction, indicating the proteins were well expressed over time.

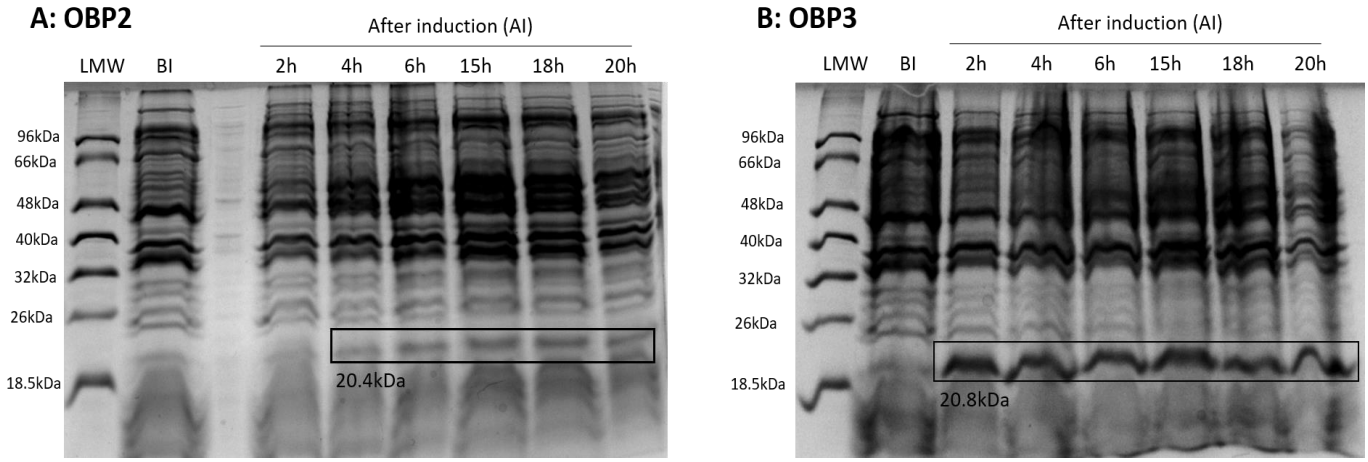


**Figure 3.15**: Time course SDS-PAGE analysis in a polyacrylamide 12.5% gel of the proteins expression: A: OBP2 and B: OBP3 in *E. coli* BL21(DE3). Recombinant proteins are highlighted with boxes. LMW – Low molecular weight marker; BI- sample before induction (t=0 h);

For further purification of the proteins, cells were harvested and lysed. Cellular fractionation was also analyzed by SDS-PAGE with a 12.5% polyacrylamide gel (Figure 3.15). Visually, the higher amount of both proteins can be found in the supernatant samples after centrifugation (SF) and ultracentrifugation (SFU). In the case of OBP3, small bands are visible in the insoluble fractions

(IF and IFU), this may explain the incomplete supernatant removing during the fractionation process. This means that it was possible to produce OBPs mostly as soluble proteins.

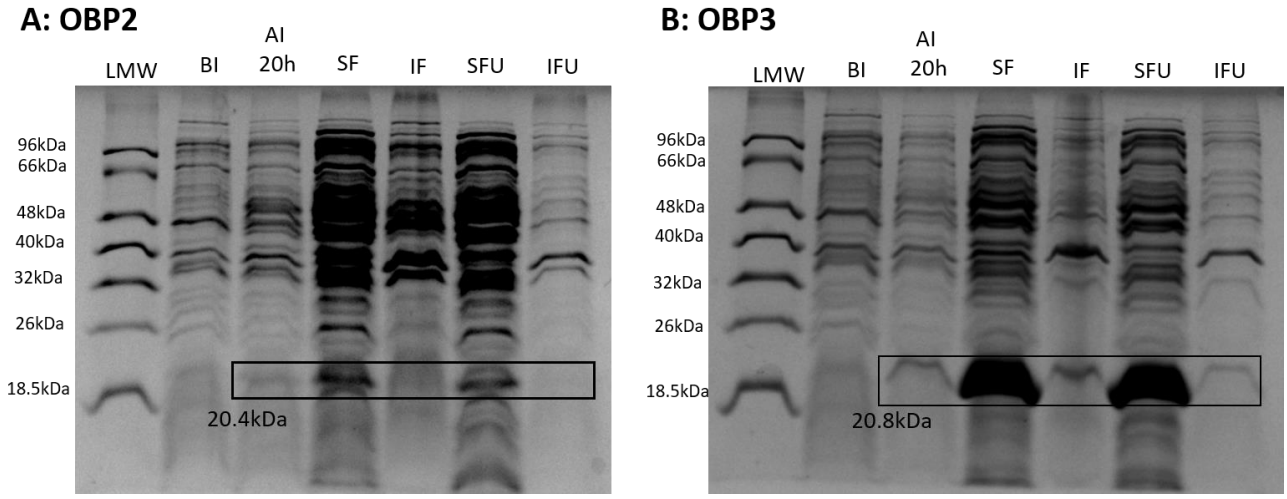


**Figure 3.16:** SDS-PAGE analysis of cellular fractionation, in a polyacrylamide 12.5% gel of the proteins expression: (A): OBP2 and (B): OBP3. LMW – Low molecular weight marker; BI- before induction; AI-20h after induction; SF- soluble fraction, supernatant after low speed centrifugation; IF-insoluble fraction, resuspended pellet after low speed centrifugation; SFU-supernatant after ultracentrifugation; IFU-resuspended pellet after ultracentrifugation.

## 3.5 Conclusions and future work

The analysis of the available experimental data demonstrated that OBP2 and OBP3 are tuned towards specific VOC ligands. The interactions between OBPs and ligands are mainly hydrophobic. However, several residues located at the binding site of each protein were identified (Lys113 for OBP2; Glu120 and Tyr122 for OBP3) as tips for site-directed mutagenesis. This may yield OBPs tuned against VOCs of interest.

Both proteins were successfully produced in the soluble form using *E. coli* expression system. Before, evaluation of their effectiveness as affinity ligands for different VOCs, proteins have to be purified. This can be performed with the IMAC chromatography system, which has a nickel metal immobilized with a high affinity towards His-tagged fusion proteins [127]. The final yield should be determined to evaluate the efficiency of the used expression system and conditions (temperature, agitation, concentration of IPTG). Moreover, OBPs folding has to be analyzed by CD to ensure that proteins remain their binding activity since incomplete folding may lead to the production of inactive proteins. To confirm that produced proteins are active, binding studies against different ligands can be performed using MST or fluorescence-based ligand binding with an appropriate probe.

Finally, the combination of rational and combinatorial approaches will be interesting for the creation of libraries of affinity reagents, tuned for binding the specific VOCs. For example, the random mutation of the specific residues involved in the binding pocket may be performed. This will generate proteins able to establish different types of interactions and/or volume of the binding pocket, which is also an important feature that affect proteins selectivity.

# CHAPTER 4

# Concluding remarks

The search for robust and versatile affinity reagents is an area of intense research for sensing, diagnostic and therapeutic approaches. This work focused on the chemical and biological production of cyclic β-hairpin peptides and rat OBPs respectively, as well as the biochemical characterization of peptides as affinity ligands. Many different fields and techniques were explored to achieve the objectives.

Cyclic-M3, cyclic-M9 and the negative control (cyclic-M0) are 14-mer synthetic peptides previously designed to mimic the interaction between the BRCT domain and phosphorylated peptides. The three peptides were successfully produced using solid-phase peptide synthesis (SPPS). The SPPS was efficient during the incorporation of non-canonical amino acid residues, namely diaminopropionic acid (Dap). The use of solid resin avoids the loss of the product during production, and facilitates its separation from the solvents and the excess of reagents, resulting in purer crudes. Moreover, an automated microwave technology lead to a higher final yield in comparison to manual synthesis. Despite all advantages, SPPS have some drawbacks associated. Despite the presence of side chain protecting groups in some amino acids, the side reaction may still occur, affecting negatively the production. Additionally, the incomplete coupling and deprotection reaction lead to the incorrect peptide assembly, as it was observed for cyclic-M3. In such cases the protocol of production could be optimized by performing double coupling of incoming amino acid residue, changing the coupling reagent or solvent system [100]. However, such optimization may increase production costs decreasing cost-efficiency. Furthermore, with an increase in peptides length, each coupling tends to be less productive, thus when double coupling is not effective, and to avoid the formation of deletion sequences, the unreacted residues have to be "blocked" by capping. This reaction will yield a truncated sequence thereby facilitating the separation of the desired product. This is frequently observed for long peptides (with more than 23 amino acids). As the peptides used in this work are relatively small, the capping was not performed.

About the affinity and selectivity of the sequences towards phosphorylated peptides, it was possible to obtain some promising results which demonstrated that rationally designed small β-hairpin peptides can be applied as affinity reagents for the purification of biological targets.

The third chapter was mainly focused in the study and production of OBPs proteins. The extensive literature review and analysis of structural data on the mammalian OBPs have demonstrated that they are mostly transport proteins, isolated in high concentrations from nasal mucus. Generally, OBPs (as bOBP, pOBP) can accommodate a wide range of different small volatile molecules belonging to different chemical classes, with different functional groups, structure and size with a micromolar affinity or clear specificity. This work focused mostly on rat OBP2 and rat OBP3 that have shown a higher specificity for a determinant class of volatile compounds. The homology modeling techniques have been successfully used for building an OBP2 structural model to analyze its binding pocket. The interaction that OBP2 and OBP3 with ligands establish are mostly hydrophobic, however the higher affinity for specific organic compounds and the constitution of the binding sites indicate that some amino acid residues may interact with small molecules through hydrogen bonds.

Given the size and complexity of OBPs (20.4 kDa for OBP2 and 20.8 kDa for OBP3) the recombinant proteins were produced using BL21(DE3) *E. coli* expression system.

Before the expression of proteins, the gene constructs were designed to facilitate cloning into suitable expression vectors and further purification. OBPs are naturally produced as extracellular proteins since they are expressed together with a signal peptide. Therefore, the signal peptides were replaced by a sequence coding for His-tag and EK cleavage site. As a result, OBP2 and OBP3 were successfully obtained as soluble and intracellular proteins.

The produced OBPs will be purified and characterized before proceeding with binding studies. Since the purpose of the exploring OBPs is their application in biosensors, the immobilization strategies will also be studied in the future.

# CHAPTER 5

# References


1. Skerra, A. Alternative non-antibody scaffolds for molecular recognition. *Curr. Opin. Biotechnol.* **18,** 295–304 (2007).

2. Carter, P. J. Potent antibody therapeutics by design. *Nat. Rev. Immunol.* **6,** 343–357 (2006).

3. Trilling, A. K., Beekwilder, J. & Zuilhof, H. Antibody orientation on biosensor surfaces: A minireview. *R. Soc. Chem.* **138,** 1619–1627 (2013).

4. Clonis, Y. D. Affinity chromatography matures as bioinformatic and combinatorial tools develop. *J. Chromatogr. A* **1101,** 1–24 (2006).

5. Ross, N. T., Katt, W. P. & Hamilton, A. D. Synthetic mimetics of protein secondary structure domains. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **368,** 989–1008 (2010).

6. Jonathan M. Haigh, Abid Hussain, Michael L. Mimmack, C. R. L. Affinity ligands for immunoglobulins based on the multicomponent Ugi reaction. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **877,** 1440–1452 (2009).

7. Filippusson, H., Erlendsson, L. S. & Lowe, C. R. Design, synthesis and evaluation of biomimetic affinity ligands for elastases. *J. Mol. Recognit.* **13,** 370–381 (2000).

8. Roque, A. C. A., Silva, C. S. O. & Taipa, M. Â. Affinity-based methodologies and ligands for antibody purification: Advances and perspectives. *J. Chromatogr. A* **1160,** 44–55 (2007).

9. Dias, A. M. G. C. & Roque, A. C. A. The future of protein scaffolds as affinity reagents for purification. *Biotechnol. Bioeng.* **114,** 481–491 (2017).

10. Maltezos, A. *et al.* Design, synthesis and application of benzyl-sulfonate biomimetic affinity adsorbents for monoclonal antibody purification from transgenic corn. *J. Mol. Recognit.* **27,** 19–31 (2014).

11. Barroso, T., Branco, R. J. F., Aguiar-Ricardo, A. & Roque, A. C. A. Structural evaluation of an alternative Protein A biomimetic ligand for antibody purification. *J. Comput. Aided.*

*Mol. Des.* **28,** 25–34 (2014).

12. Roque, A. C. A., Taipa, M. Â. & Lowe, C. R. Synthesis and screening of a rationally designed combinatorial library of affinity ligands mimicking protein L from Peptostreptococcus magnus. *J. Mol. Recognit.* **18,** 213–224 (2005).

13. Chen, C., Khoury, G. El & Lowe, C. R. Affinity ligands for glycoprotein purification based on the multi-component Ugi reaction. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **969,** 171–180 (2014).

14. Menegatti, S., Naik, A. D. & Carbonell, R. G. The hidden potential of small synthetic molecules and peptides as affinity ligands for bioseparations. *Pharm. Bioprocess.* **1,** 467–485 (2013).

15. Huang, P. Y. & Carbonell, R. G. Affinity purification of proteins using ligands derived from peptide libraries. *Biotechnol. Bioeng.* **47,** 288–297 (1995).

16. Dias, R. L. A. *et al.* Protein ligand design: From phage display to synthetic protein epitope mimetics in human antibody Fc-binding peptidomimetics. *J. Am. Chem. Soc.* **128,** 2726–2732 (2006).

17. Rutledge, S. E., Volkman, H. M. & Schepartz, A. Molecular Recognition of Protein Surfaces: High Affinity Ligands for the CBP KIX Domain. *J. Am. Chem. Soc.* **125,** 14336–14347 (2003).

18. Kasher, R. *et al.* Design and synthesis of peptides that bind α-bungarotoxin with high affinity. *Chem. Biol.* **8,** 147–155 (2001).

19. Gentilucci, L., De Marco, R. & Cerisoli, L. Chemical Modifications Designed to Improve Peptide Stability: Incorporation of Non-Natural Amino Acids, Pseudo-Peptide Bonds, and Cyclization. *Curr. Pharm. Des.* **16,** 3185–3203 (2010).

20. Choe, W., Durgannavar, T. A. & Chung, S. J. Fc-binding ligands of immunoglobulin G: An overview of high affinity proteins and peptides. *Materials (Basel).* **9,** 1–17 (2016).

21. Howell, S. M. *et al.* Serum stable natural peptides designed by mRNA display. *Sci. Rep.* **6008,** 1–5 (2014).

22. Millward, S. W., Fiacco, S., Austin, R. J. & Roberts, R. W. Design of cyclic peptides that bind protein surfaces with antibody-like affinity. *ACS Chem. Biol.* **2,** 625–634 (2007).

23. Recio, C., Maione, F., Iqbal, A. J., Mascolo, N. & De Feo, V. The potential therapeutic application of peptides and peptidomimetics in cardiovascular disease. *Front. Pharmacol.* **7,** 1–11 (2017).

24. Vlieghe, P., Lisowski, V., Martinez, J. & Khrestchatisky, M. Synthetic therapeutic peptides: science and market. *Drug Discov. Today* **15,** 40–56 (2010).

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

26. Whitby, L. R. *et al.* Design , Synthesis , and Validation of a β -Turn Mimetic Library Targeting Protein À Protein and Peptide À Receptor Interactions. *J. Am. Chem. Soc.* **133,** 10184–10194 (2011).

27. DeLano, W. L., Ultsch, M. H., De Vos, A. M. & Wells, J. A. Convergent solutions to binding at a protein-protein interface. *Science (80-. ).* **287,** 1279–1283 (2000).

28. Obrecht, D., Chevalier, E., Moehle, K. & Robinson, J. A. β-Hairpin protein epitope mimetic technology in drug discovery. *Drug Discov. Today Technol.* **9,** 63–69 (2012).

29. Raghavender, U. S., Aravinda, S., Rai, R., Shamala, N. & Balaram, P. Peptide hairpin nucleation with the obligatory Type I' β-turn Aib-DPro segment. *Org. Biomol. Chem.* **8,** 3133–3135 (2010).

30. Robinson, J. A. -Hairpin Peptidomimetics : Design , Structures and Biological Activities. *Acc. Chem. Res.* **41,** 1278–1288 (2008).

31. Descours, A., Moehle, K., Renard, A. & Robinson, J. A. A new family of beta-hairpin mimetics based on a trypsin inhibitor from sunflower seeds. *ChemBioChem* **3,** 318–323 (2002).

32. Athanassiou, Z. *et al.* Structural mimicry of retroviral Tat proteins by constrained β-hairpin peptidomimetics: Ligands with high affinity and selectivity for viral TAR RNA regulatory elements. *J. Am. Chem. Soc.* **126,** 6906–6913 (2004).

33. Gokhale, A. S. & Satyanarayanajois, S. Peptides and peptidomimetics as immunomodulators. *Immunotherapy* **6,** 755–774 (2014).

34. Jeschke, B. *et al.* RGD-peptides for tissue engineering of articular cartilage. *Biomaterials* **23,** 3455–3463 (2002).

35. Okarvi, S. M. Peptide-Based Radiopharmaceuticals: Future Tools for Diagnostic Imaging of Cancers and Other Diseases. *Med. Res. Rev.* **24,** 357–397 (2004).

36. Auernheimer, J. *et al.* Titanium implant materials with improved biocompatibility through coating with phosphonate-anchored cyclic RGD peptides. *ChemBioChem* **6,** 2034–2040 (2005).

37. Gao, X. & Matsui, H. Peptide-based nanotubes and their applications in bionanotechnology. *Adv. Mater.* **17,** 2037–2050 (2005).

38. Zhao, W. W., Shi, Q. H. & Sun, Y. Dual-ligand affinity systems with octapeptide ligands for affinity chromatography of hIgG and monoclonal antibody. *J. Chromatogr. A* **1369,** 64–72 (2014).

39. Fassina, G., Verdoliva, A., Odierna, M. R., Ruvo, M. & Cassini, G. Protein A mimetic peptide ligand for affinity purification of antibodies. *J. Mol. Recognit.* **9,** 564–9 (1996).

40. Kaufman, D. B. *et al.* Affinity purification of fibrinogen using a ligand from a peptide library. *Biotechnol. Bioeng.* **77,** 278–289 (2001).

41. Perret, G. & Boschetti, E. Aptamer affinity ligands in protein chromatography. *Biochimie* **145,** 98–112 (2018).

42. Nguyen, V. T., Kwon, Y. S. & Gu, M. B. Aptamer-based environmental biosensors for small molecule contaminants. *Curr. Opin. Biotechnol.* **45,** 15–23 (2017).

43. Ellington Andrew, S. J. In vitro selection of RNA molecules that bind spedific ligands. *Nature* **346,** 818–822 (1990).

44. Biroccio, A. *et al.* Selection of RNA Aptamers That Are Specific and High-Affinity Ligands of the Hepatitis C Virus RNA-Dependent RNA Polymerase Selection of RNA Aptamers That Are Specific and High-Affinity Ligands of the Hepatitis C Virus RNA-Dependent RNA Polymerase. *J. Virol.* **76,** 3688–3696 (2002).

45. Forier, C. *et al.* DNA aptamer affinity ligands for highly selective purification of human plasma-related proteins from multiple sources. *J. Chromatogr. A* **1489,** 39–50 (2017).

46. Chen, Z. *et al.* Discovery of aptamer ligands for hepatic stellate cells using SELEX. *Theranostics* **7,** 2982–2995 (2017).

47. Centi, S., Tombelli, S., Minunni, M. & Mascini, M. Aptamer-based detection of plasma proteins by an electrochemical assay coupled to magnetic beads. *Anal. Chem.* **79,** 1466–1473 (2007).

48. Kim, C. H., Lee, L. P., Min, J. R., Lim, M. W. & Jeong, S. H. An indirect competitive assay-based aptasensor for detection of oxytetracycline in milk. *Biosens. Bioelectron.* **51,** 426–430 (2014).

49. Huang, D. W., Niu, C. G., Qin, P. Z., Ruan, M. & Zeng, G. M. Time-resolved fluorescence aptamer-based sandwich assay for thrombin detection. *Talanta* **83,** 185–189 (2010).

50. Ruscito, A. & DeRosa, M. C. Small-Molecule Binding Aptamers: Selection Strategies, Characterization, and Applications. *Front. Chem.* **4,** 1–14 (2016).

51. Barbosa, A. J. M., Oliveira, A. R. & Roque, A. C. A. Protein- and Peptide-Based Biosensors in Artificial Olfaction. *Trends Biotechnol.* (2018). doi:10.1016/j.tibtech.2018.07.004

52. Pugliese, R. & Gelain, F. Peptidic Biomaterials: From Self-Assembling to Regenerative Medicine. *Trends Biotechnol.* **35,** 145–158 (2017).

53. Vazquez-Lombardi, R. *et al.* Challenges and opportunities for non-antibody scaffold drugs. *Drug Discov. Today* **20,** 1271–1283 (2015).

54. Pakulska, M. M., Miersch, S. & Shoichet, M. S. Designer protein delivery: From natural to engineered affinity-controlled release systems. *Science (80-. ).* **351,** (2016).

55. Richter, A., Eggenstein, E. & Skerra, A. Anticalins: Exploiting a non-Ig scaffold with hypervariable loops for the engineering of binding proteins. *FEBS Lett.* **588,** 213–218 (2014).

56. Nuttall, S. D. & Walsh, R. B. Display scaffolds: protein engineering for novel therapeutics. *Curr. Opin. Pharmacol.* **8,** 609–615 (2008).

57. Binz, H. K., Amstutz, P. & Plückthun, A. Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* **23,** 1257–1268 (2005).

58. Olson, C. A. & Roberts, R. W. Design, expression, and stability of a diverse protein library based on the human fibronectin type III domain. *Protein Sci.* **16,** 476–484 (2007).

59. Xu, L. *et al.* Directed evolution of high-affinity antibody mimics using mRNA display. *Chem. Biol.* **9,** 933–942 (2002).

60. Gunnarsson, L. C. *et al.* A carbohydrate binding module as a diversity-carrying scaffold. *Protein Eng. Des. Sel.* **17,** 213–221 (2004).

61. Engfeldt, T., Renberg, B., Brumer, H., Nygren, P. Á. & Karlström, A. E. Chemical synthesis of triple-labelled three-helix bundle binding proteins for specific fluorescent detection of unlabelled protein. *ChemBioChem* **6,** 1043–1050 (2005).

62. Russell, S. J., Blandl, T., Skelton, N. J. & Cochran, A. G. Stability of Cyclic -Hairpins : Asymmetric Contributions from Side Chains of a Hydrogen-Bonded Cross-Strand Residue Pair. 625–632 (2003).

63. Fernandez-millan, P., Autour, A., Ennifar, E. & Westhof, E. Crystal structure and

fluorescence properties of the iSpinach aptamer in complex with DFHBI. *RNA* **23,** 1788–1795 (2017).

64. Lendel, C., Dogan, J. & Härd, T. Structural Basis for Molecular Recognition in an Affibody : Affibody Complex. *J. Mol. Biol* **359,** 1293–1304 (2006).

65. Korndo, I. P., Schlehuber, S. & Skerra, A. Structural Mechanism of Specific Ligand Recognition by a Lipocalin Tailored for the Complexation of Digoxigenin. *J. Mol. Biol* **330,** 385–396 (2003).

66. Nygren, P. Å. Alternative binding proteins: Affibody binding proteins developed from a small three-helix bundle scaffold. *FEBS J.* **275,** 2668–2676 (2008).

67. Friedman, M. *et al.* Phage display selection of Affibody molecules with specific binding to the extracellular domain of the epidermal growth factor receptor. *Protein Eng. Des. Sel.* **20,** 189–199 (2007).

68. Grimm, S., Yu, F. & Nygren, P. Å. Ribosome display selection of a murine IgG1fab binding affibody molecule allowing species selective recovery of monoclonal antibodies. *Mol. Biotechnol.* **48,** 263–276 (2011).

69. Feldwisch, J. *et al.* Design of an Optimized Scaffold for Affibody Molecules. *J. Mol. Biol.* **398,** 232–247 (2010).

70. Renberg, B. *et al.* Affibody molecules in protein capture microarrays: Evaluation of multidomain ligands and different detection formats. *J. Proteome Res.* **6,** 171–179 (2007).

71. Choi, J.-W. *et al.* Two-Photon Tracer for Human Epidermal Growth Factor Receptor-2: Detection of Breast Cancer in a Live Tissue. *Anal. Chem.* **88,** 9412–9418 (2016).

72. Orlova, A. *et al.* Tumor imaging using a picomolar affinity HER2 binding Affibody molecule. *Cancer Res.* **66,** 4339–4348 (2006).

73. Seijsing, J., Yu, S., Frejd, F. Y., Höiden-Guthenberg, I. & Gräslund, T. In vivo depletion of serum IgG by an affibody molecule binding the neonatal Fc receptor. *Sci. Rep.* **8,** 1–9 (2018).

74. Zielinski, R. *et al.* HER2-affitoxin: A potent therapeutic agent for the treatment of HER2-overexpressing tumors. *Clin. Cancer Res.* **17,** 5071–5081 (2011).

75. Nord, K. *et al.* Recombinant human factor VIII-specific affinity ligands selected from phage-displayed combinatorial libraries of protein A. *Eur. J. Biochem.* **268,** 4269–4277 (2001).

76. Orelma, H. *et al.* Affibody conjugation onto bacterial cellulose tubes and bioseparation of human serum albumin. *RSC Adv.* **4,** 51440–51450 (2014).

77. Schlehuber, S. & Skerra, A. Tuning ligand affinity, specificity, and folding stability of an engineered lipocalin variant - A so-called 'anticalin' - Using a molecular random approach. *Biophys. Chem.* **96,** 213–228 (2002).

78. Flower, D. R. The lipocalin protein family: structure and function. *Biochem. J.* **318,** 1–14 (1996).

79. Åkerstrom, B., Flower, D. R. & Salier, J. P. Lipocalins: Unity in diversity. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1482,** 1–8 (2000).

80. Skerra, A. Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.* **13,** 167–187 (2000).

81. Zhu, J. *et al.* Reverse chemical ecology: Olfactory proteins from the giant panda and their interactions with putative pheromones and bamboo volatiles. *Proc. Natl. Acad. Sci.* **114,** 9802–9810 (2017).

82. Beste, G., Schmidt, F. S., Stibora, T. & Skerra, a. Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 1898–1903 (1999).

83. Rothe, C. & Skerra, A. Anticalin Proteins as Therapeutic Agents in Human Diseases. *BioDrugs* **32,** 233–243 (2018).

84. Eyer, F. *et al.* Intravenous application of an anticalin dramatically lowers plasma digoxin levels and reduces its toxic effects in rats. *Toxicol. Appl. Pharmacol.* **263,** 352–359 (2012).

85. Gebauer, M., Schiefner, A., Matschiner, G. & Skerra, A. Combinatorial design of an anticalin directed against the extra-domain b for the specific targeting of oncofetal fibronectin. *J. Mol. Biol.* **425,** 780–802 (2013).

86. Eggenstein, E., Eichinger, A., Kim, H. J. & Skerra, A. Structure-guided engineering of Anticalins with improved binding behavior and biochemical characteristics for application in radio-immuno imaging and/or therapy. *J. Struct. Biol.* **185,** 203–214 (2014).

87. Mercader, J. V & Skerra, A. Generation of anticalins with specificity for a nonsymmetric phthalic acid ester. *Anal. Biochem.* **308,** 269–277 (2002).

88. Hunter, T. Signaling--2000 and beyond. *Cell* **100,** 113–127 (2000).

89. Zavialova, M. G., Zgoda, V. G. & Nikolaev, E. N. Analysis of the Role of Protein

Phosphorylation in the Development of Diseases. *Biochemistry* **11,** 203–218 (2017).

90. Cohen, P. The role of protein phosphorylation in human health and disease. *Eur. J. Biochem.* **268,** 5001–5010 (2001).

91. Myung, N. H. *et al.* Evidence of DNA damage in Alzheimer disease: Phosphorylation of histone H2AX in astrocytes. *Age (Omaha).* **30,** 209–215 (2008).

92. Cañas Montalvo, B., López-Ferrer, D., Ramos-Fernández, A., Camafeita, E. & Calvo, E. Mass spectrometry technologies for proteomics. *Briefings Funct. Genomics Proteomics* **4,** 295–320 (2006).

93. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science (80-. ).* **312,** 212–217 (2006).

94. Mann, M. *et al.* Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol.* **20,** 261–268 (2002).

95. Leekyoung Hwang, Serife Ayaz-Guner, Zachery R. Gregorich, Wenxuan Cai, Santosh G. Valeja, Song Jin, and Y. G. Specific Enrichment of Phosphoproteins Using Functionalized Multivalent Nanoparticles. *Am Chem Soc.* **137,** 2432–2435 (2015).

96. Schmidt, S. R., Schweikart, F. & Andersson, M. E. Current methods for phosphoprotein isolation and enrichment. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **849,** 154–162 (2007).

97. Fílla, J. & Honys, D. Enrichment techniques employed in phosphoproteomics. *Amino Acids* **43,** 1025–1047 (2012).

98. Williams, R. S., Lee, M. S., Hau, D. D. & Glover, J. N. M. Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. *Nat. Struct. Mol. Biol.* **11,** 519–525 (2004).

99. Athanassiou, Z. *et al.* Structure-Guided Peptidomimetic Design Leads to Nanomolar - Hairpin Inhibitors of the Tat - TAR Interaction of Bovine Immunodeficiency Virus †,‡. 741–751 (2007).

100. Pires, D. A. T., Bemquerer, M. P. & Do Nascimento, C. J. Some mechanistic aspects on Fmoc solid phase peptide synthesis. *Int. J. Pept. Res. Ther.* **20,** 53–69 (2014).

101. Albericio, F. & Carpino, L. A. Coupling Reagents and Activation. *Methods Enzymol.* **289,** 104–126 (1997).

102. Mahalakshmi, R., Shanmugam, G., Polavarapu, P. L. & Balaram, P. Circular dichroism of

designed peptide helices and β-hairpins: Analysis of Trp- and Tyr-rich peptides. *ChemBioChem* **6,** 2152–2158 (2005).

103. Pastor, M. T., López de la Paz, M., Lacroix, E., Serrano, L. & Pérez-Payá, E. Combinatorial approaches: a new tool to search for highly structured beta-hairpin peptides. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 614–619 (2002).

104. Haines, L. A. *et al.* Light-activated hydrogel formation via the triggered folding and self-assembly of a designed peptide. *J. Am. Chem. Soc.* **127,** 17025–17029 (2005).

105. Pires, S. *et al.* Design of a peptidic turn with high affinity for HgII. *Inorg. Chem.* **51,** 11339–11348 (2012).

106. Fragoso, A. *et al.* Effect of the Peptidic Scaffold in Copper(II) Coordination and the Redox Properties of Short Histidine-Containing Peptides. *Chem. - A Eur. J.* **21,** 13100–13111 (2015).

107. Manke, I. a, Lowery, D. M., Nguyen, A. & Yaffe, M. B. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science (80-. ).* **302,** 636–639 (2003).

108. Wu, Q., Jubb, H. & Blundell, T. L. Phosphopeptide interactions with BRCA1 BRCT domains: More than just a motif. *Prog. Biophys. Mol. Biol.* **117,** 143–148 (2015).

109. Liu, X. & Ladias, J. A. A. Structural basis for the brca1 brct interaction with the proteins atrip and baat1. *Biochemistry* **52,** 7618–7627 (2013).

110. Amann, A. *et al.* The human volatilome : volatile organic compounds ( VOCs ) in exhaled breath , skin emanations , urine , feces and saliva. *J. Breath Res.* **8,** 1–17 (2014).

111. Palma, S. I. C. J. *et al.* Machine learning for the meta-analyses of microbial pathogens' volatile signatures. *Sci. Rep.* **8,** 1–15 (2018).

112. Schubert, J. K., Miekisch, W., Geiger, K. & Nöldge-Schomburg, G. F. E. Breath analysis in critically ill patients: potential and limitations. *Expert Rev. Mol. Diagn.* **4,** 619–629 (2004).

113. Miekisch, W., Schubert, J. K. & Noeldge-Schomburg, G. F. E. Diagnostic potential of breath analysis - Focus on volatile organic compounds. *Clin. Chim. Acta* **347,** 25–39 (2004).

114. Phan, N. T. *et al.* Analysis of volatile organic compounds released during food decaying processes. *Environ. Monit. Assess.* **184,** 1683–1692 (2012).

115. Duflos, G., Coin, V. M., Cornu, M., Antinelli, J. F. & Malle, P. Determination of volatile

compounds to characterize fish spoilage using headspace/mass spectrometry and solid-phase microextraction/gas chromatography/mass spectrometry. *J. Sci. Food Agric.* **86,** 600–611 (2006).

116. Mayr, D. *et al.* Rapid detection of meat spoilage by measuring volatile organic compounds by using proton transfer reaction mass spectrometry. *Appl. Environ. Microbioogy* **69,** 4697–4705 (2003).

117. Lewis, A. C. The changing face of urban air pollution. *Science (80-. ).* **359,** 744–745 (2018).

118. Wilson, A. D. Advances in electronic-nose technologies for the detection of volatile biomarker metabolites in the human breath. *Metabolites* **5,** 140–163 (2015).

119. Tegoni, M. *et al.* Mammalian odorant binding proteins. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1482,** 229–240 (2000).

120. Ramoni, R. *et al.* Control of domain swapping in bovine odorant-binding protein. *Biochem. J.* **365,** 739–48 (2002).

121. Vincent, F. *et al.* Complexes of porcine odorant binding protein with odorant molecules belonging to different chemical classes. *J. Mol. Biol.* **300,** 127–139 (2000).

122. Ramoni, R. *et al.* The protein scaffold of the lipocalin odorant-binding protein is suitable for the design of new biosensors for the detection of explosive components. *J. Phys. Condens. Matter* **19,** 1–7 (2007).

123. Yi, X. *et al.* Ligands binding and molecular simulation: The potential investigation of a biosensor based on an insect odorant binding protein. *Int. J. Biol. Sci.* **11,** 75–87 (2015).

124. Pelosi, P., Mastrogiacomo, R., Iovinella, I., Tuccori, E. & Persaud, K. C. Structure and biotechnological applications of odorant-binding proteins. *Appl. Microbiol. Biotechnol.* **98,** 61–70 (2014).

125. Di Pietrantonio, F. *et al.* Detection of odorant molecules via surface acoustic wave biosensor array based on odorant-binding proteins. *Biosens. Bioelectron.* **41,** 328–334 (2013).

126. Wei, Y., Brandazza, A. & Pelosi, P. Binding of polycyclic aromatic hydrocarbons to mutants of odorant-binding protein: A first step towards biosensors for environmental monitoring. *Biochim. Biophys. Acta - Proteins Proteomics* **1784,** 666–671 (2008).

127. Dietrich Löbel, Marion Jacob, M. V. and H. B. I. Odorants of Different Chemical Classes Interact with Distinct Odorant Binding Protein Subtypes. *Chem. Senses* **27,** 39–44 (2002).

128. Löbel, D., Marchese, S., Krieger, J., Pelosi, P. & Breer, H. Subtypes of odorant-binding proteins--heterologous expression and ligand binding. *Eur. J. Biochem.* **254,** 318–24 (1998).

129. Löbel, D., Strotmann, J., Jacob, M. & Breer, H. Identification of a third rat odorant-binding protein (OBP3). *Chem. Senses* **26,** 673–80 (2001).

130. Yabuki, M., Portman, K. L., Scott, D. J., Briand, L. & Taylor, A. J. DyBOBS: A dynamic biomimetic assay for odorant-binding to odor-binding protein. *Chemosens. Percept.* **3,** 108–117 (2010).

131. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28,** 235–242 (2000).

132. Sander, T., Freyss, J., Von Korff, M. & Rufener, C. DataWarrior: An open-source program for chemistry aware data visualization and analysis. *J. Chem. Inf. Model.* **55,** 460–473 (2015).

133. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).

134. Schiefner, A., Freier, R., Eichinger, A. & Skerra, A. Crystal structure of the human odorant binding protein, OBP. *Wiley Online Libr.* 1–5 (2015). doi:10.1002/prot.24797

135. Consortium, T. U. The universal protein resource (UniProt) in 2010. *Nucleic Acids Res.* **36,** 190–195 (2009).

136. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680 (1994).

137. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

138. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* **234,** 779–815 (1993).

139. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 283–291 (1993).

140. Vincent, F. *et al.* Crystal structures of bovine odorant-binding protein in complex with odorant molecules. *Eur. J. Biochem.* **271,** 3832–3842 (2004).

141. Jonathan PevsnerS, Vivian Hou, Adele M. Snowman, and S. H. S. Odoran-binding

Protein. *journalof Biol. Chem.* **265,** 6118–6125 (1990).

142. Ramoni, R. *et al.* The Insect Attractant 1-Octen-3-ol is the Natural Ligand of Bovine Odorant-binding Protein. *J. Biol. Chem.* **276,** 7150–7155 (2001).

143. Felicioli, A. *et al.* MULTIPLE TYPES A N D FORMS OF ODORANT-BINDING PROTEINS IN THE OLD-WORLD PORCUPINE HYSTRIX CRISTATA. *Comp. Biochem. Physiol.* **105,** 775–784 (1993).

144. Pes, D. & Pelosi, P. Odorant-binding proteins of the mouse. *Comp. Biochem. Physiol. - Part B* **112,** 471–479 (1995).

145. Briand, L. *et al.* Evidence of an odorant-binding protein in the human olfactory mucus: Location, structural characterization, and odorant-binding properties. *Biochemistry* **41,** 7241–7252 (2002).

146. Tcatchoff, L., Nespoulous, C., Pernollet, J. C. & Briand, L. A single lysyl residue defines the binding specificity of a human odorant-binding protein for aldehydes. *FEBS Lett.* **580,** 2102–2108 (2006).

147. Portman, K. L. *et al.* Enthalpy/Entropy Compensation E ff ects from Cavity Desolvation Underpin Broad Ligand Binding Selectivity for Rat Odorant Binding Protein 3. *Biochemistry* **53,** 2371–2379 (2014).

148. Bordoli, L. *et al.* Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **4,** 1–13 (2009).

149. Saxena, A., Sangwan, R. S. & Mishra, S. Fundamentals of Homology Modeling Steps and Comparison among Important Bioinformatics Tools: An Overview. *Science International* **1,** 237–252 (2013).

150. Larsson, P., Wallner, B., Lindahl, E. & Elofsson, A. Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Sci.* **17,** 990–1002 (2008).

151. Breustedt, D. A., Korndo, I. P., Redl, B. & Skerra, A. The 1 . 8-Å Crystal Structure of Human Tear Lipocalin Reveals an Extended Branched Cavity with Capacity for Multiple Ligands *. *J. Biol. Chem.* **280,** 484–493 (2005).

152. Bandaranayake, A. D. *et al.* Daedalus : a robust , turnkey platform for rapid production of decigram quantities of active recombinant proteins in human cell lines using novel lentiviral vectors. *Nucleic Acids Res.* **39,** 1–11 (2011).

153. Marc A. Martı-Renom, Ashley C. Stuart, Andras Fiser, Roberto Sanchez, Francisco Melo,

and A. S. COMPARATIVE PROTEIN STRUCTURE MODELING OF GENES AND GENOMES. *Rev. Biophys. Biomol. Struct.* **29,** 291–325 (2000).

154. Wiederstein, M. & Sippl, M. J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35,** 407–410 (2007).

155. Perez-miller, S., Zou, Q., Novotny, M. V & Hurley, T. D. High resolution X-ray structures of mouse major urinary protein nasal isoform in complex with pheromones. *Protein Sci.* **19,** 1469–1479 (2010).

156. Pertinhez, T. A. *et al.* The binding cavity of mouse major urinary protein is optimised for a variety of ligand binding modes. *Biochem. Biophys. Res. Commun.* **390,** 1266–1271 (2009).

157. Perez-Miller, S., Zou, Q., Novotny, M. V. & Hurley, T. D. High resolution X-ray structures of mouse major urinary protein nasal isoform in complex with pheromones. *Protein Sci.* **19,** 1469–1479 (2010).

158. Chen, D. & Texada, D. E. Low-usage codons and rare codons of Escherichia coli. *Gene Ther. Mol. Biol.* **10,** 1–12 (2006).

159. Gottesman, S. PROTEASES AND THEIR TARGETS IN *ESCHERICHIA COLI*. *Annu. Rev. Genet.* **30,** 465–506 (1996).

160. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.* **5,** 1–17 (2014).

161. Tegoni, M., Ramoni, R., Bignetti, E., Spinelli, S. & Cambillau, C. Domain swapping creates a third putative combining site in bovine odorant binding protein dimer. *Nat. Struct. Mol. Biol.* **3,** 863–867 (1996).

162. Bianchet, M. *et al.* The three-dimensional structure of bovine odorant binding protein and its mechanism of odor recognition. *Nat. Struct. Mol. Biol.* **3,** 934–939 (1996).

163. Ramoni, R. *et al.* Deswapping bovine odorant binding protein. *Biochim. Biophys. Acta* **1784,** 651–657 (2008).

164. Spinelli, S. *et al.* Articles The Structure of the Monomeric Porcine Odorant Binding Protein Sheds Light on the Domain Swapping Mechanism ‡. *Biochemistry* **37,** 7913–7918 (1998).

165. Scott, D. J. & Antoni, J. Structure of rat odorant-binding protein OBP1 at 1.6 A resolution. *Acta Crystallogr. Sect. D* **65,** 403–410 (2009).

166. Cennamo, N. *et al.* Easy to use plastic optical fiber-based biosensor for detection of butanal. *PLoS One* **10,** 1–12 (2015).

167. Golebiowski, J., Topin, J., Charlier, L. & Briand, L. Interaction between odorants and proteins involved in the perception of smell: The case of odorant-binding proteins probed by molecular modelling and biophysical data. *Flavour Fragr. J.* **27,** 445–453 (2012).

# CHAPTER 6

## Support information

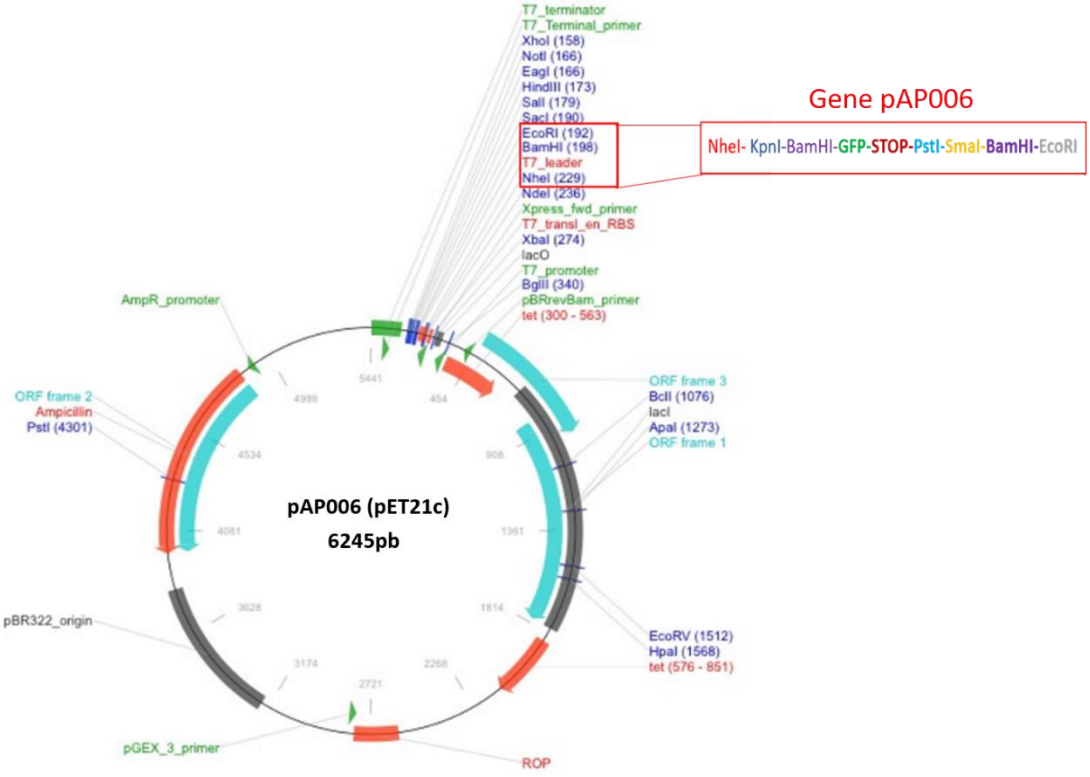


**Figure S1:** Modified pET21c plasmid pAP006. pET-21c(+) expression vector was modified at multicloning site of the plasmid. The gene pAP006 was cloned with NheI and EcorI restriction enzymes.

**Table S1**: Crystal structures of mammalian OBPs solved by X-ray crystallography and deposited in PDB.

| Source | PDB ID | Refference | Ligand ID | Ligand Formula | Ligand name |
|---|---|---|---|---|---|
| *Bos taurus* | 1G85 | [142] | 3OL | C8 H16 O | (3r)-oct-1-en-3-ol |
| | 1GT1 | [140] | ANC | C14 H11 N | anthracen-1-ylamine |
| | 1GT1 | [140] | PRZ | C9 H14 N2 O | 2-isobutyl-3-methoxypyrazine |
| | 1GT3 | [140] | DHM | C10 H20 O | 2,6-dimethyl-7-octen-2-ol |
| | 1GT4 | [140] | UNA | C11 H22 O | undecanal |
| | 1GT5 | [140] | BZQ | C13 H10 O | diphenylmethanone |
| | 1OBP | [161] | UNX | - | unknown atom or ion |
| | 1PBO | [162] | SES | C10 H18 N2 Se | 4-butyl-5-propyl-1,3-selenazol-2-amine |
| | 2HLV | [163] | LIK | C12 H20 O2 | 3,6-bis(methylene)decanoic acid |
| *Sus scrofa* | 1A3Y | [164] | - | - | - |
| | 1DZJ | [121] | SES | C10 H18 N2 Se | 4-butyl-5-propyl-1,3-selenazol-2-amine |
| | 1DZK | [121] | PRZ | C9 H14 N2 O | 2-isobutyl-3-methoxypyrazine |
| | 1DZM | [121] | BZM | C14 H12 O2 | benzoic acid phenylmethylester |
| | 1DZP | [121] | BZQ | C13 H10 O | diphenylmethanone |
| | 1E00 | [121] | DHM | C10 H20 O | 2,6-dimethyl-7-octen-2-ol |
| | 1E02 | [121] | UNA | C11 H22 O | undecanal |

| | | | | | |
|---|---|---|---|---|---|
| | 1E06 | [121] | IPB | C10 H14 O | 5-methyl-2-(1-methylethyl)phenol |
| *Rattus norvegicus* | 3FIQ | [165] | EDO | C2 H6 O2 | 1,2-ethanediol |
| | 3ZQ3 | [147] | - | - | - |
| *Homo sapiens* | 4RUN | [134] | FLC | C6 H5 O7 -3 | citrate anion |
| *Ailuropoda melanoleuca* (Giant panda) | 5NGH | [81] | TRS | C4 H12 N O3 1 | 2-amino-2-hydroxymethyl-propane-1,3-diol (tris solvent) |

**Table S2**: Binding of various chemical classes to bOBP. NB indicate that compound was tested but no binding was observed. "-" no information was found

| Ligand | Kd (uM) | Ref | IC50 (uM) | Ref |
|---|---|---|---|---|
| 1-octen-3-ol | 1.2 | [142] | - | - |
| 1-amino-anthracene (AMA) | 1 | [142] | - | - |
| Benzophenone | 0.8 | [140] | 0.6 | [141] |
| Dihydromyrcenol | 0.35 | [140] | 0.7 | [141] |
| 2-isobut yl-3-metoxypyrazine | 3.3 | [140] | - | - |
| Undecanal | 0.3 | [140] | 0.9 | [141] |
| Citralva | - | - | 0.4 | [141] |
| 3.7-Dimethyl-1-octanol | - | - | 0.3 | [141] |
| Citronellol | - | - | 1 | [141] |
| Geranyl acetone | - | - | 1.2 | [141] |
| Citronellal | - | - | 1.8 | [141] |
| Citral dimethyl acetate | - | - | 1.9 | [141] |
| 2,6-Dimethyl octane | - | - | 1.9 | [141] |
| Geranyl acetaldehyde | - | - | 2.0 | [141] |
| Retinol | - | - | 2.0 | [141] |
| b-Ionone | - | - | 2.3 | [141] |
| Dimetol | - | - | 2.6 | [141] |
| Nerol | - | - | 2.6 | [141] |
| Citronellyl acetate | - | - | 2.6 | [141] |
| Geraniol | - | - | 2.9 | [141] |
| d-carvone | - | - | 2.9 | [141] |
| Linalool | - | - | 3.1 | [141] |
| Menthone | - | - | 3.3 | [141] |
| Retinal | - | - | 3.4 | [141] |
| 3,7-Dimethyl-1-octene | - | - | 3.5 | [141] |
| a-Ionone | - | - | 4 | [141] |
| L-Carvone | - | - | 4.4 | [141] |
| Neo-alloocimene | - | - | 75 | [141] |
| Borneol | - | - | NB | [141] |
| a-Cedrene | - | - | NB | [141] |
| Cholesterol | - | - | NB | [141] |
| b-Carotene | - | - | NB | [141] |
| Amyl cinnamic aldehyde | - | - | 0.4 | [141] |
| Tetradecanal (Myristaldehyde) | - | - | 0.5 | [141] |
| Hexyl cinnamaldehyde | - | - | 0.6 | [141] |

| | | | | |
|---|---|---|---|---|
| Decanal (Decyl aldehyde) | - | - | 1.2 | [141] |
| Myrmac aldehyde | - | - | 2 | [141] |
| Cocal | - | - | 2 | [141] |
| Nonanal | - | - | 6 | [141] |
| Heptanal | - | - | 20 | [141] |
| a-Pinylisobutylaldehyde | - | - | 20 | [141] |
| Pinoacetaldehyde | - | - | 30 | [141] |
| 4-Heptenal | - | - | 60 | [141] |
| Aubepine (p-Anisaldehyde) | - | - | NB | [141] |
| 2-Methylpropanal | - | - | NB | [141] |
| Agrumen aldehyde | - | - | NB | [141] |
| Ethyl vanilin | - | - | NB | [141] |
| 3,5,5-Trimethanal | - | - | NB | [141] |
| Benzyl isovalerate | - | - | 0.2 | [141] |
| Benzyl benzoate | - | - | 0.7 | [141] |
| Octyl isovalerate | - | - | 5 | [141] |
| Octyl isobutyrate | - | - | 20 | [141] |
| Vertenex | - | - | 22 | [141] |
| Diethyl phthalate | - | - | 25 | [141] |
| Bornyl isovalerate | - | - | 30 | [141] |
| Hexyl-2-methylbutyrate | - | - | 90 | [141] |
| Isobutyl valerate | - | - | NB | [141] |
| Isoamyl isovalerate | - | - | NB | [141] |
| n-Butyl isovalerate | - | - | NB | [141] |
| Propyl isovalerate | - | - | NB | [141] |
| Musk 89 | - | - | 0.8 | [141] |
| Coniferan | - | - | 10 | [141] |
| Ambrettolide | - | - | 12 | [141] |
| Galaxolide | - | - | 12 | [141] |
| Cashmeran | - | - | 40 | [141] |
| Cis-Jasmone | - | - | 1.2 | [141] |
| Jasmal | - | - | 2 | [141] |
| Bacdanol | - | - | 6 | [141] |
| Jessemal | - | - | 6 | [141] |
| Hedione | - | - | 8 | [141] |
| Fructose | - | - | NB | [141] |
| 2-hexylpyridine | - | - | 0.5 | [141] |
| 3-hexylpyridine | - | - | 0.5 | [141] |
| 2-furfuryl mercaptan | - | - | 0.6 | [141] |

| | | | | |
|---|---|---|---|---|
| Agrumea | - | - | 8 | [141] |
| Eugenol | - | - | 12 | [141] |
| Phenethyl alcohol | - | - | 20 | [141] |
| Cinnamic aldehyde | - | - | 50 | [141] |
| Methyl isonicotinate | - | - | NB | [141] |
| 3-Ethyl-4-methylpyridine | - | - | NB | [141] |
| Quinoline | - | - | NB | [141] |

**Table S2:** Binding of various chemical classes to pOBP. NB indicate that compound was tested but no binding was observed. "-" no information was found

| **Ligand** | **Kd (uM)** | **Ref** | **IC50 (uM)** | **Ref** |
|---|---|---|---|---|
| Benzophenone | 3.1 | [140] | 3.6 | [121] |
| Dihydromyrcenol | 0.1 | [140] | 0.8 | [121] |
| Benzyl-benzoate | - | - | 3.9 | [121] |
| 2-isobutyl-3-methoxypyrazine | 0.8 | [121] | 0.9 | [121] |
| Thymol | - | - | 2.5 | [121] |
| 1-octen-3-ol | 2.7 | [140] | - | - |
| Undecanal | 2.1 | [140] | 0.7 | [121] |
| Selenazol | - | - | 1.3 | [121] |
| 1-aminoanthracene (AMA) | 1.5 | [140] | - | - |
| Butanal | 1.92 | [166] | - | - |
| Diphenyl | 0.14 | [126] | 0.4 | [126] |
| Fluoranthene | 0.5 | [126] | 1.4 | [126] |
| Pyrene | 0.18 | [126] | 0.5 | [126] |
| Phenanthrene | 0.27 | [126] | 0.75 | [126] |
| Benzo[j]fluoranthene | 0.29 | [126] | 0.8 | [126] |
| 9-phenylanthracene | 0.48 | [126] | 1.35 | [126] |
| Fluorene | 0.65 | [126] | 1.8 | [126] |
| Anthracene | 1 | [126] | 2.8 | [126] |
| Acenaphthene | NB | [126] | NB | [126] |
| Phthalazine | 0.93 | [126] | 2.6 | [126] |
| 1-pyrenebutyric acid | NB | [126] | NBV | [126] |
| Benzo[a]pyrene | 0.7 | [126] | 1.95 | [126] |

| | | | | |
|---|---|---|---|---|
| Benzo[b]fluoranthene | 0.72 | [126] | 2 | [126] |

**Table S4:** Binding of various chemical classes to pandas OBPs. NB indicate that compound was tested but no binding was observed. "-" no information was found

| | Aimel OBP3 | | Aimel OBP5 | | Ref |
|---|---|---|---|---|---|
| **Ligand** | **Kd (uM)** | **IC50 (uM)** | **Kd (uM)** | **IC50 (uM)** | |
| 9-Tetradecenal | 0.3 | 1.6 | NB | NB | [81] |
| 11-hexadecenal | 0.4 | 1.7 | NB | NB | [81] |
| 11-Hexadecenyl | NB | NB | NB | NB | [81] |
| 11-Hexadecenol | 3.8 | 18 | NB | NB | [81] |
| 9-Hexadecenal | 0.4 | 1.8 | NB | NB | [81] |
| 2-Dodecenal | 1.3 | 6 | 3.8 | 9.5 | [81] |
| Octanal | NB | NB | NB | NB | [81] |
| Decanal | NB | NB | NB | NB | [81] |
| Dodecanal | 1.3 | 6 | 8 | 20 | [81] |
| y-undecalactone | 4.7 | 22 | 8 | 20 | [81] |
| Hexadecanal | NB | NB | 4.8 | 12 | [81] |
| y-nonalactone | NB | NB | 8 | 20 | [81] |
| (-)-Carvone | 3 | 14 | 9.6 | 24 | [81] |
| (+)-Carvone | 2.3 | 11 | 9.6 | 24 | [81] |
| Citral | 0.4 | 2 | 5.6 | 14 | [81] |
| Safranal | 0.4 | 1.7 | NB | NB | [81] |
| Citronellal | 1.2 | 5.5 | NB | NB | [81] |
| Coniferyl | 3.6 | 17 | 2.8 | 7 | [81] |
| Eugenol | 6 | 28 | NB | NB | [81] |
| Citralva | 0.4 | 1.9 | 6.8 | 17 | [81] |
| Linalol | NB | NB | 9.6 | 24 | [81] |
| Geraniol | 2.1 | 10 | NB | NB | [81] |
| Phytol | NB | NB | NB | NB | [81] |
| β-Ionone | 0.3 | 1.4 | 7.2 | 18 | [81] |
| Geranyl | 0.3 | 1.4 | NB | NB | [81] |
| 3,7-Dimethyloctanol | 1.7 | 8 | NB | NB | [81] |
| Farnesol | 0.3 | 1.4 | NB | NB | [81] |
| Cedrol | 0.4 | 1.7 | 20 | 50 | [81] |
| Octanoic (caprylic) acid | NB | NB | 8 | 20 | [81] |
| Decanoic (Capric) acid | NB | NB | 5 | 12.5 | [81] |
| Dodecanoic (Lauric) acid | NB | NB | 3 | 7.5 | [81] |

| | | | | | |
|---|---|---|---|---|---|
| Tetradecanoic (Myristic) acid | NB | NB | 1.7 | 4.2 | [81] |
| Hexadecanoic (Palmitic) acid | NB | NB | 1 | 2.6 | [81] |
| Stearic acid | NB | NB | 6.4 | 16 | [81] |
| 9-hexadecenoic acid | 2.9 | 13.5 | 0.8 | 1.9 | [81] |
| Myristoleic Acid | 4.3 | 20 | 2 | 5 | [81] |
| 11-hexadecanoic acid | 4.3 | 20 | 0.8 | 2.1 | [81] |
| Z13-18-COOH | 4.3 | 20 | 0.8 | 2 | [81] |
| Oleic acid | NB | NB | 1.4 | 3.5 | [81] |
| Linoleic acid | 3.8 | 18 | 0.6 | 1.6 | [81] |
| N-phenyl-1-naphthylamine (NPN) | 0.43 | | 1.11 | | [81] |

**Table S4:** Binding of various chemical classes to rat OBPs. NB indicate that compound was tested but no binding was observed. "-" no information was found

| | OBP1 | | | | OBP2 | | | | OBP3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Ligand** | **Kdiss (uM)** | **Ref** | **IC50 (uM)** | **Ref** | **Kdiss (uM)** | **Ref** | **IC50 (uM)** | **Ref** | **Kdiss (uM)** | **Ref** | **IC50 (uM)** | **Ref** |
| (-) Borneol | 1.25 | [127] | 4.37 | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| (-) Camphor | 1.35 | [127] | 4.72 | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| (+) Camphor | 1.29 | [127] | 4.52 | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| D (-) Limonene | 1.07 | [127] | 3.85 | [127] | NB | [127] | NB | [127] | 2.41 | [127] | 8.43 | [127] |
| (+)-cis-Limonene-1,2-epoxide | 4.03 | [127] | 14.13 | [127] | NB | [127] | NB | [127] | 2.92 | [127] | 10.22 | [127] |
| (+)-trans-Limonene-1,2-epoxide | 3.98 | [127] | 14.35 | [127] | NB | [127] | NB | [127] | 2.9 | [127] | 10.15 | [127] |
| (-)-Menthone | 1.75 | [127] | 6.3 | [127] | NB | [127] | NB | [127] | 2.21 | [127] | 7.7 | [127] |
| (-)-Menthol | 2.02 | [127] | 7.3 | [127] | NB | [127] | NB | [127] | 4.12 | [127] | 14.3 | [127] |
| Phenol | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| 4-ethylphenol | 3.69 | [127] | 13.3 | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| 4-propylphenol | 0.54 | [127] | 1.89 | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| 4-isopropylphenol | 0.39 | [127] | 1.37 | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| 4-Isopropylbenzylalcohol | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| 3-isopropylphenol | 0.52 | [127] | 1.75 | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2-isopropylohenol | 0.68 | [127] | 2.38 | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| 2,6-isopropylphenol | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 0.16 | [127] | 0.58 | [127] |
| Benzaldehyde | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| 1-Decanol | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| 1-decanal | NB | [127] | NB | [127] | 1.24 | [127] | 4.35 | [127] | NB | [127] | NB | [127] |
| Decanoic (capric) Acid | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| 1-Tetreadecanol (Myristyl alcohol) | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| Tetradecanal (Myristine) aldehyde | NB | [127] | NB | [127] | 0.41 | [127] | 1.7 | [127] | NB | [127] | NB | [127] |
| Hexadecanoic (Palmitic) acid | - | - | - | - | 0.46 | [128] | 1.9 | [128] | - | - | - | - |
| Octadecanoic (Stearic) acid | - | - | - | - | 0.82 | [128] | 3.4 | [128] | - | - | - | - |
| Tetradecanoic (Myristic) acid | NB | [127] | NB | [127] | 0.29 | [127] | 1.2 | [127] | NB | [127] | NB | [127] |
| Myristyl amine | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 7.56 | [127] | 24.2 | [127] |
| Citralva | NB | [127] | NB | [127] | 2.42 | [127] | 8.5 | [127] | NB | [127] | NB | [127] |
| p-tert-Butyl-a-methyl dihydrocinnamic aldehyde | NB | [127] | NB | [127] | 1.77 | [127] | 6.2 | [127] | NB | [127] | NB | [127] |
| Pyrazine | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| 2-Methylpyrazine | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 6.97 | [127] | 24.2 | [127] |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2-Acetylpyrazine | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| 2-Methoxypyrazine | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| 2-Ethylpyrazine | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 2.15 | [127] | 8.75 | [127] |
| 2,3-Dimethylpyrazine | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 1.8 | [127] | 6.3 | [127] |
| 2,3-Diethylpyrazine | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 1.67 | [127] | 5.85 | [127] |
| 2,3,5-Trimethylpyrazine | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 1.55 | [127] | 5.42 | [127] |
| 2,3,5,6-Tetramethylpyrazine | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 1.94 | [127] | 6.8 | [127] |
| 2,6-Dimethylpyrazine | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 1.8 | [127] | 6.3 | [127] |
| 2,5-Dimethylpyrazine | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 1.32 | [127] | 4.62 | [127] |
| 2-Methyl-3-methoxypyrazine | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 8.05 | [127] | 28.2 | [127] |
| 2-Isobutyl-3-methoxypyrazine | 1.62 | [127] | 5.67 | [127] | NB | [127] | NB | [127] | 2.88 | [127] | 10.1 | [127] |
| Thiazole | NB | [127] | NB | [127] | NB | [127] | NB | [127] | - | [127] | NB | [127] |
| 5-Methylthiazole | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 3.85 | [127] | 13.45 | [127] |
| 2-Acetylthiazole | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 4.72 | [127] | 16.52 | [127] |
| 2-Ethoxythiazole | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| 2-Isobutylthiazole | 1.25 | [127] | 4.3 | [127] | NB | [127] | NB | [127] | 0.58 | [127] | 2.03 | [127] |
| 4,5-Methylthiazole | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 1.88 | [127] | 6.62 | [127] |
| 2,4,5-Methylthiazole | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 0.87 | [127] | 3.05 | [127] |
| Benzothiazole | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 1.67 | [127] | 5.85 | [127] |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4-Methyl-5-ethanol-thiazole | NB | [127] | NB | [127] | NB | [127] | NB | [127] | 7.98 | [127] | 27.95 | [127] |
| 4-Methyl-5-Acetic acid-ethylester-thiazole | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] | NB | [127] |
| Musk Ketone | - | - | - | - | - | - | NB | [127] | 1.88 | [127] | 6.62 | [127] |
| 2-isobutyl-3-methoxypyrazine | 1.7 | [167] | - | - | - | - | - | - | - | - | - | - |
| 1-amino-anthracene (AMA) | 0.6 | [167] | - | - | - | - | 10.1 | [127] | - | - | - | - |
| Eugenol | 0.046 | [167] | - | - | - | - | - | - | - | - | - | - |
| 2-Isobutyl-3-methoxypyrazine | 0.069 | [167] | - | - | - | - | - | - | - | - | - | - |
| (+)-Menthol | 0.092 | [167] | - | - | - | - | - | - | - | - | - | - |
| y-Decalactone | 0.1 | [167] | - | - | - | - | - | - | - | - | - | - |
| (-)-Menthol | 0.115 | [167] | - | - | - | - | - | - | - | - | - | - |
| Thymol | 0.125 | [167] | - | - | - | - | - | - | - | - | - | - |
| (R) -Limonene | 0.14 | [167] | - | - | - | - | - | - | - | - | - | - |
| Linanol | 0.15 | [167] | - | - | - | - | - | - | - | - | - | - |
| Coumarin | 0.21 | [167] | - | - | - | - | - | - | - | - | - | - |
| Cineole | 0.23 | [167] | - | - | - | - | - | - | - | - | - | - |
| a-Pinene | 0.29 | [167] | - | - | - | - | - | - | - | - | - | - |
| Anisole | 0.43 | [167] | - | - | - | - | - | - | - | - | - | - |
| Vanilin | 1.11 | [167] | - | - | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Guaiacol | 1.42 | [167] | - | - | - | - | - | - | - | - | - | - |
| p-Anisaldehyde | 1.42 | [167] | - | - | - | - | - | - | - | - | - | - |
| Acetoin | 10 | [167] | - | - | - | - | - | - | - | - | - | - |