Universidade do Minho
Escola de Engenharia

Ana Paula Pacheco Marques da Silva

Selection of specific colon cancer tumor peptides from combinatorial phage display libraries

Dissertação de Mestrado
Mestrado Integrado em Engenharia Biológica
Ramo de Tecnologia Química e Alimentar

Trabalho efetuado sob a orientação da
Professora Doutora Lígia Raquel Marona Rodrigues

Trabalho na empresa efetuado sob a orientação do
Doutor Leonardus Kluskens

janeiro de 2015

Nome: Ana Paula Pacheco Marques da Silva

Curso: Mestrado Integrado em Engenharia Biológica Ano de conclusão da dissertação: 2015

Área de Especialização: Ramo de Tecnologia Química e Alimentar

Escola de Engenharia, Departamento/Centro: Engenharia Biológica

TÍTULO DA DISSERTAÇÃO/TRABALHO DE PROJETO

Título em PT: Seleção de péptidos específicos para células do cancro do colon através do uso de bibliotecas de péptidos expressas na superfície de fagos

Título em EN: Selection of specific colon cancer tumor peptides from combinatorial phage display libraries

Orientador: Doutora Lígia Raquel Marona Rodrigues

Coorientador: Doutor Leonardus Kluskens

Universidade do Minho, _____/_____/_____

Assinatura:_____

“Learn from yesterday, live for today, hope for tomorrow.
The important thing is to not stop questioning.”

Albert Einstein

AGRADECIMENTOS

Aos meus orientadores, Professora Doutora Lígia Rodrigues e Doutor Leon Kluskens agradeço primeiramente por me terem dado a incrível oportunidade de trabalhar com este tema, um desafio que se mostrou imensamente gratificante e que me abriu novos horizontes e perspetivas para o futuro. Quero também agradecer toda a disponibilidade e amabilidade que sempre demonstraram ao longo destes meses de trabalho.

Ao Franklin Nóbrega e à Ivone Martins, que me acompanharam durante todo o trabalho, e que se mostraram incansáveis em atender a todas as minhas dúvidas. Sem eles não teria conseguido chegar até aqui, e por isso lhes estou extremamente grata.

Um muito obrigada a todos aqueles que tive o prazer de conhecer e que durante estes meses se tornaram verdadeiros amigos! Débora, obrigada por seres a pessoa que és, por me teres ajudado em todos os momentos e por teres esse coração enorme! João, obrigada por te disponibilizares sempre para me ajudar em tudo o que podias, por teres sempre uma coisa parva para dizer que me fazia rir mesmo nos momentos de maior stress, e por seres uma das melhores (e mais parvas) pessoas que conheci ☺. Mariana, obrigada por seres essa pessoa bondosa, sempre preocupada e sempre com uma palavra amiga para dar. Adelaide, fizeste de mim uma "filha" no laboratório :p. Obrigada por estares sempre lá comigo, por me ajudares a ultrapassar momentos difíceis, e por me apoiares em tudo, mesmo tudo o que precisei. És uma pessoa incrível, a quem devo muito. Ritinha, obrigada por teres sempre um sorriso na cara, por teres sempre um otimismo fantástico e um coração do tamanho do mundo. Obrigada, ao Rui pela amizade, ao Zé, por ter animado os meus dias com os factos mais (des)interessantes ;), e pelas conversas que ajudavam a passar as longas horas de trabalho. Ao Sílvio, por ter sempre alegria para dar (ou cantar :p). Ao Franklin, por tudo.

Aos meus amigos destes 5 anos, que se tornaram a minha família e que vou levar comigo para sempre na minha vida. Vocês, que desde o início me acompanharam, que me ajudaram a crescer e que são o melhor que levo desta Universidade.

Aos meus amigos de sempre, um grande obrigada por serem quem são, com todas as particularidades que nos fazem ser nós. Um obrigada em especial à Dulce Silva, és como uma irmã, que cresceu comigo, viveu comigo e que sempre me ajudou, e que tenho a certeza me vai acompanhar pela vida fora.

Ao Carlos Martins, obrigada por seres quem és e por fazeres parte de quem sou. Por toda a paciência que tiveste comigo durante estes meses e porque és o único que me sabe fazer sorrir, sempre.

Obrigada à minha mãe, por fazeres por mim tudo o que podes, por me fazeres sempre ambicionar ser mais e melhor, por me fazeres acreditar que sou capaz, por me ensinares o valor do esforço. Ao meu pai, que tem o maior coração do mundo, por todos os valores que sempre me transmitiste. Ao meu irmão, por me aturares sempre, nas minhas maluqueiras e nos meus stresses. Um grande obrigada a toda a minha família, que esteve sempre presente e unida durante o meu crescimento e que sempre foi um pilar para mim. Quero agradecer particularmente à minha madrinha e ao meu padrinho, que sempre me apoiaram em tudo, tanto no percurso académico como pessoal!

Por último, o meu maior obrigada ao meu avô Zé, que partiu mas que viverá sempre nas nossas memórias. Que estará sempre presente enquanto a minha família se mantiver unida, enquanto vivermos com base nos seus princípios que atravessaram e, com certeza, continuarão a atravessar gerações. Um grande obrigada, por me teres dado a minha família.

Até sempre Avô.

ABSTRACT

Colorectal cancer (CRC), localized to the large intestine and rectum, is the third most commonly diagnosed cancer and the second leading cause of cancer death. At the molecular level, CRC is a heterogeneous disease. This heterogeneity, currently not covered by available screening tests, translates into differences in the disease progression and also in the response to chemotherapeutic agents which proves that there is an undeniable need for improved diagnostic and screening tools capable of distinguish cancer at a molecular level.

The main goal of this thesis was the development of a multifunctional phage-based nanoparticle to detect and report the presence of cancer cells on a given sample through the use of a biosensor. Particularly, M13KE phage particle was genetically modified to accomplish a recombinant product capable of recognizing poorly differentiated colon cancer cells (RKO cell line) and report their presence using bioluminescence. The element of this particle that allowed targeting RKO cells was obtained using phage-displayed random peptide libraries to screen the surface of these cells towards the identification of biomarkers (i.e. membrane proteins) and specific peptide ligands. For that purpose, two types of random peptide libraries (presenting seven linear peptides - Ph.D. 7 and twelve linear peptides - Ph.D. 12) were cloned into the genome of the M13KE filamentous phage. Constructed libraries and a commercial one (Ph.D. C7C™ from New England Biolabs®) were used to perform two methodologies of phage display and biopanning in order to find specific peptide ligands with affinity to RKO cells. The affinity of the phage pools obtained through biopanning was assessed using ELISA methodology. As a result of these experiments, a new specific peptide, CIGNSNTLC, with high affinity towards RKO cells was discovered. In order to confer to the phage the ability to produce a bioluminescent signal, the insertion of NanoLuc™ gene into the M13KE genome was also attempted, at the 5' end of gene VIII in order to obtain 2700 copies of NanoLuc™ luciferase coupled to pVIII M13KE major coat protein. The presence of this enzyme on the phage would function as a reporter of the presence of a specific biomarker on a given sample and the amount of signal could be related to the amount of that biomarker.

In summary, the results herein gathered highlight that phage display is a powerful tool to find new peptide ligands capable of targeting cancer cells. In addition, the construction of a phage particle capable of recognizing and report the presence of biomarkers on a given sample will be a major advance on the use of biosensors for cancer diagnosis.

SUMÁRIO

O cancro colorectal, localizado no intestino grosso e no recto, é o terceiro cancro mais diagnosticado e a segunda causa de morte relacionada com doenças oncológicas. O cancro colorectal é uma doença muito heterogénea a nível molecular. Esta heterogeneidade que se traduz em diferenças na progressão da doença e na resposta dos pacientes aos agentes quimioterápicos, não é, atualmente, abrangida pelos testes diagnósticos usados o que prova, inegavelmente, que há uma necessidade de encontrar métodos de diagnóstico mais eficazes e capazes de distinguir esta doença ao nível molecular. O objetivo principal desta tese foi o desenvolvimento de uma nanopartícula fágica multifuncional para detetar e reportar a presença de células cancerígenas numa dada amostra, através do uso de um biossensor. Em particular, foi planeada a modificação do fago M13KE para obter um produto recombinante capaz de reconhecer células pouco diferenciadas do cancro do colon (de uma linhagem celular denominada RKO), e reportar a sua presença através de bioluminescência. O elemento desta partícula fágica que possibilita o reconhecimento das células RKO foi obtido usando bibliotecas de fagos acopladas à cápside da partícula fágica para rastrear a superfície celular destas células e identificar biomarcadores (proteínas membranares) e ligandos específicos. Com esse objetivo, dois tipos de bibliotecas fágicas (com uma sequência peptídica linear de 7 peptidos – Ph.D. 7, e de 12 péptidos – Ph.D. 12) foram geneticamente construídas usando o fago filamentoso M13KE. As bibliotecas construídas, em conjunto com uma terceira biblioteca comercial (Ph.D. C7C™ da New England Biolabs®) foram usadas para proceder a duas metodologias de *phage display* e *biopanning* para encontrar ligandos peptídicos específicos com afinidade para as células RKO. A afinidade dos complexos fágicos obtidos através dos ensaios de *biopanning* foi avaliada através de ensaios de ELISA. Como resultado das atividades experimentais apresentadas descobriu-se uma nova sequência peptídica, CIGNSNTLC, com grande afinidade para as células RKO. Com o objetivo de conferir a capacidade de produzir um sinal luminescente ao fago, tentou-se inserir o gene NanoLuc™ no vetor M13KE por forma a conseguir o acoplamento da NanoLuc® luciferase às 2700 cópias da proteína VIII da cápside do fago. Em suma, os resultados desta tese evidenciaram que o *phage display* é uma ferramenta poderosa para encontrar novos ligandos capazes de reconhecer células cancerígenas. Adicionalmente, a construção de uma partícula fágica capaz de reconhecer e reportar a presença de biomarcadores numa dada amostra será um grande avanço no uso de biosensores para o diagnóstico de cancro.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 2: STATE OF THE ART

CHAPTER 3: MATERIALS AND METHODS

CHAPTER 4: RESULTS AND DISCUSSION

LIST OF TABLES

CHAPTER 2: STATE OF THE ART

CHAPTER 3: MATERIALS AND METHODS

CHAPTER 4: RESULTS AND DISCUSSION

LIST OF ABREVIATIONS

2D-PAGE	Two Dimension Polyacrylamide Gel Electrophoresis
abs	Absorbance
ACF	Aberrant Crypt Foci
APC	Adenomatous Polyposis Coli
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BRASIL	Biopanning and rapid analysis of selective interactive ligands
BSA	Bovine serum albumin
CA 19.9	Carbohydrate Antigen 19-9
cDNA	Complementary DNA
CEA	Carcinoembryonic antigen
CRC	Colorectal Cancer
DLA	Double Layer Agar
DMEM	Dulbecco's Modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxy nucleoside triphosphates
dsDNA	Double-stranded DNA
dsM13KE	Double-stranded M13KE
E. coli	*Escherichia coli*
EDTA	Ethylenediamine tetraacetic acid
ELISA	Enzyme-Linked Imunnosorbent Assay
EU	European Union
FBS	Fetal Bovine Serum
FDA	Food and Drug Administration
FOBT	Faecal Occult Blood Test
FS	Flexible Sigmoidoscopy
FW	Forward
GEP	Gene Expression Profile
gFOBT	Guaiac-based faecal occult blood test
HER-2/neu	Human Epidermal growth factor Receptor 2
HRP	Horseradish peroxidase
ICAM- 1	Intercellular Adhesion Molecule 1
ICAT	Isotope-coded Affinity Tag
iFOBT	Immunochemical faecal occult blood test
IPTG	Isopropyl β-D-1-thiogalactopyranoside
iTRAQ	isobaric Tags for Relative and Absolute Quantitation
lacZα	lacZ alpha fragment
lacZω	lacZ omega fragment
LB	Luria-Bertani

L-DNA	Long form – DNA
LRP	Luciferase Reporter Phages
-mer	Referent to the number of amino acids
MMR	Mismatch Repair System
mRNA	Messenger Ribonucleic Acid
MS	Mass Spectrometry
MSI	Microsatellite Instability
MudPIT	Multidimensional Protein Identification Technology
OPD	o-phenylenediamine dihydrochloride
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene Glycol
PES	Polyethersulfone
pfu	Plaque-forming units
Ph.D.	Phage Display
RF	Replicative Form
RV	Reverse
SELDI-ToF	Surface-enhanced laser desorption/ionization-Time of flight
SM	Sodium - Magnesium
SOC	Super Optimal Broth with Catabolite repression
TAE	Tris-acetate-EDTA
TBST	Tris-Buffered Saline and Tween 20
TE	Tris-EDTA
TIMP-1	TIMP metallopeptidase inhibitor 1
TNM	Tumor, Nodes, Metastasis
TSS	Transformation - Storage Solution
UK	United Kingdom
USA	United States of America
wtM13KE	Wild-type M13KE

CHAPTER 1: MOTIVATION AND OBJECTIVES

1. MOTIVATION

Colon and rectal cancer have several similar features and are most often discussed together as CRC – colorectal cancer. There are approximately 1,000,000 new cases of CRC and 500,000 deaths associated with it each year. In fact, CRC represents one of the primary causes of cancer deaths in EU and the USA. Mortality rates for CRC have declined over the past two decades. Data from 2006 to 2010 indicates that this rate declined by 2.5% per year in men and by 3.0% per year in women. This decrease reflects both the declining incidence rates and improvements in early detection - meaning screening programs for population over 50 years - and treatment (American Cancer Society 2014). However, the available screening and diagnostic tests frequently lack in sensitivity for early diagnosis and often encompass a high economic burden that limits their use for screening general population worldwide.

Recent advances on genomics and proteomics have proven that CRC is a very heterogeneous disease at molecular level which translates into differences in disease progression, survival and response to chemotherapeutic agents (Newton et al. 2011; Marisa et al. 2013). The current screening methods for this disease overlook these important molecular features, meaning that many patients are misdiagnosed and that the treatment options chosen may not be the best for a particular case.

The emerging molecular methods being used to study cancer are leading to a better understanding of the disease, as well as to the discovery of potential new genomic and proteomic biomarkers which are the basis to develop biosensor platforms that can overcome the challenges in cancer diagnosis and early detection (Tothill 2009). Their principal components are a sensing platform that interacts with an analyte i.e. biomarker and a signal processor that transduces the binding impulse into a measurable signal. Biosensors integrated with bioselective bacteriophage represent a pioneering approach of this methodology, as recombinant phage with displayed peptides provides a source of high quality detection agents. In addition, it is already demonstrated that phage display libraries contain many potential probes for various types of biomolecules, including surface markers of cells and blood components (Petrenko 2008).

2. OBJECTIVES

This thesis is part of a much bigger project ongoing at Inception –LifeSciences Research and Development, a start-up company of the University of Minho, which global aim is the development of a virus-based kit that will enable the detection of all types of cancer on a fast, specific and reliable way, through the use of virus bioluminescence. The goal is to develop a new biosensor platform using luminescent virus as sensing interface that report the presence, or absence, of a specific cancer biomarker on a complex sample such as serum, blood, and urine or biopsy tissue.

In particular, the aim of this thesis is the development of a multifunctional recombinant phage particle using M13KE filamentous phage as a template. At the end, it is expected to have a genetically engineered M13KE displaying a specific peptide to RKO colon cancer cell line coupled to pIII minor coat protein and also NanoLuc™ luciferase coupled to pVIII major coat protein. The resulting phage-based nanoparticle will enable the detection of RKO cancer cells on a given sample giving a bioluminescent signal in response to their presence.

The search of the RKO-specific peptide will be performed using phage display and biopanning technologies. For those, two types of libraries containing a variety of phage-displayed peptides will be constructed by the insertion of random DNA sequences into the N-terminus of pIII gene. These two phage display libraries encompass a diversity of phage particles displaying different peptide sequences with 7 and 12 amino acids (Ph.D. 7 and Ph.D. 12, respectively) and it will be used, as well as a third commercial one (Ph.D. C7C™ from New England Biolabs) to screen the surface of RKO cells in order to discover one or more peptides with high affinity towards those cells. The screening of RKO surface will be performed using two phage display methodologies, conventional and BRASIL. The affinity of phage pools retrieved from biopanning rounds will be assessed using ELISA methodology. Finally, some clones randomly picked from the phage pools that present more affinity will be sequenced to obtain the specific peptide encoding sequences.

At the same time NanoLuc™ gene will be inserted at the N-terminus of pVIII gene in order to confer luminescence to M13KE. When accomplished, the resulting recombinant phage particle will present 2700 copies of NanoLuc™ luciferase, an enzyme capable of generating bioluminescent light through a chemical reaction with oxygen and a substrate.

1. COLON CANCER

Colorectal carcinoma (CRC) arises from the epithelial cell on the internal layer of the large intestine and it is caused by an accumulation of genetic and epigenetic changes that affect tumor suppressor genes as well as oncogenes (85% of CRC cases), or mismatch repair genes (MMR, in 15% of the cases), over several years (Fearon and Vogelstein 1990, Deschoolmeester et al. 2010). Those alterations can lead to changes in gene function which influences protein expression, structure and activity causing abnormal behavior on epithelial cells like altered metabolism, proliferation and apoptosis that are characteristic in tumor cells (Bendardaf et al. 2004). The first recognizable manifestation of epithelial alteration during colorectal tumor development occurs primarily on the crypts (invaginations of the mucosa), and is known as Aberrant Crypt Foci (ACF). Accumulation of these abnormal cells leads to the formation of adenomatous polyps – adenoma (Fig. 1 a) (Alberici 2007).

About 95% of CRC cases develop from adenomas, but only approximately 5% of adenomas grow into tumors (Shinya & Wolff 1979). When an adenoma progresses to a tumor mass, a significant number of undifferentiated cells appear – dysplasia (Fig. 1 b) (Fearon & Vogelstein 1990), with a marked pleomorphism (cell differentiation in size and shape) and an atypical nucleus size (nuclear:cytoplasm ratio close to 1) – *in situ* carcinoma (Fig. 1 c). When the mass grows substantially it can acquire the capacity to infiltrate and even destroy the closest surrounding tissues and finally to migrate to distant organs like the liver (metastatic capacity) – invasive carcinoma (Fig. 1 d) (Armaghany et al. 2012).

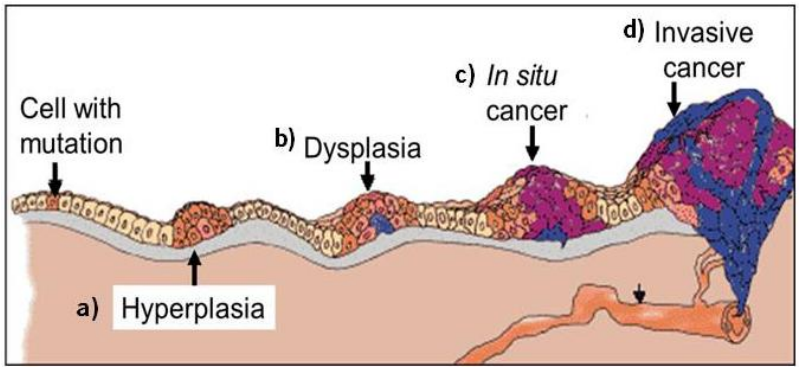

Fig. 1 – The biology of cancer: tumorigenesis process.
[taken from: http://www.ndhealthfacts.org/wiki/Oncology_%28Cancer%29 on November 10th, 2014]

The classification of CRCs, as many others, still relies on histological studies based on tumor characteristics such as differentiation status and tumor staging like, depth or size of the tumor (T), involvement of regional lymph nodes (N) and occurrence, or not, of metastasis (M). This is called the TNM staging system and it is universally used for survival prediction, treatment selection (need for radiotherapy, chemotherapy, surgical resection, among others), patient sorting for clinical trials, accurate communication among healthcare providers and also to provide uniformity on treatment options and cancer management (Ludwig & Weinstein 2005; Lin et al. 2011). The TNM classification system includes conventional prognostic factors for patient survival. Often patients at the same TNM stage for CRC have very different disease-related outcomes. For some, surgical resection of the primary tumor leads to complete recovery, while for others the metastasis and recurrence events occur even with adjuvant treatment. This shows that CRC is a very heterogeneous disease even at the same TNM stage and therefore it becomes crucial to understand the molecular processes that distinguish potential cancer subtypes. The discovery of serum and cell biomarkers that can serve to a more accurate prognosis can personalize the treatment and turn it much more effective (Duffy & Crown 2008; Walther et al. 2009).

Over the last two decades, a whole range of new technologies have been introduced in clinical practice to diagnose and treat the disease, with therapeutic modalities extending to advanced stages of the disease (Deschoolmeester et al. 2010). However, prevention remains the key to reduce morbidity and mortality since it is well established that the stage of the disease at diagnosis greatly impacts colon cancer survival rates. There are two types of screening and diagnosis techniques currently in use in the majority of EU countries, UK and USA: stool tests

(FOBT – Faecal Occult Blood Tests) and endoscopic examinations, such as flexible sigmoidoscopy (FS) and colonoscopy.

Guaiac-based faecal occult blood test (gFOBT) is, at the time, the most frequently used method in screening programs. It is a simple, inexpensive and non-invasive approach that has proven its value (Zavoral et al. 2009). With its use, a decrease in mortality rates for CRC by 15 to 33% has been verified. However, gFOBT presents relatively high false negative and false positive rates, and it has poor sensitivity for the detection of early-stage lesions (García-Bilbao et al. 2012; Burch et al. 2007). In an attempt to improve on the false positive rates of gFOBT, a new Faecal Immunochemical testing (iFOBT) has been developed. It has slightly superior performance characteristics but this comes with greatly increased financial costs which caused the failure of iFOBT implementation as a screening test for general population (García-Bilbao et al. 2012; Zavoral et al. 2009; Newton et al. 2011). On the other hand, endoscopic examinations, used many times when abnormalities appear on gFOBT results, offers significant improvements in detection rates for CRC but it also has important disadvantages associated such as high economic burden and potential major complications like bleeding or perforation (García-Bilbao et al. 2012; Winawer et al. 2003).

All this emphasizes the imperative need of new diagnostic approaches to improve the outcome of CRC screening programs. Particularly, there is a clinical need for identifying specific biomarkers for early detection of CRC, as the risk of recurrence and subsequent death due to CRC is intimately related to the stage of the disease at the time of the diagnosis (Kim et al. 2008).

2. CANCER DIAGNOSIS –THE BIOSENSORS ERA

The new strategies for cancer diagnosis at molecular level involve the development of biosensors for biomarker analysis. A biosensor (illustrated on Fig. 2) is, in a traditional sense, a bioanalytical device that incorporates a biological material, which can be microorganisms, enzymes, antibodies, among others, and have a physicochemical transducer, which may be optical, electrochemical, thermometric, piezoelectric or magnetic. The objective of using a biosensor is the production of either discrete or continuous digital electronic signals that can translate qualitative and/or quantitative information about the analyte or group of analytes (Tothill 2009; Soper et al. 2006).

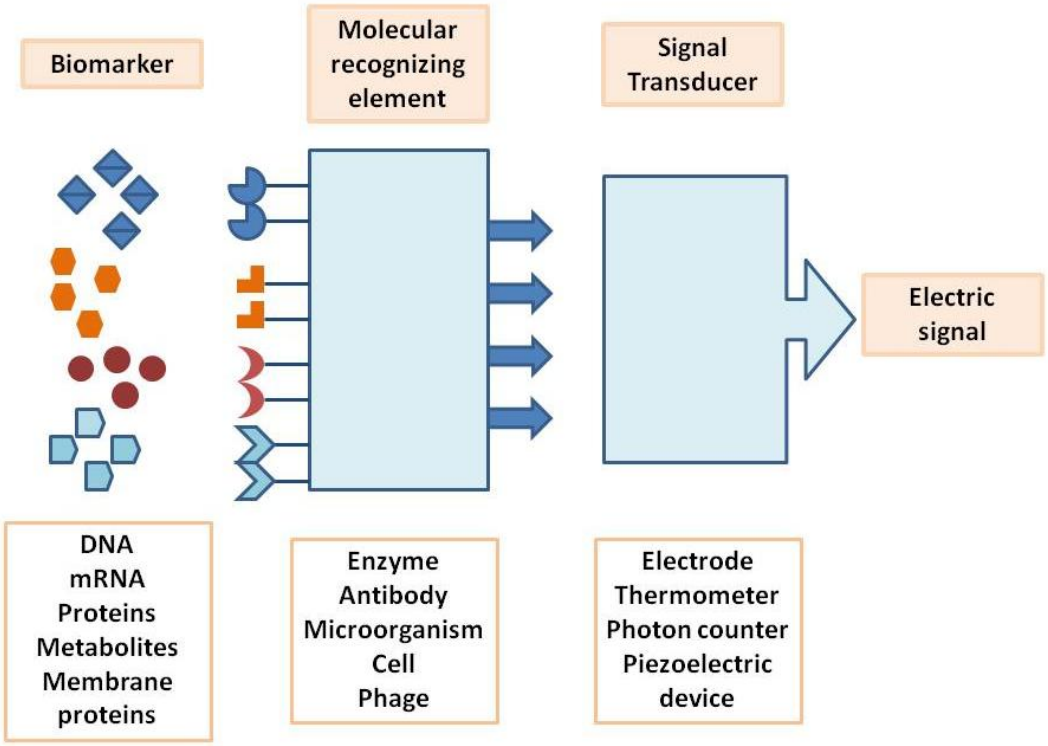

Fig. 2 – Schematic representation of a single element biosensor containing a molecular recognizing element, a transducer and the physical output whose magnitude is related with the biomarker concentration.

2.1. BIOMARKERS

The discovery of reliable biomarkers that can give trustworthy information about the presence of a tumor, but also the stage of tumorigenesis is the first step to the design of a biosensor (Soper et al. 2006). Biomarkers can be defined as measurable substances that potentially indicate a biological or pathological state. Tumor biomarkers, in particular, can be used to detect the presence of a tumor mass, predict prognosis and the response to a therapeutic agent. In order to have clinical relevance, a biomarker should lead to an improvement in life expectancy or quality of life (Newton et al. 2011; Ludwig & Weinstein 2005; Duffy & Crown 2008). Cancer biomarkers can be DNA, mRNA, proteins, metabolites, whole cells, and even processes like apoptosis, angiogenesis and cell proliferation (Kulasingam & Diamandis 2008). These can be produced by the tumor mass or by other tissues in response to the presence of the foreign mass or associated conditions such as inflammation. Such biomarkers can be found in body fluids like blood, serum or urine, on tissue and cell lines (Even-Desrumeaux et al. 2011). Tumor markers can be divided in some groups according with their specifications and

applications. For instance, there are diagnostic, prognostic and predictive biomarkers (Kulasingam & Diamandis 2008). A diagnostic marker is a marker that is used to detect and identify a certain type of cancer in any individual. These kind of markers are expected to have high specificity and sensitivity meaning that their presence on a given sample indicates the existence of cancer, or that cancer will occur with nearly a 100% certainty within a specified time interval (Srivastava et al. 2001). Several criteria must be met before biomarkers can be approved as markers for early detection involving the quality of biomarkers and also the assays for their measurement: (a) the biomarker must be expressed in a different manner in normal, premalignant or high-risk, and tumor tissue; (b) the marker and its assay must provide acceptable predictive accuracy for risk or for the presence of cancer; and (c) the variation of the detection tests and the intra- and inter-laboratory variance must be known (Srivastava et al. 2001). Prognostic biomarkers are expected to predict a likely course of the disease since their status is already established including the probability of recurrence. Hence, a prognostic marker can aid the definition of the therapy to be used (Bendardaf et al. 2004). At last, predictive markers are those that can predict the response to a drug treatment before its application. They allow distinguishing individuals as "responders" and "no responders" to a given treatment (Newton et al. 2011). These biomarkers mainly arise from array-type experiments that make it possible to predict the clinical outcome from the molecular characteristics of a patient's tumor (Kulasingam & Diamandis 2008).

An ideal cancer biomarker should be measured easily, reliably and cost-effectively by the use of an assay with high analytical sensitivity and specificity (Even-Desrumeaux et al. 2011). Currently, a central concern on the clinical use of biomarkers is that, generally, they lack diagnostic specificity and sensitivity and so they are used as a complement of imaging, biopsy and associated clinic pathological information before a clinical decision is made (Kulasingam & Diamandis 2008).

2.1.1. STRATEGIES FOR THE DISCOVERY OF NOVEL CANCER BIOMARKERS

The last decade has been marked with an impressive growth in the field of large-scale and high-throughput biology, which contributed to an era of new technological development. The accomplishment of a number of genome-sequencing projects, the discovery of oncogenes and tumor-suppressor genes, and recent advances in genomic and proteomic technologies, jointly with potent bioinformatics tools, are now having a great impact on the cancer biomarkers

research field (Kulasingam & Diamandis 2008). Modern technologies enable performing parallel rather than serial analysis, therefore being able to discriminate different patterns and multiple markers simultaneously. Fig. 3 summarizes a number of new strategies that have been using emergent technologies towards tumor markers research.

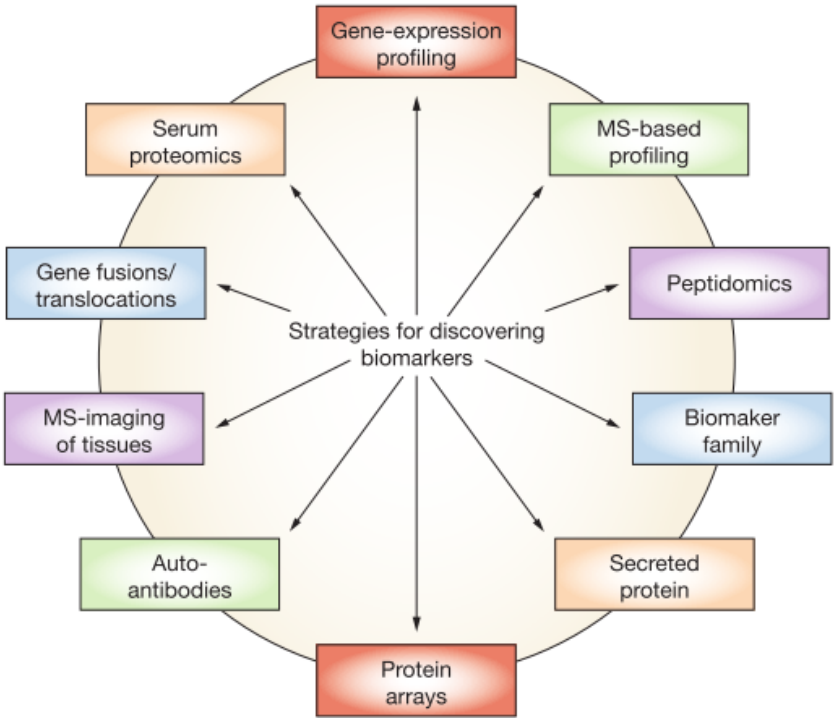

Fig. 3 – Outline of strategies for biomarker discovery through utilization of emerging technologies. Abbreviation: MS, mass spectrometry [taken from: (Kulasingam & Diamandis 2008)].

2.1.2. DNA AND RNA AS CANCER BIOMARKERS – GENE-EXPRESSION PROFILING

On the latest years a new type of tumor classification hypothesis arose. This classification is based on genomic microarrays, a powerful technology for gene-expression studies and it states on the fact that gene-expression patterns identified with DNA microarrays can predict the clinical behavior of tumors (Mohr 2002; Marisa et al. 2013). This method can provide complete gene expression profiles of tumor and normal cells. Comparison of normal and tumor profiles of gene expression (GEPs) can lead to the discovery of different levels of expression of one or more genes, thus indicating a possible DNA or mRNA biomarker (Mohr 2002).

In spite of the positive results of this technology for cancer sub-classification, there are only two multigene panel tests approved by FDA (US Food and Drug Administration), both of them to predict the recurrence of breast cancer (Kulasingam & Diamandis 2008). For CRC,

among all molecular markers that have been extensively studied for its characterization and prognosis, microsatellite instability (MSI), caused by defects on the Mismatch Repair System of DNA (MMR), is the only marker that was found to be a significant prognostic factor in early CRC (Marisa et al. 2013). Microarray technology has been used in the recent years to investigate GEPs in CRC, but they are found to be poorly reproducible, possibly because CRC can be developed through multiple pathways being composed of distinct molecular entities. In fact, specific GEP studies, including genetic and epigenetic analysis, have identified at least three distinct molecular subtypes of colon cancer (Shen et al. 2007), proving again that this should no longer be considered an homogenous disease.

2.1.3. PROTEINS AS CANCER BIOMARKERS - ONCOPROTEOMICS

Gene expression data gives limited information about the biological processes occurring within a cell since proteins are the main functional units performing all of them. Also, it is known that mRNA suffers post-transcriptional events and proteins, post-translational modifications. Therefore, the direct analysis of proteins, using proteomics has several advantages regardless of requiring more tissue and being more time-consuming (Even-Desrumeaux et al. 2011; Wulfkuhle et al. 2003). For this reason, the use of proteomic patterns for cancer diagnosis and tumor sub-classification seems promising (Kulasingam & Diamandis 2008). The concept of oncoproteomics arises as the study of proteins and their interactions in cancer cells using proteomic technologies. Cells and tissue phenotype is ultimately dependent on which proteins are being expressed and how much of it is being produced at a given time. Hence, all alterations that occur at the cell level during carcinoma process can be monitored evaluating cell protein profiles both qualitatively and quantitatively (Cho 2007). Protein signatures in cancer provide helpful details that may enable more effective diagnosis, prognosis, and response to therapy information.

Several proteomics technologies including 2D-PAGE, Mass Spectrometry (MS), surface enhanced laser desorption/ionization time of flight (SELDI-ToF) (Wulfkuhle et al. 2003), protein arrays and mass-spectrometry techniques such as isotope coded affinity tags (ICAT), iTRAQ and multidimensional protein identification technology (MudPIT) are the approaches being implemented in cancer research to a qualitative analysis (Even-Desrumeaux et al. 2011) aiming at the discovery of novel biomarkers that can be measured using enzyme-linked immunosorbent assay (ELISA) or immunohistochemistry (Cho 2007).

ELISA is the most common method used for protein quantification. This system represents the most reliable, sensitive and widely available protein-based methodology for biomedical research and clinical diagnostic (Even-Desrumeaux et al. 2011; Wulfkuhle et al. 2003). ELISA systems can be used for the detection of specific antibodies, soluble antigens adsorbed onto a plastic microtiter plate or cell-surface antigens, which are incubated with reactants covalently coupled to an enzyme. Several washes are done to remove the unbound conjugates and a chromogenic or fluorogenic substrate is added. As the substrate is hydrolyzed by the bound enzyme conjugate, a colored or fluorescent product is generated that can be measured either visually or using a microtiter plate reader (Hornbeck 1991). To test for the presence of disease, these tests requires a scrupulously validated protein biomarker as well as an well-characterized, high-affinity antibody that can detect the protein of interest (Wulfkuhle et al. 2003).

2.2. NON-INVASIVE BIOMARKERS FOR THE EARLY DETECTION OF CRC

Several markers that include faecal, genetic, epigenetic and serum or blood markers available for the detection of CRC are presented on table 1. These are either already in use for screening of general population (a), in clinical trials (b) or in a preclinical state (c).

Faecal hemoglobin has proved its value as a marker for CRC screening as it is the protein measured on FOBT, i.e. the most widely screening modality available. In addition, DNA stool test, named cologuard test, was approved by FDA on 2014 (Simon 2014), for analysis of cancer faecal colonocytes (colonic epithelium cells). These cells are shed into the faecal stream providing revealing material that can be used to detect mutations on genetic markers and epigenetic markers (Loktionov et al. 1998) as *K-ras*, p53 (Losi et al. 1996; Shivapurkar et al. 1997), adenomatous polyposis coli (APC) (Smith et al. 1994), microsatellite instability (MSI) or long-form DNA (L-DNA) (Kim et al. 2008). This DNA panel was found to be more sensitive and specific for the detection of CRC than FOBT (Imperiale et al. 2004).

Of all serum or blood markers shown in table 1, the carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA 19.9) are the only tumor markers that are currently in clinical use. CEA is a high molecular weight glycoprotein that has been used for many years as a biomarker for CRCs and other cancers. This marker is mostly used for monitoring the patient reaction to treatment, such as resection surgery, but does not provide sufficient sensitivity and

reliability for early cancer detection (Bates 2008), being instead a prognostic marker since higher levels are indicative of a more aggressive disease and poorer prognosis (Kim et al. 2008). CA 19.9, the second most investigated gastrointestinal tumor marker, is less sensitive than CEA when used as a prognostic marker for CRC patients. This and other carbohydrate antigens have been extensively studied, but due to their low sensitivity, stage dependency and specificity, they are not useful for detection of CRC (Hundt et al. 2007). Moreover, tissue inhibitor of metalloproteinase type 1 (TIMP-1) total levels in patients with CRC have shown to be significantly higher when compared to the ones of healthy blood donors with very narrow range of plasma TIMP-1 levels. More importantly, it has been proven that it can be detected at early stages of CRC (Holten-Andersen et al. 2004). In spite of the promising results, more studies are needed to validate the use of TIMP-1 as a diagnostic and prognostic biomarker for CRC (Kim et al. 2008).

Table 1 – Non-invasive molecular markers for the detection of CRC [adapted from: (Kim et al. 2008; Tanaka et al. 2010)]

Analyte	Sample	Type	Status
Faecal Hemoglobin	Stool	Protein	(a)
K-ras	Stool	DNA	(b)
APC	Stool	DNA	(b)
L-DNA	Stool	DNA	(b)
p53	Stool	DNA	(b)
CEA	Serum	Protein	(a)
CA 19.9	Serum	Carbohydrate	(a)
TIMP-1	Serum	Protein	(b)
Spondin-2, DcR3, Trail-R2, Reg IV, MIC 1	Serum	Protein	(c)
PSME3	Serum	Protein	(c)
NNMT	Serum	Protein	(c)
CRMP-2	Serum	Protein	(c)
SELDI (apolipoprotein C1, C3a-desArg, alpha1-antitrypsin, transferring)	Serum	Protein	(c)
HNP 1-3	Serum	Protein	(c)
MIF	Serum	Protein	(c)
M-CSF	Serum	Protein	(c)
M2-PK	Serum	Protein	(c)
Prolactin	Serum	Protein	(c)
Septin 9	Plasma	DNA	(c)
Five-gene panel (CDA,BANK1,BCNP1,MS4A1, MGC20553)	Blood	DNA	(c)
CCSA-2,-3,-4	Serum	Protein	(c)
MMP-7,-9	Serum	Protein	(c)
Laminin	Serum	Protein	(c)

There are evidences that all of the markers mentioned and that are currently in pre-clinical development can be used alone, or in combination with others, for the detection of CRC, since they show a greater level of sensitivity and specificity when compared to CEA. However, large-scale studies are needed to evaluate the potential of using biomarkers that have recently been discovered through genomics and proteomics advances in routine analysis (Tanaka et al. 2010).

2.3. MOLECULAR RECOGNIZING ELEMENTS

The progress of biosensors for the detection of cancer is dependent on the availability of high affinity and specific ligands for the desired cell type and/or biomarker (Mcguire et al. 2009). The ability to recognize a cell or biomarker in a mixed population is viewed as the critical step in any diagnostic assays. The biomarker that can be as complex as a whole cell, or as simple as a single molecule, must be recognized and collected from a heterogeneous population, regardless of the complexity of the sample matrix (Soper et al. 2006).

The most commonly used recognition elements in biosensors are antibodies but, more recently, synthetic based recognition elements such as aptamers, peptides, surface-imprinted polymers, carbohydrates, nucleic acids and molecularly imprinted polymers are being used as a replacement for antibodies (Soper et al. 2006; Petrenko 2008).

For a cell-based recognition, typically very little is known about the specific landscape of the membrane surface making it impossible to design specific cell-targeting ligands, such as specific antibodies. The development of methods to identify specific ligands for tumor biomarkers that can discriminate between normal and cancerous cells is an effective way of cancer cell targeting that can be useful for diagnostic purposes and targeted drug delivery (Zhang et al. 2007), as it has been shown that endothelial cells in the vasculature of tumors differ from normal endothelial cells, and that tumor blood vessels express proteins that are not produced in motionless vascular endothelium (Rasmussen et al. 2002).

The biopanning approach using phage-displayed ligands consists in an unbiased screen technique in which there is no selective pressure towards binding a particular macromolecule (Soper et al. 2006). Phage display screening of random peptide libraries do not require a knowledge of the cell surface nature, and as a result many protocols to identify cell-specific binding peptides have been developed (Barry et al. 1996; Brown 2000; Oyama et al. 2003; McGuire et al. 2004). Peptides selected using biopanning are extremely cell-specific meaning that affinity selection using random peptide libraries can be used to identify different types of tumors that are currently similarly classified (Soper et al. 2006; Oyama et al. 2003). This high discriminating power of the selected phage also suggests that peptides could be used to target cancer cells *in vivo*, since they can distinguish between normal and cancer cells, being an extremely remarkable characteristic that is useful for target chemotherapy and diagnosis purposes. Also, these peptides can be used for example to drive the delivery of fluorescent nanoparticles, as well as cell capture reagents for cell enrichment, and as antibody replacements

for flow cytometry. Another appealing aspect of peptide libraries to screen cancer cells surface is that the peptide found can be characterized as modular, meaning that they retain functionality more or less independently of the protein context. A phage display-derived peptide is, by definition, a peptide that binds to a given target in the context of being coupled to a coat protein of a phage (Voss et al. 2002). A modular peptide is one that was found to be specific to a target in that context, but retains its binding affinity when fused to an heterologous proteins such as glutathione S-transferase, agarose (Frangioni et al. 2000) or a contrast agent. In other words, a specific peptide found using combinatorial phage peptide libraries, can be coupled to a variety of different functional entities. This characteristic has already been demonstrated for target chemotherapy (Pasqualini & Ruoslahti 1996; Arap et al. 1998a; Arap et al. 1998b). Finally, as peptides are amenable to derivatization, it is possible to anticipate that these cell-specific ligands will find utility in a variety of different biosensor platforms (Mcguire et al. 2009).

Synthetic based recognition elements, such as aptamers and peptides, have particular advantages over antibodies for its use as ligands. Aptamers and peptides have robust structures that can be placed in varied conditions without losing their specificity and are easy to modify structurally to support the addition of reporters or immobilization to sensing elements. In addition, their inexpensive chemical synthesis makes their production cost-effective and appropriated for large scale preparation contrarily to the labor-intensive and time-consuming production of antibodies (Soper et al. 2006). In particular, peptides present some advantages especially over antibodies and proteins, namely regarding their use in a clinical context. The small size of peptides makes them very interesting for drug delivery purposes since the degree of penetration of a molecule into a living cell is markedly associated with the size of the compound used for targeting. Moreover, peptides are non-immunogenic, thus promoting their safety profile and making them suitable for clinical uses (Bakhshinejad et al. 2014; Zhang et al. 2007).

Peptides with high affinity for specific tumor antigens like HER_2/neu receptor, oligonucleotide receptor and ICAM- 1 have been isolated by phage display (Bélizaire et al. 2003; Karasseva et al. 2002; Schatzlein et al. 2001). Specifically in the case of CRC, peptide libraries have been used successfully for the identification of colon tumor binding peptides (Zhang et al. 2007; Kelly & Jones 2003; Rasmussen et al. 2002).

3. SPECIFIC CANCER CELLS LIGANDS IDENTIFICATION BY PHAGE DISPLAY TECHNOLOGY

Within the last three decades, biopanning of phage displayed peptide libraries on intact cells proven to be a successful route for the identification of cell-specific ligands (Mcguire et al. 2009). Phage display is a powerful *in vitro* selection technology that uses bacteriophages (phages) to the display of exogenous (poly)peptides on its surfaces coupled to the phage coat proteins (Huang et al. 2011).

The concept was first introduced in 1985 by George Smith (Smith 1985), who used filamentous f1 phage as an expression vector to insert exogenous DNA sequences into specific sites of its genome where the genes encoding the phage coat proteins are localized, enabling the display of peptides and proteins as a fusion to one of the phage coat proteins. Smith proved that it was possible to introduce such alterations on the phage genome, and consequently on the proteins of the coat, without loss of infectivity capacity (Smith 1985; Smith & Petrenko 1997).

Phage display is a combinatorial technology that has developed enormously in the last decades with the construction of a large number of libraries of peptides, proteins, antibodies, cDNA and mRNA (Bratkovič 2009). It had a major influence on several discoveries in fields like drug discovery (Alirezapour et al. 2013; Christensen et al. 2001), tumor cell-targeting (Rasmussen et al. 2002), molecular imaging and cancer diagnosis (Deutscher 2011), and the identification of several cancer cells specific peptides useful for diagnosis and targeted drug delivery (Oyama et al. 2003; Kelly & Jones 2003; Zhang et al. 2007; Ghosh et al. 2012; Bakhshinejad et al. 2014).

This technology takes advantage of the extensive genetic flexibility of phages to create a huge variety of modifications on their surface, thus allowing the construction of combinatorial libraries composed by phage particles displaying ligands (peptides, proteins, among others) for different targets eliminating the need of genetically or chemically engineer each one of them (Arap 2005). Using standard molecular biology techniques, a phage display library, containing up to 10^{10} variants can be constructed simultaneously on the surface of phages, and can be used for the selection of any desired target even when the target and its specifications are unknown (Huang et al. 2011). This is the great strength of phage technology; i.e. the ability to identify interactive regions of proteins and other molecules without preexisting notions about the nature of the interaction. For this reason, different screening methods allowed the isolation and

characterization of peptides binding to several molecules *in vitro*, in the context of living cells, on both animals and humans (Arap 2005).

3.1. FILAMENTOUS PHAGES FOR PHAGE DISPLAY: PHAGE STRUCTURE AND PROPAGATION

A variety of phages, obligate intracellular parasites that use the biosynthetic machinery of the bacterial cell host (gram-negative cells) for reproduction, have been used for phage display purposes. M13, f1 and fd phages from Ff family (usually referred as Ff or M13-like phages) are mostly used for phage display technology (Arap 2005; Russel et al. 2004). Filamentous phages are ideal to use as expression vectors and for display in particular. On phage display expression systems (e.g. on Fig. 4) the foreign DNA sequence is introduced into the gene of one of the phage coat proteins so that the respective amino acid sequence is displayed coupled to the endogenous amino acids of the phage coat protein. The hybrid coat protein is incorporated into the viral particles as they are released from the cell host, yielding a recombinant phage with the foreign peptide or protein domain displayed on the outer surface. (Smith & Petrenko 1997).

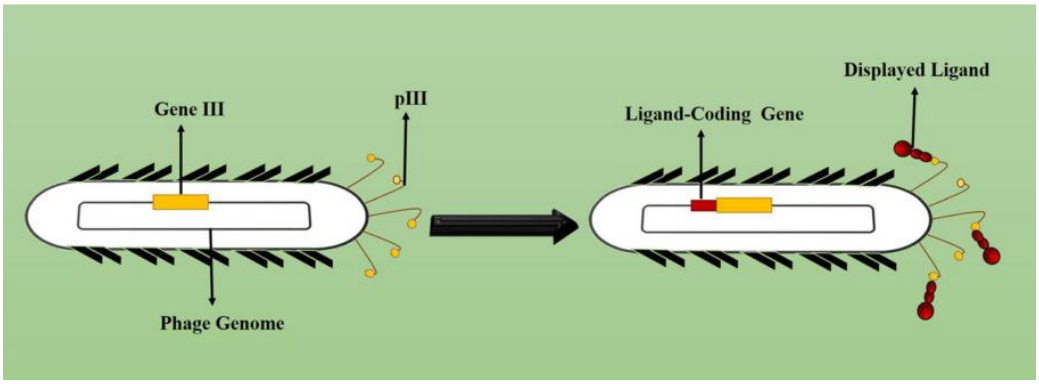

Fig. 4 – Phage Display expression system (M13 phage): the ligand-encoding oligonucleotide inserted next to gene III that encodes a coat protein of the phage generates a recombinant phage, presented on the right, with the ligand displayed coupled to the respective coat protein [taken from: (Bakhshinejad et al. 2014)].

Ff phages defining characteristic is a circular single-stranded DNA genome enclosed in a flexible tube composed of thousands of copies of a single major coat protein and few minor proteins at the tips. An important feature for phage display is that genomic DNA of these phages can be obtained both as single-stranded and as double-stranded forms which facilitates cloning and library construction as it can be used as a plasmid-based vector. Additionally, phage's small

genome can tolerate large insertions on non-essential regions with no effect on infection capacity as the enlargement of the genome size is accompanied by an equal increase in the length of the phage particle, as well as in the number of pVIII (major coat protein) copies (Bakhshinejad et al. 2014; Ebrahimizadeh & Rajabibazl 2014).

M13 is an *E. coli*-specific filamentous bacteriophage and is the most commonly used for phage display purposes. It has a flexible tube shaped structure and a circular genome with 6-8 kb enclosed in a coat composed of five different proteins (Sidhu 2001) (Fig. 5).

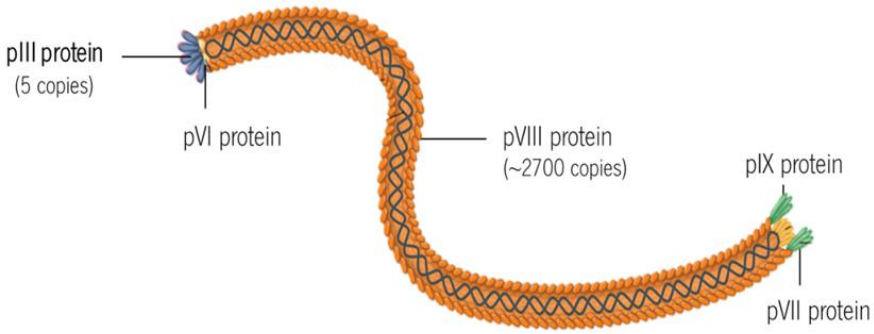

Fig. 5 – The general structure of M13 phage particle.
[adapted from: http://www.wwnorton.com/college/biology/microbiology2/ch/11/etopics.aspx, on December 17th, 2014]

The M13 genome contains 11 genes; five of which encode the coat proteins described below, while the others encode proteins necessary for viral replication and assembly. Several thousand copies of pVIII, the major coat protein, cover the length of the phage with five surface-exposed N-terminal residues. On one side of the phage particle, where phage assembly begins, there are 3-5 copies of proteins pVII and pIX whose absence will make impossible to form a phage particle. On the other end there are about 5 copies of both pIII and pVI that are needed for the phage to detach from the host cell membrane (Russel et al., 2004).

All of the five coat proteins of the phage have been used to the display of foreign peptides and proteins, but pIII and pVIII are the most important and each of them has different features (Bakhshinejad et al. 2014). Protein pIII is more suitable for display of a smaller number of larger proteins, while pVIII is appropriate for the display of more copies of the same small peptide (Willats 2002). PIII libraries display 3-5 copies of each individual peptide, whereas pVIII libraries can display up to 2700 copies of small peptides (Kay et al. 1996). However, this cannot be used as a general assumption for phage display. In fact, choosing the coat protein for displaying

depends on the type of ligand and the specificity needed for the screening of the target (Bakhshinejad et al. 2014).

Fd particles are able to infect a variety of gram-negative bacteria. F' and F+ strains (which have an F pilus) of *E. coli* are widely used to obtain virus particles for phage display. Filamentous phage infection does not produce lytic infection in *E. coli*, contrariwise induces a state in which the infected bacteria produces and releases phage particles into the growing medium, without killing the cell (Russel et al. 2004). Possessing a non-lytic cycle propagation mechanism is an important feature for the use of M13-like phages as expression systems as it enables the accumulation of a large number of virus particles in the host cell and growth medium (Bakhshinejad et al. 2014). Fig. 6 represents the infection life cycle of filamentous phage that begins with the infection of bacterial cells when the N-terminus of pIII attaches to the F pilus of male *E. coli* (Sidhu 2001). After this, the single stranded phage genome enters the bacteria and is converted to the double stranded replicative form (RF) by the host DNA replication machinery. RF genome is the template for the production of viral coat proteins and single stranded DNA progeny. The viral coat proteins encapsulate the single stranded DNA during its extrusion from the host cell (Pande et al. 2010).

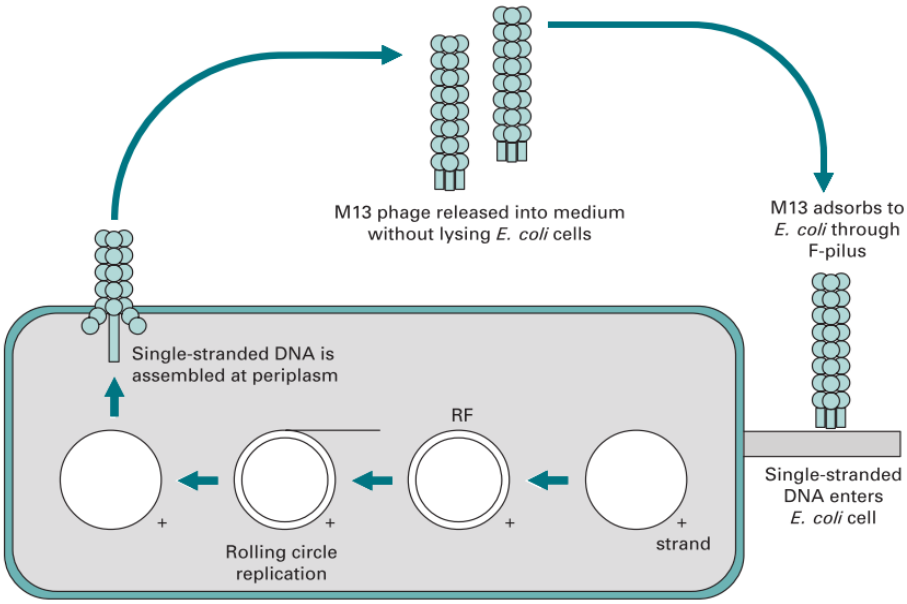

Fig. 6 – M13 particles life cycle [taken from: (Wilson & Walker 2010)].

The existence of F pilus is very important for phage infection and propagation as, even though infection may occur without appropriate pili, the process is extremely inefficient. There are only a few F pilus per each bacterial cell and the target size at their ends is small. For that reason, the rate and efficiency of infection of a bacterial culture is improved at high multiplicity

phage per cell and high cell density. Nevertheless, high density of bacterial cells has to be achieved during log phase because after this point (at stationary phase) pilus expression is decreased and infectivity is compromised. Therefore, to accomplish a higher titer of virus, it is important to have a well aerated bacterial culture so a high cell density is achieved immediately after entering the log phase. Also, F pilus do not assemble at low temperatures, thus temperatures equal or above 34°C are needed to assure plaque formation (Russel et al. 2004).

Under optimal conditions, the first progeny particles appear in the culture supernatant 10 min after infection. Phage number increases exponentially on the first 40 min after the infection, after which the rate of propagation becomes linear. In one generation of bacteria, approximately 1000 phage particles are produced and liberated to the surrounding medium (Wilson & Walker 2010). Under optimal conditions infected cells can continue to grow and divide, as well as producing phages, indefinitely (Russel et al. 2004).

3.2. PHAGE DISPLAY RANDOM PEPTIDE LIBRARIES

Random peptide libraries are the most common type of combinatorial phage libraries (Pande et al. 2010), which provide a rich source of molecules that can be chemically or genetically synthesized and screened to obtain a desired function or affinity. The first combinatorial library was developed five years after the introduction of the phage display concept (Huang et al. 2011). It was constructed by inserting random oligonucleotide sequences into the gene encoding one of the coat proteins of M13. In a phage library, each phage particle displays a unique ligand, but the entire library encompasses a great variety of billions of peptides. The genetic encoding of a library, meaning the presence of the DNA and the peptide of interest in the same virion makes possible to re-synthesize and re-screen binding ligands of interest. Combinatorial libraries turned phage display into a high throughput system that can be used to screen a large pool of potential binding motifs and select those with highest affinity to the target structure (Bakhshinejad et al. 2014). Also, given that a single event of phage/target binding can be detected, this technology can be used to discover binding peptides when the target is ranging from a purified protein sample to a complex mix of proteins, as in the case of a living cell surface (Voss et al. 2002).

For the construction of pIII random peptide libraries, the random oligonucleotides are inserted between the coding sequence for the signal peptide and the N-terminus of the coat

protein pIII (Fig. 7). The DNA inserts are derived from "degenerate" oligonucleotides, which are chemically synthesized by adding mixtures of nucleotides (in spite of single nucleotides) to a growing nucleotide chain (Smith & Petrenko 1997).

Fig. 7 – Schematic diagram of how foreign peptide domains are fused to coat proteins pIII in phage-display vectors [adapted from: (Smith & Petrenko 1997)].

In general, the length of the displayed linear peptides varies from 6 up to 45 amino acids (Bakhshinejad et al. 2014) having been successfully expressed as N-terminal pIII fusions (McConnell et al. 1996; Burritt et al. 1995). Since then, phage-displayed combinatorial peptides have been the source of peptides that selectively bind to cell and tissue surfaces, cytosolic proteins, receptors, inert materials, metals, and toxins (Huang et al. 2011)

Beyond linear random peptide libraries, there are also phage display libraries with a loop scaffold that have been developed by flanking the random peptide sequences by a pair of cysteine residues that form a cross-bridge (C7C). These structurally constrained libraries have been successfully used for the selection of targets that cannot be selected using a linear random peptide library (Lane & Stephen 1993; Clackson & Wells 1994; O'Neil et al. 1992). Such libraries are particularly useful for targets whose native ligands are in the context of a surface loop, such as antibodies with structural epitopes.

The detailed procedure for the synthesis and amplification of the libraries will be described on CHAPTER 3, point 1.3.

3.3. SELECTION OF LIGAND-RECEPTORS IN COMPLEX BIOLOGICAL SYSTEMS - BIOPANNING

In general, biopanning involving phage display random peptide libraries involves five fundamental steps being the first the preparation of the library or amplification of an existing one (point 3.2). After that, a biopanning round involves multiple steps: target immobilization, phage

binding [1], washing [2], phage elution [3] and finally *E. coli*-infection for phage amplification [4, 5] and one round is usually repeated three to five times [6] (Fig. 8) (Huang et al. 2011).

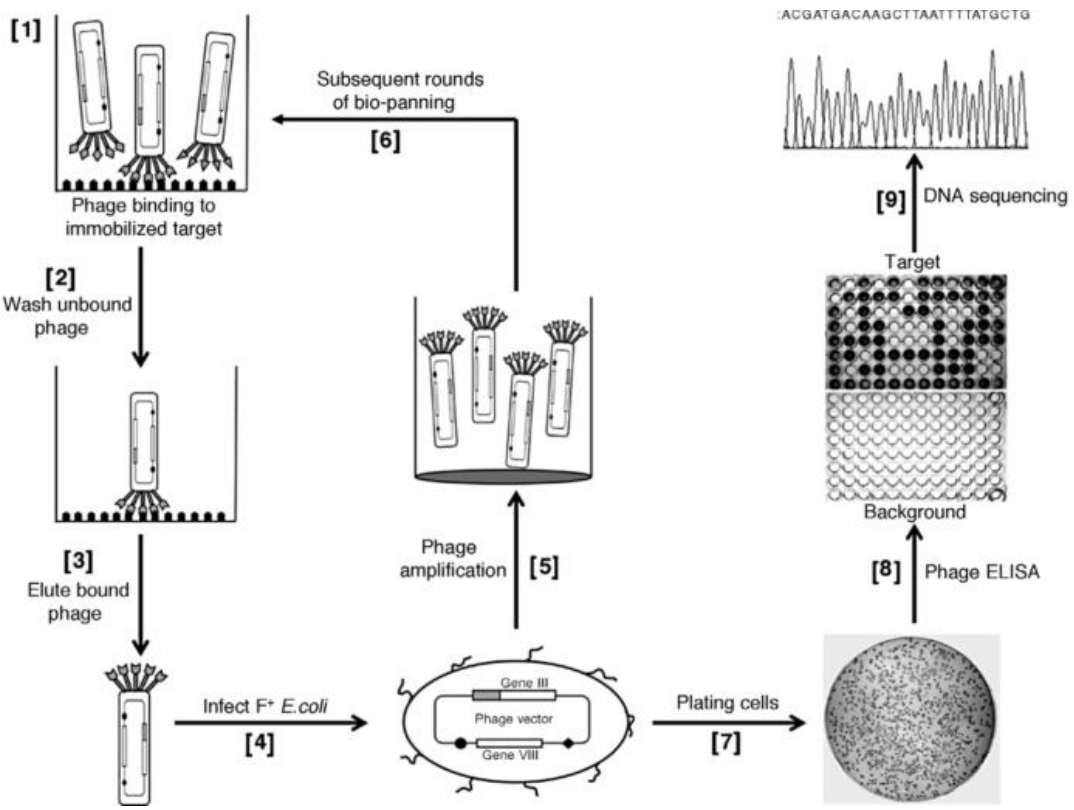

Fig. 8 – Schematic representation of the screening of a phage peptide library through biopanning and cloning analysis [taken from: (Huang et al. 2011)].

In the affinity screening methodology presented on Fig. 8, a well plate is covered with the target that can be a purified protein, adherent cultured cells, tumor tissues (obtained through biopsy), among others (Bakhshinejad et al. 2014). When using adherent cells on a well plate, the target is already immobilized (Fig. 8 [1]), but there is a need to wash the cells to reduce unspecific binding to common molecules like bovine serum albumin (BSA) (commonly used as a complement to the growth medium (Arap 2005)) and also dead cells. To discover peptide ligands against a given target, in this case a cell, it is important to verify that the morphology of the cell is normal.

Then, the phage display random peptide library is added to the target coated well in a solution that allows stability of the target and minimal non-specific binding of the phage (Fig. 8 [1]) that can be growth medium without BSA, for example. For the first round of selection it is very important to have a highly diverse library to enhance the possibilities of finding a peptide of

interest (Pande et al. 2010). A restrict pool of peptides will decrease immensely the positive results of phage display affinity selection. After this it is necessary to wash away all the phage displaying peptides that did not bind to the target (Fig. 8 [2]). The first round of biopanning requires higher yield of the fittest phage clones over the background and for this reason the washes need to be less stringent (Smith & Petrenko 1997). The stringency of this washing step needs to be increased over the rounds, with more extensive washes to isolate the phages with higher affinity (Pande et al. 2010). After this washing step, only the bounded phages are in the well plates and need to be eluted (Fig. 8 [3]) for further amplification. Specific elution of the phage (without the target) can be carried out in a solution containing either free target or a competing ligand (Pande et al. 2010). Phage particles withstand very harsh conditions, such as low pH and low temperatures, without losing bacterial infectivity thus, protocols using low pH and high concentration urea have been used to dissociate bound phage from a target (Arap 2005). Non-specific elution of the target bound phage can also be performed using extreme pH, denaturants, ionic strength, limited proteolysis or sonication (Pande et al. 2010).

In addition, bound phages may not need to be eluted from the well or animal tissue before bacterial infection, since infection can proceed after addition of bacteria directly into the well or to the homogenized cell tissue (detached from the well using trypsin) (Arap 2005).

The most popular and versatile technique for initial characterization of the selected phage clones properties is ELISA (Fig. 8 [8]). In this method, the target of interest is immobilized, at high density, on a 96-microwell plate and incubated with individual clones or the entire phage complexes obtained from the biopanning rounds. The specific binding is then detected using an anti-M13 antibody (Russel et al. 2004). Since the amount of target coated on the plate is not quantifiable, and it is present at sufficiently high density to allow multivalent binding to the phage, this method will not determine whether the selected phage binds with high or low affinity. The method is useful for qualitative determination of relative binding affinities for a number of selected clones, or phage complexes, in parallel, and will distinguish true target binding from binding to the plastic support i.e. background (Russel et al. 2004; Adey et al. 1995). Following the analysis of the phages relative affinity towards the target using ELISA, the phage clones can be analyzed by DNA sequencing (Fig. 8 [9]) to identify the sequence of the peptide that interacts with the target (Paschke 2006).

In 2001, Giordano and colleagues presented a new approach for screening and selection of cell-surface binding peptides. The method named BRASIL – biopanning and rapid analysis of

selective interactive ligands, is based on differential centrifugations in which a cell suspension previously incubated with phages displaying the pool of peptides for four hours at 4 °C, is placed in an aqueous upper phase and is then centrifuged through a non-miscible organic lower phage (Giordano et al. 2001). Centrifugation drives the cells and consequently the phage that bounded to them from a hydrophilic phase into the organic phase. The passage of cells from a hydrophilic to a hydrophobic setting will separate water-soluble materials, such as the unbound phage that stays on the hydrophilic phase (Fig. 9). The cell/phage pellet can then be recovered from the bottom of the tube and can be used directly for bacterial infection to recover the phages that will be used on the subsequent round and finally for ELISA binding assays and DNA sequencing as on the conventional affinity selection protocol.

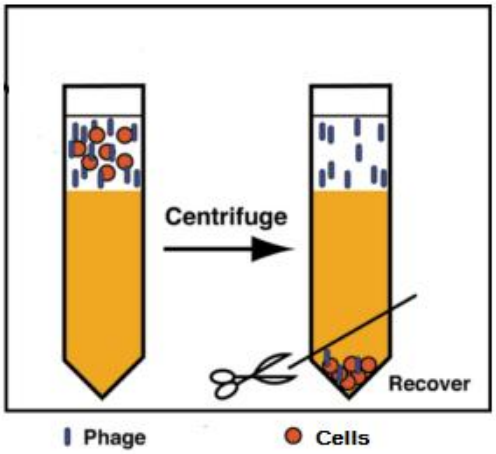

Fig. 9 – Schematic representation of a BRASIL selection round [adapted from: (Giordano et al. 2001)].

On their report, Giordano and colleagues used RGD-4C phage displaying a known specific peptide for α_v integrins to incubate with α_v integrin-expressing Kaposi's sarcoma cells (KS1767) using, in parallel, a control insertless phage, named fd-tet, which was also incubated with KS1767. Using both biopanning methodologies, conventional and BRASIL, they proved that the recovery of specific phages was higher with BRASIL, which, at the same time, allowed a lower recovery of unspecific phages (control phages). The authors concluded that the single-step organic phase separation is faster, more sensitive and more specific than current methods that rely on washing steps or limiting dilution (Giordano et al. 2001). An extensive explanation of these protocols is presented on CHAPTER 3, point 1.4.

4. REPORTER PHAGES

Bacteriophages are an ideal tool for bacterial detection as they are extremely host sensitive being actively used as probes for pathogen detection (Singh et al. 2013). Besides using unaltered phage particles, molecular cloning techniques enabled the engineering of phages to carry specific reporter genes. When expressed, after the infection of the bacterial cells, and usually in the presence of specific substrates they produce a compound that can be detected by measuring bioluminescence, fluorescence, or enzymatic conversion of a chromogenic substrate. Many genes such as *luxAB, luxI*, *luc*, *lacZ* (see CHAPTER 3, point 1.2.), and *gfp*, among others, have been used to genetically engineer phages to confer them the ability to detect pathogens in a fast and sensitive manner (Schmelcher & Loessner 2014).

To construct such phages, the reporter gene can be introduced by direct cloning, transposition and homologous recombination. Direct cloning, used for the construction of phage libraries, is only feasible for a limited number of phages (Ulitzur & Kuhn 1987). Although being extremely useful, there are certain difficulties to construct such phages, since it requires labor-intensive and detailed genetic knowledge of the phage. Besides, as it was already mentioned, the volume of the phage capsid sets a natural limit to the amount of genetic material that can be introduced into the phage genome (Schmelcher & Loessner 2014).

Among all report phages described to date, luciferase reporter phages (LRP), which carry a luciferase gene for bacterial detection, are the most used. Major advantages of using luciferase genes as reporters are the highly sensitive detection of the bioluminescent signal they generate and the fact that they ensure low background noises (Brovko et al. 2012). The main application of these bacteriophages is the detection and control of bacterial pathogens in food and food-processing environment, but reporter genes have also already been used for molecular imaging in cancer research (Ray et al. 2003). In this specific field, optical reporter genes, like luciferase-encoding genes are probably the most used (Liu et al. 2012).

Luciferases are a class of oxidative enzymes that are capable to generate bioluminescent light through a chemical reaction with oxygen and a substrate. The most commonly exploited luciferases in optical imaging are firefly and renilla luciferases. They catalyze a reaction using different substrates and emit light at different regions of the spectra (Shao et al. 2012). The bioluminescence released by both reactions can be measured using a luminometer (Hazbo et al. 2003). Bioluminescence imaging arises as a modality that has advantages when compared for e.g. to fluorescent imaging, since contrarily to the last it does not require external excitation light.

In addition, as previously mentioned, luciferase-based bioluminescence displays low background and high sensitivity since mammalian cells and tissues do not produce significant bioluminescence during normal cellular processes (Shao et al. 2012).

In this thesis, efforts are conducted to merge both fields by constructing a reporter phage using luciferase gene to detect the presence of cancer cells on a given sample. For that purpose, NanoLuc™ gene will be inserted on the M13KE genome by direct cloning (CHAPTER 3, section 2).

1. PHAGE DISPLAY AND *IN VITRO* BIOPANNING

Phage display and biopanning experiments involve the manipulation of both mammalian and bacterial cells, as well as phage particles. The procedures and culture techniques used during this work to perform phage display and affinity selection, including the development of combinatorial random peptide libraries are described on the next sections.

1.1. COLON CANCER CELLS

The cell line used for the selection of phage displayed peptides and for the proteomic study was RKO human epithelial colon carcinoma [ATCC® CRL2577] at cell culture passages between 10 and 30 – Fig. 10.

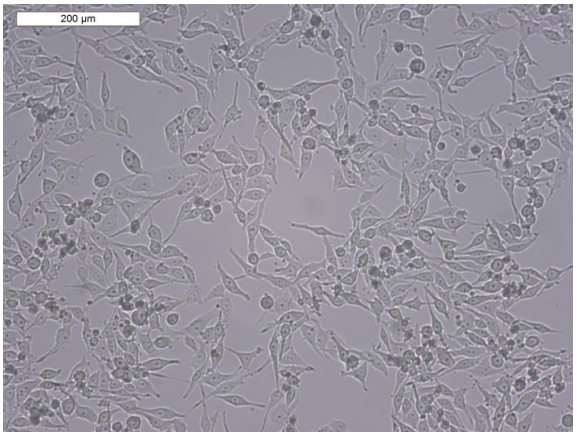

Fig. 10 – RKO cell line, passage 13 at ≈ 60% confluence (Scale bar 200 µm).

1.1.1. CELL CULTURE INITIATION AND SUBCULTURE METHODOLOGIES

Cell cultures were obtained cryopreserved in a medium composed by 5% dimethyl sulfoxide - DMSO [Sigma-Aldrich Co. LLC] and Fetal Bovine Serum (FBS). Cell culture was initiated according to ATCC protocol from ATCC (American Type Culture Collection) Animal Cell Culture Guide (ATCC 2012).

The cells were cultured in complete growth medium composed of Dulbecco's Modified Eagle Medium (DMEM) [Biochrom] supplemented with 10% (v/v) FBS [Biochrom] and 1% (v/v) of penicillin-streptomycin [Biochrom] on tissue culture treated flasks and well plates [SPL Life Sciences Co., Ltd.]. Cells were maintained at 37 °C with 5% CO_2 and humid atmosphere on a Hera Cell incubator. Subculturing was performed when the culture achieved 80-90% of confluence, with a split ratio dependent on the growth area of the flask or well to subculture (normally, subculturing from a T25 flask to another with a 1:2 or 1:3 ratio). For subculturing, the cell monolayer was washed with sterile phosphate buffered-saline (PBS), pH 7.4 (appendix 1), without Ca^{2+} and Mg^{2+} and was then covered with Trypsin/EDTA (ethylenediaminetetraacetic acid) solution 0.05%/0.2% (w/v) in PBS without Ca^{2+}/Mg^{2+} [Biochrom, L2143] to detach the cells. Trypsin is serine protease that cleavages peptide chains mainly at the carboxyl side of lysine or arginine (Roth® 2009). The enzyme was left to act at 37 °C for 5-7 min and no longer since the contact of cells with trypsin for a longer period can damage their membranes and ultimately killed them. Using inverted contrasting microscope [LEICA DM IL], cell detachment was confirmed when cells exhibit movement and a spherical shape. To inactivate the trypsin effect, an equal volume of fresh complete growth medium was added to the culture flask. The volumes of growth medium and trypsin used for the different vessels, and also the average number of cells at 80-90% confluency measured several times during the experimental work are presented on table 2.

Table 2 – Volumes of complete growth medium and Trypsin/EDTA solution; number of cells at 80% - 90% confluency for the correspondent culture vessels used to RKO cells culture

Culture vessel	Complete growth medium (mL)	Trypsin/EDTA solution (mL)	Number of cells at confluency (average)
96-well plate	0.2	0.05	0.05×10^6
12-well plate	1.0	0.50	1.2×10^6
6-well plate	3.0	1.0	2.4×10^6
T-25	4.0	2.0	5.0×10^6
T-75	8.0	5.0	12.0×10^6
T-150	15.0	10.0	20.5×10^6

Following trypsin inactivation, the cell suspension was recovered to a 15 mL falcon tube and then centrifuged at 250 x g for 7-10 min to make a solid cell pellet. The cell pellet was washed using complete growth medium in order to remove all the trypsin that could damage the

cells. After washing and centrifuging, the cells were ressuspended again on fresh complete medium and, after counted (CHAPTER 3, point 1.1.2) they were split according to the experimental needs.

When, after two or three days, the culture was not ready to subculture (e.g. when using a greater split ratio) the complete medium was renewed to replenish nutrients and maintain a correct pH level.

1.1.2. CELL COUNTING AND CELL VIABILITY ASSESSMENT

Counting the cells was a necessary procedure to evaluate the cell concentration before the subculturing and biopanning experiments (CHAPTER 3, point 1.4). Cells' counting was performed using a Neubauer chamber. Cell viability was evaluated using the trypan blue exclusion test. This test is based on the fact that, contrarily to dead cells, living viable cells have intact cell membranes that are impermeable to certain dyes such as trypan blue, eosin or propidium. Thus, mixing one of those dyes with a suspension of cells it is possible to visualize which cells absorb or exclude the dye (Strober 2001). After detaching the cell monolayer using trypsin/EDTA solution as described above, the cells were ressuspended on a known volume of complete growth medium. A dilution of the cell suspension was prepared using 0.4% trypan blue solution [Sigma-Aldrich Co. LLC, Portugal] and about 2-3 min after mixing, 10 μL of the above mentioned cell-trypan blue solution was transferred to the Neubauer chamber.

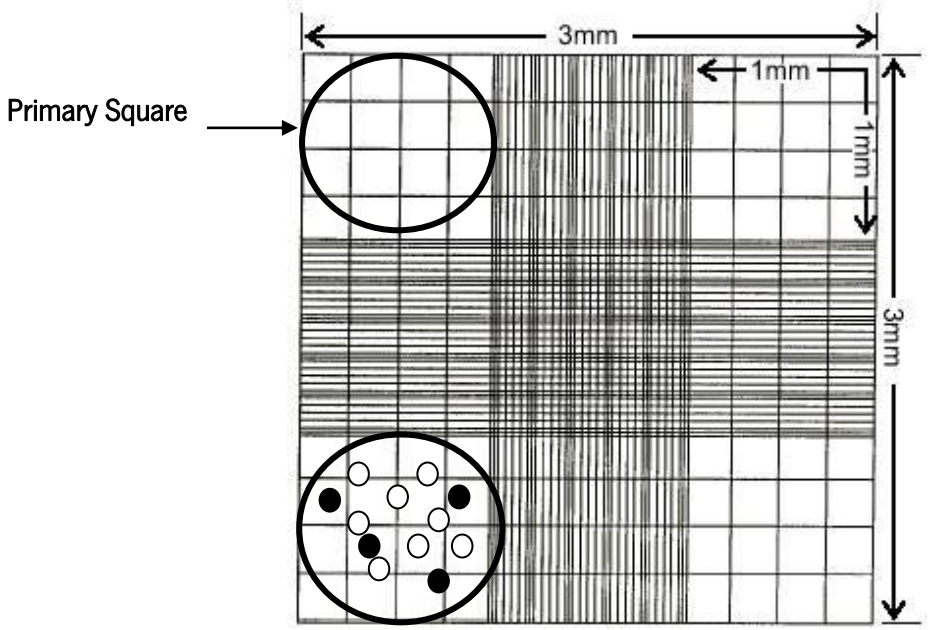

Fig. 11 – Trypan blue exclusion test – the cells that present dark coloration are excluded i.e. not counted.

Each primary square of the Neubauer chamber (with cover slip in place) represents a total volume of 0.1 mm^3 (Fig. 11). Since 1 cm^3 is equivalent to 1 mL, the subsequent cell concentration per mL (and the total number of cells) was determined using the following equations:

$$\% \text{ Cell Viability} = \frac{Total\ of\ viable\ cells\ (unstained)}{Total\ cells} \times 100 \quad (1)$$

$$\text{Viable Cells /mL} = \frac{number\ of\ cells\ counted}{number\ of\ primary\ squares\ counted} \times dilution\ factor \times 10^4 \quad (2)$$

$$\text{Number of viable cells on a sample} = number\ of\ viable\ cells \times sample\ volume \quad (3)$$

$$\text{Dilution factor} = \frac{sample\ volume}{total\ volume\ (sample\ volume + dilution\ volume)} \quad (4)$$

1.2. M13KE PHAGE

M13KE phage [N0316S] purchased from New England Biolabs® was used in this work. M13KE is a phage system derived from M13mp19 in which cloning sites have been introduced at the 5' end of gene III to facilitate the insertion of short peptide codifying sequences as N-terminal pIII fusions. As it is a phage and not a phagemid vector, all the pIII copies would present the cloned peptide. According to the manufacture datasheet, displaying of peptides longer than 20-30 amino acids would have deleterious effects on the phage infection capacity. Furthermore,

M13KE do not present a plasmid replicon or antibiotic resistance, thus it has to be propagated as a phage (by *E. coli* infection) which also means that the titer is determined by counting plaques and not colonies. This phage is specially designed for the display of small peptides on the virus capsid, and consequently to be used as a platform for random peptide libraries for affinity selection screenings. Also, M13KE presents a *LacZα* gene fragment that codifies a peptide containing the amino terminus of β-galactosidase. This alpha fragment lacks enzymatic activity but can associate with ω fragments to form proteins whose β-galactosidase activity has been restored (Ausubel et al. 2003). *E. coli* strain ER2738 is a *LacZα* complementing strain since its plasmid carries a complementing ω-fragment gene that allows the assembly of an active complex.

X-Gal (5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside) is a chromogenic substrate that is broken down by β-galactosidase in the presence of IPTG (Isopropyl-β-D-thiogalactopyranoside), a highly stable synthetic analog of lactose which inactivates lac repressor and thus de-represses the ω peptide synthesis (Ausubel et al. 2003). The hydrolysis of X-Gal produces blue precipitates that allow a straightforward visual differentiation of *E. coli* colonies that were not infected or effectively transformed (see points 1.2.2 and 1.3.9) by M13KE phage or vector, respectively.

1.2.1. BACTERIAL CELL CULTURE

For phage propagation, male *E. coli* strain ER2738, designated as ER2738 from this point forward, an F' strain with resistance to the antibiotic tetracycline from New England Biolabs® [E4104S], was used. For daily uses, ER2738 was cultured in 3-5 mL of Luria-Bertani (LB) broth [25 g/L LB broth, NzyTech, Lda] with a final concentration of 5 μg/mL of tetracycline [Sigma-Aldrich Co. LLC, 87128] on 15 mL Falcon type tubes [Labbox] for 5-6 hours, at 37 °C with an agitation of 200 rpm.

ER2738 was also maintained and weekly renewed, in solid culture LB agar plates (25 g/L of LB broth with 20 g/L of agar from Liofilchem®) with tetracycline at a final concentration of 5 μg/mL. The solid culture, renewed monthly from a stock storage at -80 °C, was used to transfer ER2738 to liquid culture on the start of every work week.

1.2.2. PHAGE PRODUCTION AND ISOLATION

Phages are easily propagated by bacterial infection (CHAPTER 2: point 3.1). To maintain and produce phages during this experimental work both solid and liquid bacterial cell cultures were used. It is possible to propagate phages from liquid suspension to solid plates and from solid plates to liquid by picking individual plaques or recovering all plated phages using a suspension buffer.

Initially, a phage suspension of M13KE purchased from New England Biolabs® was used for propagation on liquid or solid medium.

For phage production and isolation on solid medium the Double-Agar Layer (DLA) technique was used. This technique was formalized by Mark Adams (Adams 1959) and is the basis for virus isolation and enumeration of individual phage plaques (further ahead on plaque assay). Briefly, the plates were divided in two layers, the base layer of LB agar and the top layer of soft agar (0.7% LB Agar - 25 g/L of LB broth with 12.25 g/L of agar). The host bacteria ER2738 forms a lawn on the soft agar thereon X-Gal/IPTG LB agar plate (LB agar plates supplemented with 0.05 g/L X-gal [Nzytech, Lda.] and 0.25 mM IPTG [Nzytech, Lda.]) except where infectious phage particles lyses (for lytic phages) or inhibits the growth of the cells (lysogenic phages), resulting in a localized clear or translucent zone, termed a plaque (Kropinski et al. 2009). ER2738 infected with M13KE forms blue plaques since it can degrade X-Gal. Each phage plaque is termed phage-forming unit (pfu) as all the phage particles on the plaque are derived from the same virion that initial infected one bacteria cell (Kropinski et al. 2009). It is important to note, for phage display purposes, that phage particles from the same plaque have the same DNA and therefore display the same peptide coupled to its coat protein.

Hence, to produce phages on solid medium, a mix of fresh ER2738 (100 μL) and phage suspension (variable volume) with 3-5 mL of soft agar was made and dropped on the top of a X-Gal/IPTG LB agar plate. After overnight incubation at 37 °C on an incubator it was possible to see individual plaques or a phage "lawn" corresponding to a high concentration of pfu/mL. This technique enabled picking individual plaques for phage production in liquid and also to recover all the phages on the lawn using SM (Sodium-Magnesium) buffer, pH 7.5 (appendix 1). To recover all plated phages, it was necessary to cover the soft agar with 2-3 mL of SM buffer and let it with agitation (80 rpm) at 4 °C overnight. The recovered SM buffer contains the phages and could be stored at 4 °C without loss of titter for several weeks.

Another technique used to isolate or produce phage was performed using sterile toothpicks to pick a plaque or to immerse on a phage suspension and pick an X-Gal/IPTG LB Agar plate with a soft agar top layer with 100 μL of fresh ER2738. The toothpick was used to pick the plate several times and then sterilized strips of paper were used to spread the phage along the plate.

To produce virus particles in liquid medium it was necessary to dilute an overnight culture of ER2738 on LB broth (1:100) and then add the phage (in suspension or a picked individual plaque). As described in CHAPTER 1 point 3.1, for obtaining higher titers at phage production, it is relevant to have a short doubling time of the bacterial population. For that, cells need a great amount of oxygen and so aeration conditions of the culture are very important. Therefore, phage production in liquid medium was performed on a 250 mL Erlenmeyer flask containing 100 mL of LB broth for 4.5-5 hours with an agitation of 250 rpm. At the end of the growth time, the medium was either centrifuged at 4000-6000 g for 10 min [Thermo Scientific – CL31R Multispeed] or filtered using a 0.2 μm, 25 mm, polyethersulfone (PES) membrane filter [Fiorini] to obtain only the phage particles and discard the bacteria. The phage suspension obtained was stored at 4 °C for several weeks without loss of titer. When there was no need for a great titer (e.g. for DNA extraction) the same procedure was followed using 15 mL falcon tubes and 5 mL of LB broth.

1.2.3. DROP AND TITER ASSAYS

Drop and titer assay are two essential techniques to confirm the presence of M13KE and to count infectious phage particles, respectively. For those it is necessary to use DLA technique as described above. To perform the drop assay, an X-Gal/IPTG LB Agar plate with a soft agar top layer with 100 μL of fresh ER2738 was used to make a drop of phage suspension and let to dry inside the laminal chamber [Braun]. The plate was incubated overnight at 37 °C, and the presence of M13KE phage was confirmed when blue plaques appeared on the bacterial lawn.

To perform the titer assay, several dilutions of the phage suspension from 10^{1} to 10^{12} (well 1 to 12) were prepared on a sterile microtiter 96-well plate [Fisher Scientific]. First, all twelve wells from a line were filled up with 90 μL of SM buffer. Then, 10 μL of the phage suspension were diluted on well number 1 (10^{-1}). The next dilutions were made using 10 μL of the previous dilution (see Fig. 12).

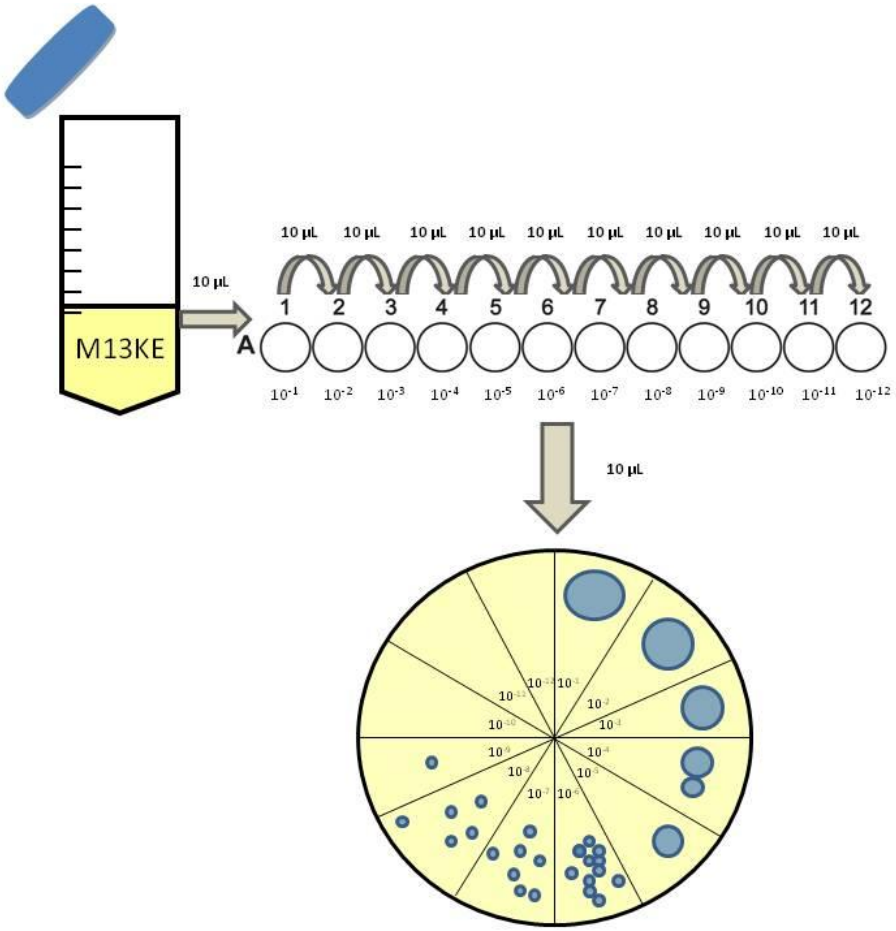

Fig. 12 – Schematic representation of the titer assay.

A drop of every dilution was plated as described above. After incubation, infectious phage particles were counted on the dilution containing 3-100 plaques. Phage titer was calculated using equation 5.

$$[\text{M13KE}]\ (pfu/\text{mL}) = \frac{number\ of\ blue\ plaques \times dilution\ factor}{drop\ volume\ (\text{mL})} \quad (5)$$

1.3. RANDOM PEPTIDE LIBRARIES CONSTRUCTION

Libraries construction was performed using M13KE phage. This process involves the cloning of random sequences on N-terminus of pIII gene of the viral vector. M13KE was extracted on its double stranded form to be further used as a cloning vector (point 1.3.1), and then digested using specific restriction enzymes (1.3.3). The inserts containing the random sequences were constructed using designed primers, digested using restriction enzymes and linked to

M13KE plasmid (1.3.6). Finally, M13KE phages containing the inserts were introduced on chemically competent ER2738 to be multiplied (1.3.9).

1.3.1. M13KE DOUBLE-STRANDED DNA EXTRACTION

To extract M13KE replicative form it was necessary to produce M13KE in liquid medium as described on point 1.2.2. After growth, 1 mL of the culture was transferred to a 2 mL eppendorf [Labbox] and centrifuged at maximum speed [Eppendorf – Centrifuge Eppendorf 5418] for 1 min to obtain the bacterial pellet. Next, the pellet was resuspended on 200 μL of pre-chilled P1 solution (appendix 1). Then, 400 μL of P2 solution were added (appendix 1) and the eppendorf was inverted 5-10 times to mix. Last, 300 μL of P3 solution (appendix 1) were added and the mix was lightly mixed. To finalize the sensitization of the bacteria cell membrane the mix was incubated on ice for 10 min. This procedure, adapted from the work of Birnboim and Doly (Birnboim & Doly 1979) allows obtaining only the circular closed DNA of M13KE that stays in the supernatant whereas the chromosomal bacterial DNA forms an insoluble clot.

Finally, M13KE double stranded DNA was recovered from the supernatant using ethanol precipitation. After the incubation time on ice, the mix was centrifuged for 5 min at maximum speed, and the supernatant was transferred to a new eppendorf containing 1 mL of iced absolute ethanol [Fisher Scientific]. Afterwards, this solution was mixed and centrifuged again for 5 min at maximum speed. Absolute ethanol was discarded and 1 mL of iced 70% ethanol was added to wash the DNA pellet. The vortexing and centrifugation steps were repeated. The supernatant was removed entirely and the pellet was typically dried completely for 2 hours at room temperature. The DNA pellet was resuspended on 15-30 μL TE (Tris-EDTA) buffer (appendix 1). DNA concentration was assessed using NanoDrop 1000 [Thermo Scientific] and stored at -20 °C.

1.3.2. M13KE SIZE CONFIRMATION

Before proceeding to the subsequent steps of the random peptide libraries construction, to the confirmation that the extracted DNA sample was indubitably the M13KE vector, was required. Therefore, the DNA sample obtained was linearized using HindIII restriction enzyme from Thermo Scientific (1 U/μL) which cuts the vector at 6234 bp. The linearization was performed at 37 °C on a bench heat block [Alfagene] for a period of 2 hours, minimum. On Table 3 it is presented the standard mix for this digestion reaction. To confirm the size of the

DNA sample 5 μL of the reaction mix were used to run on a 1% agarose gel (CHAPTER 3, point 1.3.8).

Table 3 - Standard mixture for digestion of dsM13KE samples

Component	Volume	Final Concentration
Microbiology water	16 μL	N/A
10X FastDigest Buffer	2 μL	1X
FastDigest Enzyme	1 μL	1 U
dsM13KE	1 μL	(up 1 μg)
Total Volume	20 μL	-

1.3.3. M13KE DOUBLE DIGESTION

M13KE vector was double digested using fast digestion enzymes Acc65I and EagI [Thermo Scientific, FD0904 and FD0334]. Double digestion was performed at 37 °C on a bench heat block [Alfagene] for a period of 2 hours, minimum. These enzymes cut at 1611 bp and 1631 bp respectively (see Fig. 13) just on the N-terminal of gene III leaving sticky ends. None of the used enzymes cuts the vector anywhere else.

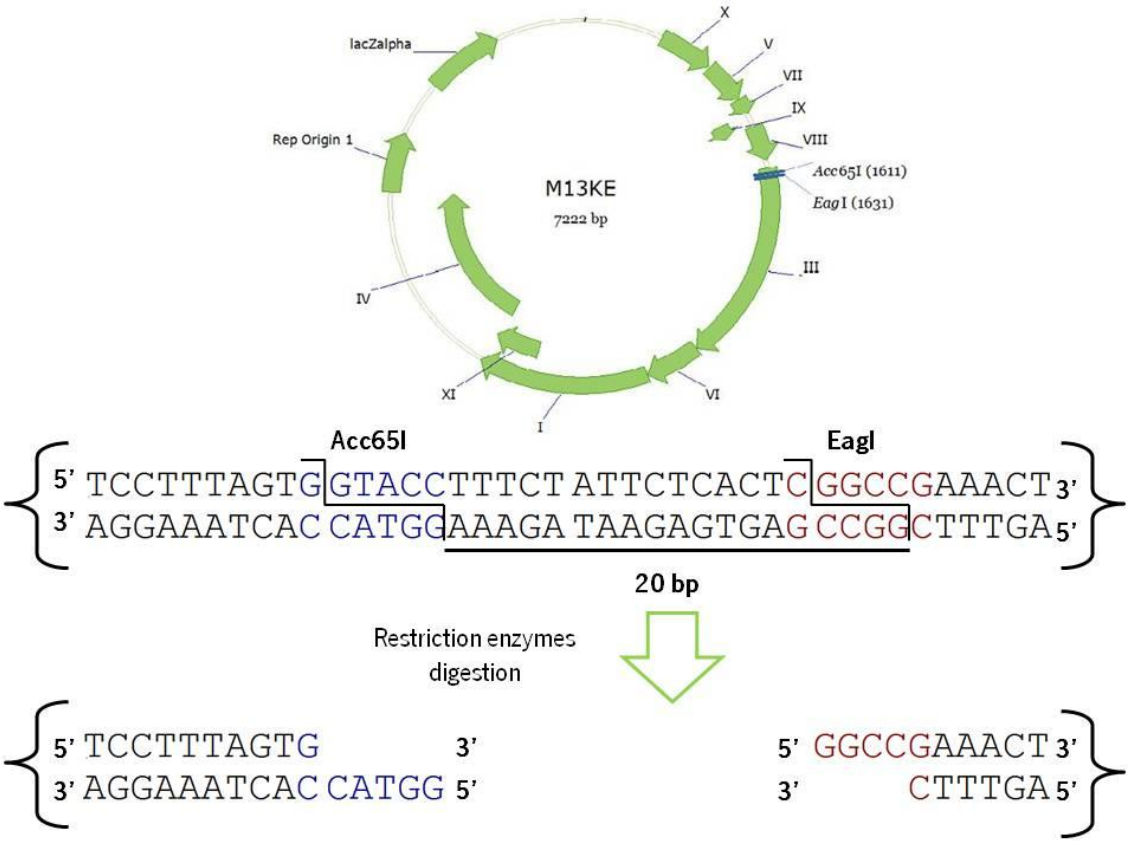

Fig. 13 – M13KE DNA restriction map of Acc65I and EagI restriction enzymes. Schematic representations of M13KE double digestion with Acc65I and EagI; the excluded fragment has a length of 20 bp.

The protocol for double digestion of plasmid DNA from Thermo Scientific was followed for double digestion of dsM13KE sample such as scaling up rules. On table 4 it is presented the standard mix for this digestion reaction. When using more than 1 μg of DNA, sample volumes and times of reaction were adapted taking into account that the combined volume of both enzymes cannot exceed 1/10 of total reaction volume.

Table 4 – Standard mixture for double digestion of dsM13KE samples with Acc65I and EagI from Thermo Scientific

Component	Volume	Final Concentration
Microbiology water	15 μL	N/A
10X FastDigest Buffer	2 μL	1X
FastDigest Enzymes	1 μL of each enzyme	1U
dsM13KE	1 μL	Up 1 μg
Total Volume	20 μL	-

Double digestion was followed by enzyme inactivation at 65 °C for 5 min. The double digested DNA was the precipitated (as explained on point 1.3.5.), in order to remove the restriction enzymes. To confirm the digestion of the viral vector, 10 μL of the volume sample were used to run on a 1% agarose gel (point 1.3.8.).

1.3.4. INSERTS CONSTRUCTION

7-mer and 12-mer codifying sequences were constructed following the protocol provided for New England Biolabs® for libraries construction (NEW ENGLAND BIOLABS ® n.d.). The primers presented on table 5 were supplied by FreeLab at a concentration of 100 μM.

Table 5 – Primer's sequence and melting temperature used for libraries construction. The melting temperatures above were calculated using Modified Breslauer's thermodynamics, dH and dS parameters. The underlined zones on the reverse primers are the restriction sites for EagI and *Acc65I*, respectively. Bold zones are the overlapping sequences

Name	Length (bp)	Sequence (5' – end start)	Melting Temperature (TM)°C	Stock concentration (μg/μL)
PhD.FW	25	CATG**CCCG***GGTACC*TTTCTATTCTC	54.4	8.25
PhD7.RV	47	CATGTTTCGGCCGA(NNN)$_7$AGAGT**GAG AATAGAAA***GGTACC*CGGG	54.4	15.51
PhD12.RV	82	CATGTTTCGGCCGAGCCCTGAAAATAA AGATTCTCACCTCCACC(NNN)$_{12}$AGAGT **GAGAATAGAAA***GGTACC*CGGG	54.4	27.06

Preparing the inserts for the random peptide libraries construction involved three steps – primer annealing, primer extension and finally, the insert digestion with the same enzymes used to digest the M13KE vector (EagI and Acc65I fast digestion enzymes). For the construction of a 90-nucleotide library oligonucleotide, approximately 4 μg of the primer containing the random zone are recommended (PhD7.RV and PhD12.RV) (NEW ENGLAND BIOLABS ® n.d.). For the annealing, 5 μg of the libraries oligonucleotides were incubated with three molar equivalents of the universal extension primer (PhD.FW) in a 50 μL volume of TE buffer containing 100 mM sodium chloride (NaCl). The mix was transferred to a container of boiling water and let slowly cool to less than 37 °C.

The extension of the annealed duplex was performed using the Klenow Fragment, exo- enzyme from Thermo Scientific [EP0421] at a concentration of 10 U/μL and a mix of deoxynucleotides (dNTPs) from KAPA Biosytems [Kn1009] at a concentration of 10 mM of each dNTP. This enzyme presents 100% activity on FastDigest Buffer from Thermo Scientific allowing the use of this buffer on the extension step eliminating the need of adding it for the subsequent digestion reaction. The extension reaction mixture is presented on table 6.

Table 6 – Extension reaction mixture for the random inserts construction

Component	Volume	Final Concentration
Microbiology Water	120 μL	N/A
10X FastDigest Buffer	20 μL	1X
Annealed duplex	50 μL	-
10 mM dNTPs	8 μL	0.4 mM
Klenow Fragment	2 μL	20 U
Total Volume	200 μL	-

The mix was incubated for 10 min at 37 °C and afterwards the Klenow Fragment was inactivated at 75 °C for 15 min. At this point, 4 μL of each insert were saved for further analysis by agarose gel electrophoresis.

As mentioned, the last step is the digestion of the extended duplex using fast digestion enzymes Acc65I and EagI. For that, 1 μL of each enzyme was added to the previous mix. After incubation (2 hours at 37 °C), double digestion was followed by enzyme inactivation at 65 °C for 5 min. The DNA inserts were precipitated as explained on point 1.3.5. and further analyzed by agarose gel electrophoresis before proceeding to ligation.

1.3.5. ETHANOL PRECIPITATION OF NUCLEIC ACIDS

As previously mentioned, restriction enzymes of the DNA samples were removed to assure a correct quantification of the DNA using NanoDrop. Ethanol precipitation of nucleic acids is a simple method used to concentrate and/or purify DNA samples (Green & Sambrook 2012). For that, 3 M sodium acetate was added to the samples to a final concentration of 0.3 M per sample. Afterwards, 2.5 volumes of ice cold absolute ethanol were added to the mix. Next, the samples were put on ice for 15-30 min. DNA was harvested by centrifugation at 14000 rpm for 15 min and the pellet was then washed using ice cold 70% ethanol and spin down for 5 min at

14000 rpm. All the ethanol was removed using a P20 tip. The DNA pellets were let to dry for 2 hours at the bench. The DNA pellet was ressuspended with 15-30 μL of TE buffer. DNA samples could then be quantified and stored at -20 °C.

1.3.6. LIGATION OF M13KE::7-MER AND 12-MER OLIGONUCLEOTIDES

The ligation of the M13KE double digested vector with the inserts for 7-mer and 12-mer libraries was carried out using T4 DNA ligase (5 U/μL) [Thermo Scientific, EL0014]. The ligation protocol from Thermo Scientific for DNA inserts ligation into DNA vectors with sticky ends was followed.

A ratio of 3:1 of insert molar excess and 20-100 ng of vector were recommended to the ligation protocol. The amount of inserts needed was calculated using equation 6.

$$\text{Insert amount (ng)} = \frac{insert}{vector} lenght\ (kb) \times ng\ of\ vector\ \times \frac{insert}{vector}\ desired\ ratio \quad (6)$$

Two different incubation times and temperatures for the ligation were tested: 2 hours incubation at 22 °C and overnight incubation at 16 °C. The ligation mix was followed as presented on table 7.

Table 7 – Ligation reaction mixture for M13KE::7-mer and M13KE::12-mer libraries construction

	7-mer library (0.32 kb)	12-mer library (0.69 kb)	
Component			**Final Concentration**
Microbiology Water	Up to 20 μL	Up to 20 μL	N/A
10X T4 DNA Ligase Buffer	2 μL	2 μL	1X
M13KE Vector (7.2 kb)	100 ng	100 ng	-
DNA Insert	≈ 15 ng	≈ 30 ng	-
T4 DNA Ligase	1 μL	1 μL	5 U
Total Volume	20 μL	20 μL	-

T4 DNA ligase was inactivated at 65 °C for 5 min. Before transforming the chemically competent cells, the ligation was confirmed by PCR amplification.

1.3.7. CLONING CONFIRMATION BY PCR AMPLIFICATION

The ligation of inserts into a DNA vector can be confirmed by PCR, using the appropriate primers to amplify the sequence flanking the insertion site. When the cloning is successfully achieved the DNA fragment created by PCR includes the flanking zone plus the insert zone (see Fig. 14). Hence, performing a PCR with a control without insert and the ligation samples for insertion confirmation is expected to create DNA fragments with variable sizes that can be analyzed by gel electrophoresis.

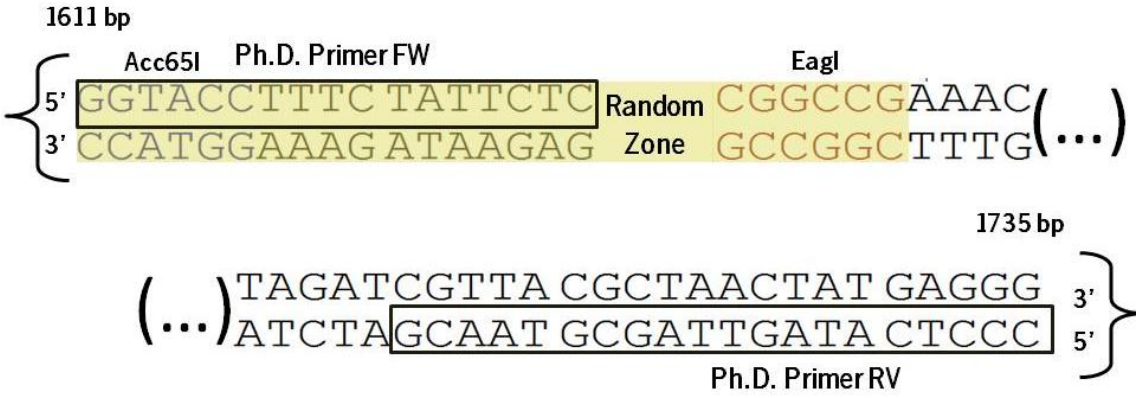

Fig. 14 – Excision of M13KE genome map. In the boxes it is presented the Ph.D. Primer FW and RV used for PCR cloning confirmation. The zone in color corresponds to the cloning site for the construction of Ph.D. libraries 7 and 12.

To confirm the insertions at the N-terminus of pIII gene of M13KE by PCR amplification the primers presented on table 8 were used.

Table 8 – Primers used for insertion confirmation on pIII gene. PhD.FW – Primer Forward; PhD.RV – Primer Reverse. The melting temperatures were calculated using Modified Breslauer's thermodynamics, dH and dS parameters

Name	Length (bp)	Sequence (5' – end start)	Melting Temperature (TM)°C
PhD.FW	25	CATGCCCGGGTACCTTTCTATTCTC	54.4
PhD.RV	20	CCCTCATAGTTAGCGTAACG	51.8

The PCR reaction was performed in a MyCyclerTM Thermal Cycler [Bio-Rad Laboratories, Inc.] using KAPA Taq DNA Polymerase (5 U/µL) [KapaBiosystems, KE1000]. PCR mix and PCR cycling protocol are presented on table 9 and 10, respectively.

Table 9 – PCR standard reaction mixture with KAPA Taq DNA Polymerase

Component	Volume	Final Concentration
Microbiology Water	38.8 μL	N/A
10X KAPA Taq Buffer A	5.0 μL	1X
10 mM dNTPs Mix	1.0 μL	0.2 mM
10 μM Primer FW	2.0 μL	0.4 μM
10 μM Primer RV	2.0 μL	0.4 μM
KAPA Taq DNA Polymerase	0.2 μL	1 U
Template DNA	1.0 μL	-
Total Volume	50 μL	-

Table 10 – PCR cycling protocol to amplify M13KE pIII cloning site

Step	Temperature (°C)	Duration	Cycles
Initial denaturation	95	5 min	1
Denaturation	95	30 s	35
Annealing	53	30 s	
Extension	72	1 min	
Final Extension	72	1 min	1
Hold	4	∞	1

The size of PCR resulting double stranded DNA fragments was analyzed by DNA gel electrophoresis with an agarose and SGTB gel as explained on point 1.3.8.

1.3.8. AGAROSE GEL ELECTROPHORESIS

DNA electrophoresis was performed to confirm the digestion of M13KE samples. Agarose gels (1% w/v) were typically prepared for a total volume of 50 mL using 0.5 g of agarose [Fisher Scientific] to 50 mL of 1X Tris-acetate-EDTA (TAE) buffer (appendix 1) and the mixture was heated until the agarose dissolve. For this volume, 1.25 μL of Thiazole Orange [Sigma-Aldrich Co. LLC, Fluka Analytical, 17237] was added to the solution. The electrophoresis was typically carried out at 80 V for 60 min using 1X TAE as running buffer. Samples were loaded with 50% glycerol [Fisher Scientific] and ladder III [NZYTech] was used for easy quantification and size determination of the DNA sample. Gel analysis was carried out using ChemiDoc XRS [Bio-Rad Laboratories, Inc.]

For small DNA fragments, such as DNA fragments resulting of PCR amplification of insertions on gene III of M13KE, agarose gels (1% w/v) with SGTB [GRiSP] that allowed a higher resolution and a better separation of the fragments, were prepared. For a 50 mL gel, 0.5 g of agarose plus 5 mL of SGTB were diluted on 45 mL of deionized water. For this volume, 1.25 μL of Thiazole Orange was added to the solution. The electrophoresis was typically carried out at 200 V for 30 min using SGTB 10% (v/v) as running buffer. Samples were loaded with 50% glycerol and ladder VI [NZYTech] was used for easy quantification and size determination of the DNA sample. Similarly, the gels were analyzed using ChemiDoc XRS.

1.3.9. CHEMICAL COMPETENT CELLS PREPARATION AND TRANSFORMATION

The process of introducing DNA into a bacteria cell is called transformation. The bacterial cells that can receive foreign DNA are called competent. None of the *E. coli* strains are naturally competent, so there are two main approaches to make them competent; chemical transformation and eletroporation. For the purpose of library construction chemically competent *E. coli* were prepared.

Chemical competent ER2738 were prepared diluting (1:100) an overnight culture of ER2738 on 50 mL of LB medium on a 250 mL Erlenmeyer flask. The diluted culture was grown until an OD of 0.2-0.5 and then put on ice for 10 min to stop the growth. After this, the bacterial culture was centrifuged at 4 °C to obtain a bacterial pellet. The pellet was resuspended on filtered, ice-cold TSS (Transformation – Storage Solution) buffer (appendix 1). Aliquots of the bacterial competent were cells can be stored for long periods at -70 °C without losing of its capacity.

All the steps of chemical competent cells transformation were carried out at 4 °C unless otherwise mentioned. Competent cells were unfrozen and 1-5 ng of the DNA of interest was mixed and let for 30 min. After that, cells suffered a heat shock, for 30 s at 42 °C and then again for 2 min at 4 °C. To grow the bacteria and consequently multiply the phage, SOC (Super Optimal Broth with Catabolite repression) medium (31.54 g/L of SOC broth from NzyTech, MB11901) was added at room temperature and put at 37 °C, 200 rpm for 45-60 min. To confirm the transformation some of the SOC outgrowth was plated on X-Gal/IPTG LB agar plate and incubated overnight at 37 °C. Using a metal rod to spread the outgrowth volume on the plate it is possible to distinguish the bacteria that do not incorporate the foreign DNA (white colonies) and those that did (blue colonies). When the presence of positive clones was confirmed the SOC

outgrowth was amplified in liquid medium and the obtained phage particles purified as explained on point 1.2.3.

1.4. *IN VITRO* BIOPANNING

The choice of library to be used in a particular biopanning experiment is dependent upon many factors. Not knowing detailed information about the target-ligand interaction structure makes impossible to predict which type of library will yield the most productive ligands. Therefore, for the purpose of this work, three types of libraries were used displaying randomized linear 7-mer, 12-mer and also C7C commercial library (Ph.D.™ C7C Phage Display Peptide Library from New England Biolabs®) were used to perform the *in vitro* biopanning procedures. 7-mers may be useful for targets requiring binding elements concentrated in a short stretch of amino acids and 12-mers are long enough to fold into short structural elements, which may be useful when panning against targets that require structured ligands (NEW ENGLAND BIOLABS ® n.d.).

Two different methodologies were used to obtain the phages presenting the RKO cell surface specific peptides. The main differences among the methodologies used was the fact that one only relies on repeated washes of the cells monolayer to separate the phages that bounded from the phages that did not bound to the surface of RKO cells; while the other, an adaptation from BRASIL method mentioned on CHAPTER 2, point 3.3, relies on centrifugations in order to pursue the same purpose. To facilitate differentiation, the two different procedures will be named as conventional panning and BRASIL panning, respectively. Four rounds of selection were done using both methodologies and RKO cell line.

1.4.1. CONVENTIONAL PANNING

Conventional panning was done using adherent cells on 6-well plates at 80-90% confluency. The RKO cell seeding was performed on the day preceding each round of selection. To ensure homogeneity of the cell number on each well for every round and for all libraries, the same amount of cells was put to grow at the same time. The well named control (Fig. 15) was used to evaluate the number of cells on each well at the time of the panning procedure.

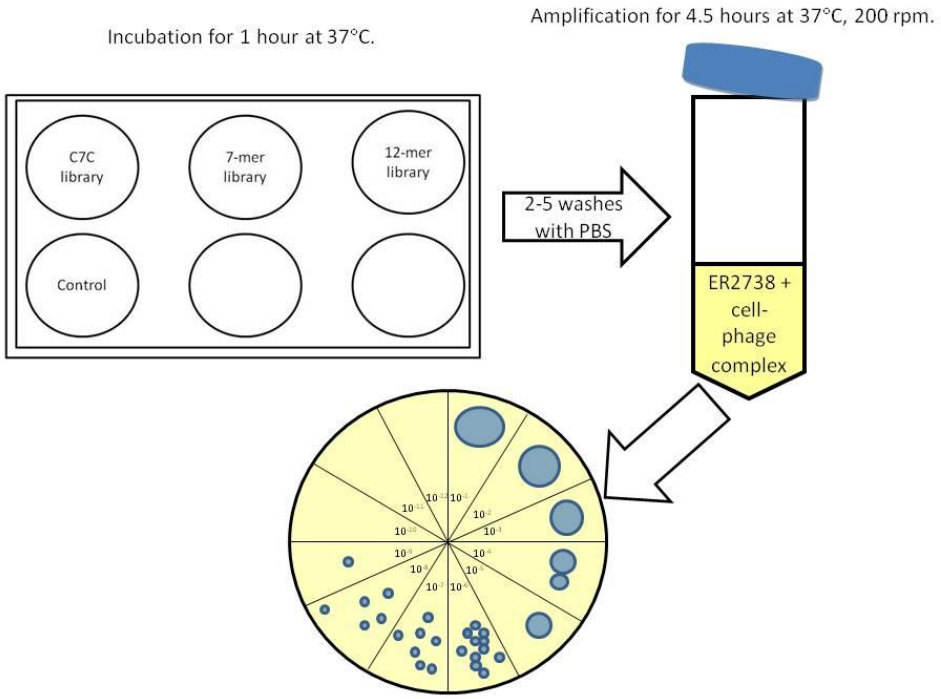

Fig. 15 – Schematic representation of a conventional biopanning round.

The cell monolayer on each well was washed with 1X PBS buffer and incubated with DMEM + 3% BSA [NzyTech, MB04601] for 1 hour at 37 °C in a humidified atmosphere with 5% CO_2. This incubation with serum-free medium was conducted to remove the attachment factors present on FBS, to consequently ensure that more locals from the cell surface were available for ligation with the phage displayed peptides. The incubation with combinatorial phage libraries was done diluting the phage particles on 1 mL of DMEM + 5% BSA for 1 hour at 37 °C in a humidified atmosphere with 5% CO_2. BSA is frequently used as a blocking agent to avoid nonspecific bindings to free surface areas in immunoassays like enzyme-linked immunosorbent assays (Xiao & Isaacs 2013) and herein it was used with the same purpose. Combinatorial libraries used were Ph.D.™ C7C from New England Biolabs® [E8121L] and self-made Ph.D. 7 and Ph.D. 12 as explained above. The initial incubation was done with $2x10^{10}$, $5,7x10^{10}$ and $1,5X10^{10}$ phage particles for C7C, 7-mer and 12-mer libraries, respectively. The number of phage particles used was calculated dividing the phage concentration by the volume of the phage library used for incubation.

Next, unbounded phages were wiped out by washing the monolayer with 1X PBS buffer (2-5 times). The number of washes was increased from round 1 (two washes) to round 4 (five washes). The bound phages were recovered, together with the cells using Trypsin/EDTA solution.

On each round a volume of collected phages was saved to do the phage tittering and ELISA (points 1.2.3 and 1.4.3, respectively). For the amplification of selected phage clones to be used in the next round of panning, the remaining cells and phages were amplified on liquid medium as explained on point 1.2.2. and the concentration of phage particles was assessed before the subsequent round (titer assay, point 1.2.3.).

1.4.2. BRASIL BIOPANNING

An adaptation of BRASIL method introduced by Giordano and colleagues (2001) was followed during this thesis, as they reported the use of cells in suspension and here biopanning procedure was followed using adherent cells in 96-well plates at 80-90% confluency. For every round, incubation was done in triplicate for all the libraries. As in conventional biopanning, seeding was done in the day preceding each round and a well was used as a control to evaluate that an approximate number of cells is available for every round and to all libraries. Similarly to conventional panning, the cell monolayer on each well was washed with 1X PBS buffer (appendix 1) and incubated with DMEM + 3% BSA for 1 hour at 37 °C in a humidified atmosphere with 5% CO_2. In this BRASIL adaptation, incubation with combinatorial phage libraries was done for 2 hours at room temperature with agitation (80 rpm) in DMEM + 5% BSA. The initial incubation was done with $2x10^{11}$, $5,7x10^{11}$ and $1,5X10^{11}$ phage particles for C7C, 7-mer and 12-mer libraries, respectively. After incubation, 10 µL of the incubation medium were saved for titer. Then, the cells-phage complexes were scrapped from the surface of each well using a P20 tip and the mix was placed in an aqueous upper phase and then centrifuged through a non-miscible organic lower phage. To prepare the tube for BRASIL the bottom of a cryovial was excised using a red-hot spatula, and the resulting tube ('inner tube') was placed into a 15 mL falcon tube. The organic phase used was dibutyl phthalate:cyclohexane 9:1 (v:v), a combination and ratio optimized by Giordano and colleagues (Giordano et al. 2001). The organic phase was placed into the BRASIL tubes and it was made a bubble of 500 µL using chilled DMEM + 1% BSA (non-organic phase) where the sample was introduced (Derda et al. 2010) (see Fig. 16).

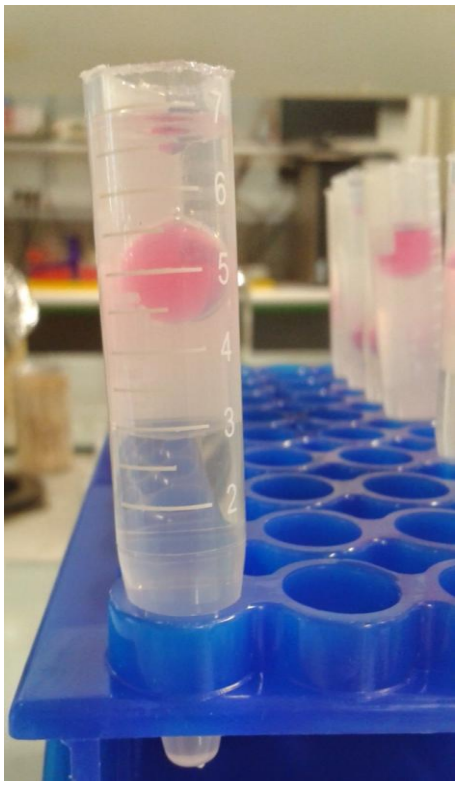

Fig. 16 – BRASIL tube containing the organic phase of dibutyl phthalate:cyclohexane 9:1 (v:v) and the non-organic phase of 500 µL of DMEM + 1% BSA.

When each sample was placed inside the bubble, the centrifugation step was carried out for 10 min at 10000 x g and 4 °C – "BRASIL wash" (Derda et al. 2010). At the end of centrifugation the phages that bounded to the cells were presented on the cell-phage complex pellet and the phages that did not bound were still inside the bubble. To prevent cross-contaminations the bubble was piped out and discarded. Subsequently, the organic phase was also discarded and the pellet was ressuspended in 50 µL of 1 M Tris-HCl Buffer, pH 9 (appendix 1). The obtained pellet was directly used to proceed to the subsequent round without amplification, saving a sample from every round to perform titer assay and ELISAs (points 1.2.3. and 1.4.3).

1.4.3. ELISA

ELISA assays were performed using RKO cells in suspension and the amplified outputs of each round of both biopanning methodologies. In the conventional panning the output of each round was already amplified for the ELISA assays, but this was not the case of the outputs of each round of BRASIL. The amplification was performed on liquid medium as in point 1.2.2.

Also, before ELISA the phage pools were purified to guarantee that there were no contaminants present on the samples that could alter ELISA results.

The phage purification was performed according to New England Biolabs® recommendations. The resulting phage particles selected on each round of biopanning were collected using polyethylene glycol (PEG) precipitation (Fig. 17).

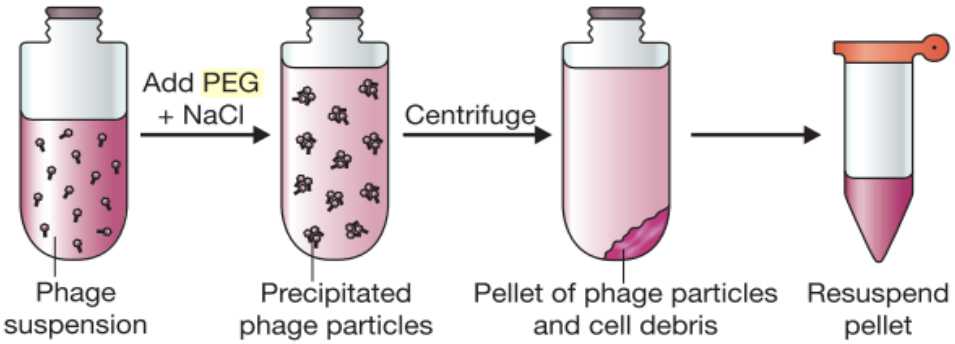

Fig. 17 – Collection of phage particles by polyethylene glycol (PEG) precipitation [taken from: (Brown 2010)].

After purifying the phage pools from all rounds of biopanning experiments it was possible to proceed to the ELISA assays (Fig. 18) which consist on several consecutive incubations as explained below. All incubations were followed by a centrifugation step at 300 x g, 4 °C for 10 min to obtain a cell pellet. The cell pellet obtained was washed using 1X TBST (Tris-Buffered Saline and Tween 20) buffer (appendix 1) before proceeding to the subsequent incubation. This washing step was followed by a similar centrifugation step.

RKO cells were detached, counted (points 1.1.1 and 1.1.2.) and an equal number of cells (1 X 10^5) were transferred to an eppendorf tube. The cells were first incubated for 1 hour at 4 °C in blocking buffer - DMEM containing 5 mg/mL of BSA and 0.01% of PBS + 0.1% EDTA (appendix 1). BSA is used to eliminate the residual binding capacity of the plastic (Xiao & Isaacs 2013) and it can also stabilize the biomolecules and reduce non-specific interactions (Gibbs & Kennebunk 2001).

After this time, the cell pellet (centrifuged and washed as mentioned above) was incubated for 1 hour at room temperature with agitation (80 rpm), with 100 µL of diluted phage pools (diluted on a 1:2 ratio using 1X TBST). The cell pellet was carefully homogenized pipetting the mix up and down to ensure that all surface proteins of RKO membrane were available to

ligation with the peptide present on the phage coat. Following the incubation of RKO cells with the phage clones displaying the specific RKO-cell peptide, RKO-phage complexes were incubated with HRP/Anti-M13 Monoclonal Conjugate [GE Healthcare Life Sciences, 27942101]. This monoclonal antibody reacts specifically with the bacteriophage M13 major coat protein product of gene VIII as it binds to an epitope present on pVIII covering the N-Terminal region of this protein (GE Healthcare 2007). HRP/Anti-M13 Monoclonal Conjugate was diluted (1:5000) on a blocking buffer (DMEM containing 5 mg/mL of BSA) and 200 μL of that preparation was used to the incubation with RKO-phage complexes for 1 hour at room temperature with agitation (80 rpm). The final incubation step was performed using OPD (o-phenylenediamine dihydrochloride) substrate [Thermo Scientific, 34005]. OPD is a water-soluble substrate for horseradish peroxidase (HRP) that produces a yellow-orange product and it is detectable on a range of 445-495 nm on ELISA plate readers. OPD substrate was prepared using 5 mg of OPD powder that were dissolved on 9 mL of deionized water and 1 mL of Stable Peroxide Buffer [Thermo Scientific, 34062] immediately before use. Incubation was conducted using 100 μL of OPD substrate, for 15 min at room temperature. Next, the content of the eppendorfs was immediately transferred to a 96-well plate and the absorbance was measured at 445 nm.

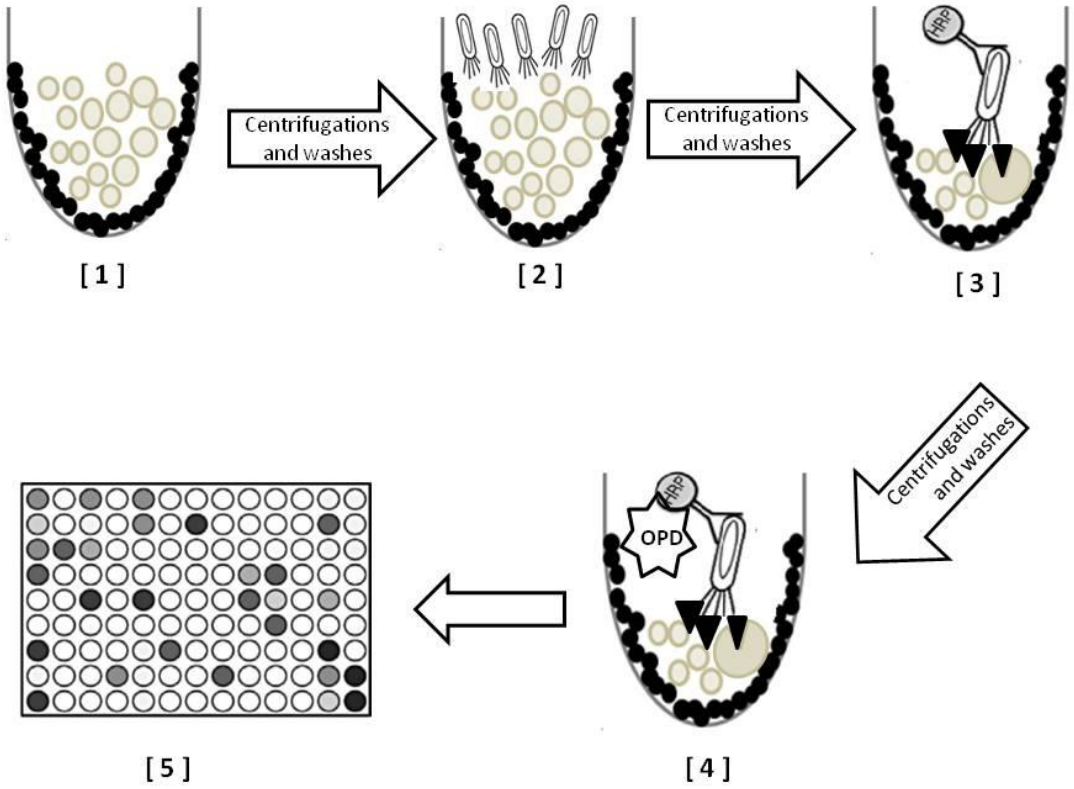

Fig. 18 – Schematic representation of phage ELISA. RKO cells incubated with DMEM + 5mg/mL of BSA and 0.01% of PBS + 0.1% EDTA - RKO cells (○), BSA (●) [1]; RKO cells incubated with phage pool [2]; RKO-phage complexes incubated with HRP/Anti-M13 Monoclonal Conjugate – RKO cell specific peptide [3]; RKO-phage complexes incubated with HRP/Anti-M13 Monoclonal Conjugate and OPD substrate [4] and quantitative analysis performed through absorbance measurements [5].

Phage pools were analyzed in triplicate. A negative control (assay performed on the same conditions but without RKO cells) was also used to subtract the background noise, i.e. exclude false positive results or results that can be wrongly considered good, as it enables the evaluation of phage ligation to plastic – plastic binders. A positive control, namely M13KE without any displayed peptide, was also used as a lower limit on the analysis of absorbance (abs) results.

1.4.4. LIGAND SEQUENCING: SAMPLES PREPARATION

Twelve random picked phage clones resulting from the pools evaluated by ELISA were randomly picked to send to sequencing. Two clones i.e. plaques from round 4 of each selection method and for each library were picked from X-Gal/IPTG plates. Individual phage clones were picked from plates stored at 4 °C and put to grow, as explained on point 1.2.3. After filtering, 500 μL of phage stock were transferred to an eppendorf tube.

Afterwards, phage clones were precipitated, for 10 min at room temperature after adding 200 µL of 20% PEG 8000/2.5 M NaCl (in deionized water). This mixture was then centrifuged at 14,000 rpm for 10 minutes at 4 °C and the supernatant was discarded. At this point, the phage pellet was thoroughly ressuspended on iodide buffer (appendix 1) and 250 µL of ice cold absolute ethanol were added and let to incubate for 10 min at the bench. The mixture was centrifuged under the same conditions described above and the pellet was washed with ice cold 70% ethanol. DNA pellet was dried at the bench and ressuspended on TE buffer. Further, the DNA was quantified on NanoDrop.

Succeeding DNA extraction and purification, the insert zone was amplified through PCR using the primers presented on table 11.

Table 11 – Primers used to amplify the cloning region. Ph.D.sequencing.FW – Primer Forward; Ph.D.RV – Primer Reverse. The melting temperatures were calculated using Modified Breslauer's thermodynamics, dH and dS parameters

Name	Length (bp)	Sequence (5' – end start)	Melting Temperature (TM) °C
Ph.D.sequencing.FW	20	TTAACTCCCTGCAAGCCTCA	51.8
Ph.D.RV	20	CCCTCATAGTTAGCGTAACG	51.8

The resulting PCR fragments have variable sizes as they are representative of Ph.D. C7C, 7 and 12 libraries. Without insert, the resulting fragment has a length of approximately 335 bp.

PCR conditions were optimized as the ideal annealing temperature of the primers was unknown. According to that, firstly a temperature gradient PCR (from 45 °C to 70 °C) was conducted to evaluate the optimal annealing temperature that was found to be 46.8 °C. PCR protocols were performed in a MyCyclerTM Thermal Cycler using KAPA Taq DNA Polymerase [KapaBiosystems, KE1000]. The composition of the standard mix used to amplify all sequencing templates is provided in table 12. The PCR cycling protocol was followed as presented on table 9 (point 1.2.6.) except for the annealing temperature used (46.8 °C).

Table 12 – PCR mixture used to amplify sequencing templates

Component	Volume	Final Concentration
Microbiology Water	Up to 20 μL	N/A
10X KAPA Taq Buffer B	2.00 μL	1X
10 mM dNTPs Mix	0.40 μL	0.20 mM
10 μM Primer FW	0.80 μL	0.40 μM
10 μM Primer RV	0.80 μL	0.40 μM
KAPA Taq DNA Polymerase	0.08 μL	0.40 U
Template DNA	5.00 μL	-
Total Volume	20 μL	-

Following PCR, the DNA fragments length was confirmed by gel electrophoresis, on a SGTB gel as described on point 1.3.8. Next, the PCR products were cleaned, concentrated and quantified as described in point 1.3.5.

Finally, a 10 μL sample containing 5 μL of PCR product (with a concentration of 50 ng/μL), 5 μL of 5 pmol/μL primer forward (from table 11) was prepared on a 1.5 mL eppendorf. The samples were sequenced by Macrogen, Amsterdam, Netherlands.

2. NANOLUC™ GENE INSERTION ON M13KE

NanoLuc™ gene is a 512 bp (base pairs) gene that was amplified from pNL1.1 [*Nluc*] Vector [Promega, N1001] (Fig. 19). This gene codifies NanoLuc™, a 19.1 kDa, ATP-independent luciferase that utilizes furimazine as a substrate to produce high intensity, bioluminescence at 460 nm (Binkowski 2012).

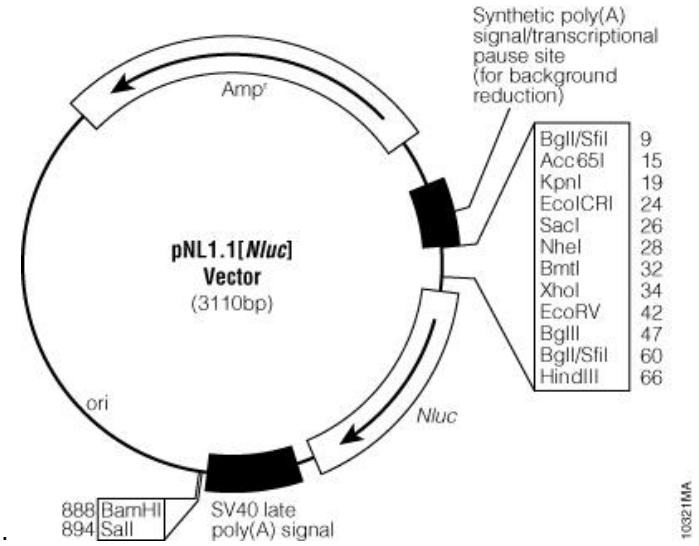

Fig. 19 – pNL1.1 *[Nluc]* Vector [Promega, N1001].

NanoLuc™ luciferase is much smaller than renilla and firefly luciferases, and produces a 150-fold increase in specific activity (Fig. 20) (Binkowski 2012).

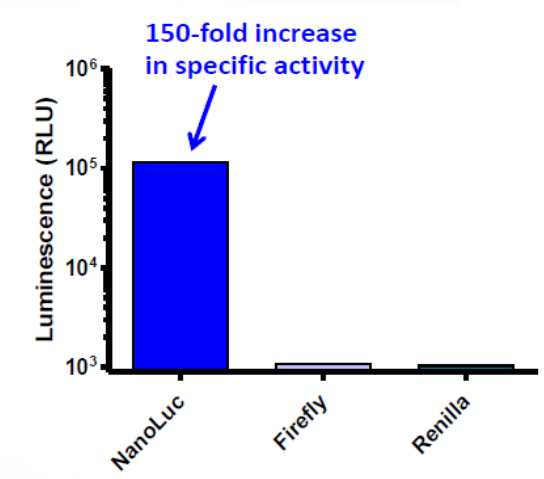

Fig. 20 – Detection of 50 attomoles of purified enzyme; Reagents: NanoLuc™/Nano-Glo™, Firefly/ONE-Glo™ Renilla/Renilla-Glo™ [taken from: (Binkowski 2012)].

The insertion of NanoLuc™ gene into the 5' end of gene VIII comprises 4 major steps; namely the amplification of NanoLuc™ gene using specific primers to insert specific restriction sites that will enable the ligation to M13KE; NanoLuc™ gene and M13KE double digestion, followed by M13KE DNA::NanoLuc™ Ligation. Afterwards, the ligation was introduced on chemically competent ER2738. This procedure is described in detail in following sections.

2.1. NANOLUC™ GENE AMPLIFICATION

NanoLuc™ gene was obtained as a part of a commercial plasmid, pNL1.1 [*Nluc*] Vector from Promega. NanoLuc™ was amplified using two specific primers (table 13) supplied by FreeLab at a concentration of 100 μM.

Table 13 – Primers used to amplify NanoLuc™ gene from pNL1.1. NlucM13.FW – Primer Forward with the restriction site for SnaBI (underlined); NlucM13.RV – Primer Reverse with the restriction site for BspHI (underlined). The melting temperatures were calculated using Modified Breslauer's thermodynamics, dH and dS parameters

Name	Length (bp)	Sequence (5' – end start)	Melting Temperature (TM)°C
NlucM13.FW	41	TATATATACGTAATGGTCTTCACACTCGAAGATTT CGTTGG	71.03
NlucM13.RV	31	TATATATCATGACGCCAGAATGCGTTCGCAC	72.13

Using both primers, the NanoLuc™ gene was amplified and the two restriction sites were inserted at each end resulting in a PCR product presenting NanoLuc™ gene with the restriction site for SnaBI at the 5' end and the restriction site for BspHI on the 3' end.

The amplification was done using DNA polymerase KAPA HiFi (1 U/μL) [KAPABiosystems, KK2102] in MyCyclerTM Thermal Cycler. PCR conditions were optimized since the ideal annealing temperature of the primers was unknown. A range of annealing temperatures from 60 °C to 75 °C was tested. The optimal temperature that enabled a greater amplification without unspecific bounding of the primers was found to be 60 °C. The composition of the standard mix used to amplify NanoLuc™ gene is provided on table 14.

Table 14 – PCR mixture used to amplify NanoLuc™ gene

Component	Volume	Final Concentration
PCR grade Water	Up to 25 µL	N/A
5X KAPA HiFi Fidelity Buffer	5.00 µL	1X
10 mM dNTPs Mix	0.75 µL	0.30 mM
10 µM Primer FW	0.75 µL	0.30 µM
10 µM Primer RV	0.75 µL	0.30 µM
KAPA HiFi DNA Polymerase	0.50 µL	0.50 U
Template DNA	1-10 ng	-
Total Volume	25 µL	-

The PCR cycling protocol used to perform NanoLuc® gene amplification is presented on table 15.

Table 15 – PCR cycling protocol for NanoLuc® amplification with DNA polymerase KAPA HiFi

Step	Temperature (°C)	Duration	Cycles
Initial denaturation	95	5 min	1
Denaturation	98	20 s	25
Annealing	60	15 s	
Extension	72	15 s	
Final Extension	72	5 min	1
Hold	4	∞	1

To ensure that the gene was correctly amplified, 5 µL of the PCR product were loaded on a 1% agarose gel and the size of the DNA fragment was confirmed by DNA electrophoresis.

2.2. NANOLUC® GENE CLEAN-UP

Before proceeding to the subsequent steps, the resulting PCR product had was cleaned as the pNL1.1 vector was still present on the mix and this could be a major drawback to achieve a successful cloning. Therefore, all volume resulting from PCR was loaded on an agarose gel. After DNA electrophoresis the bands corresponding to NanoLuc® were cut and DNA samples were purified using NucleoSpin® Gel and PCR Clean-up kit [Macherey-Nagel GmbH & Co. KG]. The eluted DNA was split in aliquots and stored at -20 °C.

2.3. M13KE AND NANOLUC™ DOUBLE DIGESTION

The M13KE vector and amplified NanoLuc™ gene were double digested using SnaBI and BspHI [New England Biolabs®, R0517S and R0130S, respectively]. Double digestion was performed at 37 °C on a bench heat block for a maximum period of 3 hours. These enzymes cut at 1271 bp and 1300 bp of M13KE vector on an overlapping zone that contains the end of gene IX and the beginning of gene VIII (Fig. 21). None of the used enzymes cuts the vector anywhere else.

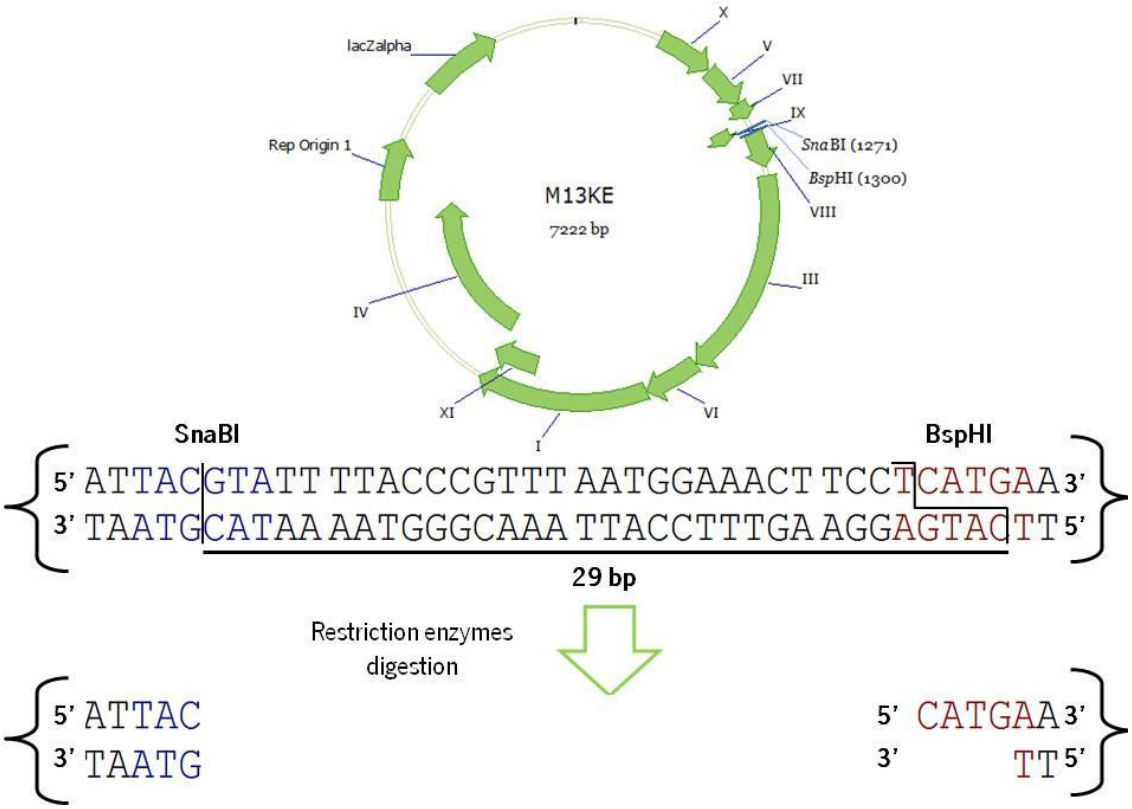

Fig. 21 – M13KE DNA restriction map of SnaBI and BspHI restriction enzymes. Schematic representations of M13KE double digestion with SnaBI and BspHI; the excluded fragment has a length of 29 bp.

The composition of the standard mix used for the double digestion of M13KE vector and amplified NanoLuc™ gene is provided on table 16.

Table 16 – Standard mix for M13KE vector and NanoLuc™ gene double digestion with SnaBI and BspHI from New England Biolabs®

Component	Volume	Final Concentration
Microbiology water	42.5 μL	N/A
10X CutSmart™ Buffer	5.0 μL	1X
DNA sample	1.0 μL	Up to 1 μg
SnaBI	1.0 μL	1 U
BspHI	0.5 μL	1 U
Total Volume	50 μL	-

Double digestion was followed by enzyme inactivation at 80 °C for 20 min. Before proceeding, the DNA samples were precipitated as described on point 1.3.5.

2.4. DNA DEPHOSPHORYLATION

Next to DNA digestion and to prevent self-ligation of the restricted fragment resulting from M13KE double digestion, 1 μL volume of 10X Antarctic Phosphatase reaction buffer [New England BioLabs, M0289S] and 1 μL volume of Antarctic Phosphatase (5 U) [New England BioLabs, M0289] were added to approximately 5 μg of M13KE phage DNA digested with SnaBI and BspHI in a 10 μL reaction. The mixture was incubated at 37 °C for 60 min and then the enzyme was heat inactivated at 70 °C for 5 min.

2.5. M13KE DNA::NANOLUC™ LIGATION

The ligation of the M13KE double digested vector and NanoLuc™ gene was carried out using T4 DNA ligase (5 U/μL).

A standard 3:1 of molar insert excess was performed using 50 ng of vector. The amount of inserts needed was calculated using equation 6.

The ligation mix used is presented on table 17. Ligation was performed overnight at 16 °C.

Table 17 – Ligation reaction mixture for insertion of NanoLuc™ gene on M13KE vector

Component	Volume	Final Concentration
Microbiology Water	Up to 20 μL	N/A
10X T4 DNA Ligase Buffer	2 μL	1X
M13KE Vector (7.2 kb)	50 ng	-
DNA Insert (0.5 kb)	≈ 11 ng	-
T4 DNA Ligase	1 μL	5 U
Total Volume	20 μL	-

T4 DNA ligase was inactivated for 10 min at 65 °C and no further purification of DNA samples was performed before transformation.

2.6. BACTERIA CHEMICAL COMPETENT CELLS TRANSFORMATION

ER2738 chemically competent cells were transformed as described on CHAPTER 3, point 1.3.9.

2.7. DNA EXTRACTION AND CLONING CONFIRMATION

Following plating of the SOC outgrowth, 11 blue plaques present on X-Gal/IPTG plate were randomly picked to confirm the presence of NanoLuc™ gene. For that, the picked clones were amplified on liquid medium and the phage was purified. Next, the phage clones (100 μL) were treated with 10 μL Proteinase K [Fisher Scientific, BPI700-100] at a concentration of 0.5 mg/mL. This protease cleaves peptide bonds adjacent to the carboxylic group of aliphatic and aromatic amino acids, destroying the coat of the phage and making possible to access to the single stranded DNA. Proteinase K was left to act for 1 hour at 37 °C and then inactivated for 20 min at 100 °C.

After that, a PCR was performed (as described on point 1.3.7) but using different primers (table 18) and annealing temperature (45 °C).

Table 18 – Primers used to confirm the insertion on pVIII gene. M13KE_pVIII.FW – Primer Forward; NlucM13.RV – Primer Reverse. The melting temperatures were calculated using Modified Breslauer's thermodynamics, dH and dS parameters

Name	Length (bp)	Sequence (5' – end start)	Melting Temperature (TM)°C
M13KE_pVIII.FW	20	CTCTTTCGTTTTAGGTTGGT	47.7
M13KE_pVIII.RV	20	GAGCCTTTAATTGTATCGGT	47.7

In the case of a successful insertion, as confirmed by DNA electrophoresis, the DNA fragment created by PCR includes the flanking zone of pVIII gene plus the insert with 512 bp (Fig. 22), being that the resulting fragment should have 828 bp whereas the control (M13KE without insert) presents a band in the range of 300 bp.

Fig. 22 – Excision of M13KE genome map. In the boxes it is presented the Ph.D. Primer FW and RV used for PCR cloning confirmation. The zone in color corresponds to the cloning site for the construction of insertion of NanoLuc™ gene.

1. RANDOM PEPTIDE LIBRARIES CONSTRUCTION

Combinatorial peptide libraries can be constructed in one of three ways. In one method, an oligonucleotide, containing the mutation is annealed to sticky ends of a DNA vector, converted to double-stranded DNA *in vitro*, prior to introducing the DNA into eletrocompetent *E. coli*. In another method, one oligonucleotide is annealed to another, generating double-stranded DNA (after fill-in with DNA polymerase), which is digested with restriction endonuclease enzymes, resulting in insert fragments that are then linked with double-stranded M13 vector and inserted into eletrocompetent *E. coli* (Huang et al. 2011). A third widely used technique, so-called 'Kunkel mutagenesis', utilizes uracil-inserted, circular, single-stranded DNA as a template to synthesize double-stranded DNA *in vitro* with an oligonucleotide primer that brings in a mutation. Subsequently, dsDNA is introduced into bacteria and recombinant clones predominate due to cleavage of the uracilated strand *in vivo* (Huang et al. 2012).

Electroporation is the method of choice to generate large-scale random peptide libraries in *E. coli* as is creates libraries with 10^9-10^{11} members (Russel et al. 2004). This is a very critical step of the library construction, as each vector contains a different foreign sequence and the quality and diversity of the library is dependent on how many plasmids are incorporated into *E. coli*.

In this thesis, random peptide libraries were constructed adopting the second methodology above mentioned, where the specific primers designed for the construction of each library containing the random zones were annealed and filled-in with DNTs in a reaction catalyzed by Klenow Fragment exo-. After double digestion of the inserts and M13KE cloning vector the ligation was performed using T4 DNA ligase. The ligation was achieved and confirmed through PCR amplification using specific primers that linked to the flanking zones of M13KE cloning site.

Afterwards, the resulting recombinant M13KE vectors were inserted into chemically competent cells. The choice of using chemical competent *E. coli* was taken even thought that electroporation generally gives higher transformation efficiencies (measured in colonies formed per μg of DNA). Electroporation is highly sensitive to the presence of salt and DNA ligase in the DNA samples, and often requires the use of spin columns to clean up ligation reactions (NEW ENGLAND BIOLABS ® n.d.). The process of cleaning DNA samples usually undertakes a great

loss of DNA and therefore it can lead to the loss of the recombinant M13KE vectors, thus causing a reduction on the library diversity. Hence, transformation of chemically competent cells was preferred as it does not require cleaning up of the ligation mix.

1.1. M13KE VECTOR AND INSERTS CONSTRUCTION

Random peptide libraries were constructed using M13KE vector digested with NZYTech fast digest restriction enzymes Acc65I and EagI and self-made inserts that were constructed as above mentioned.

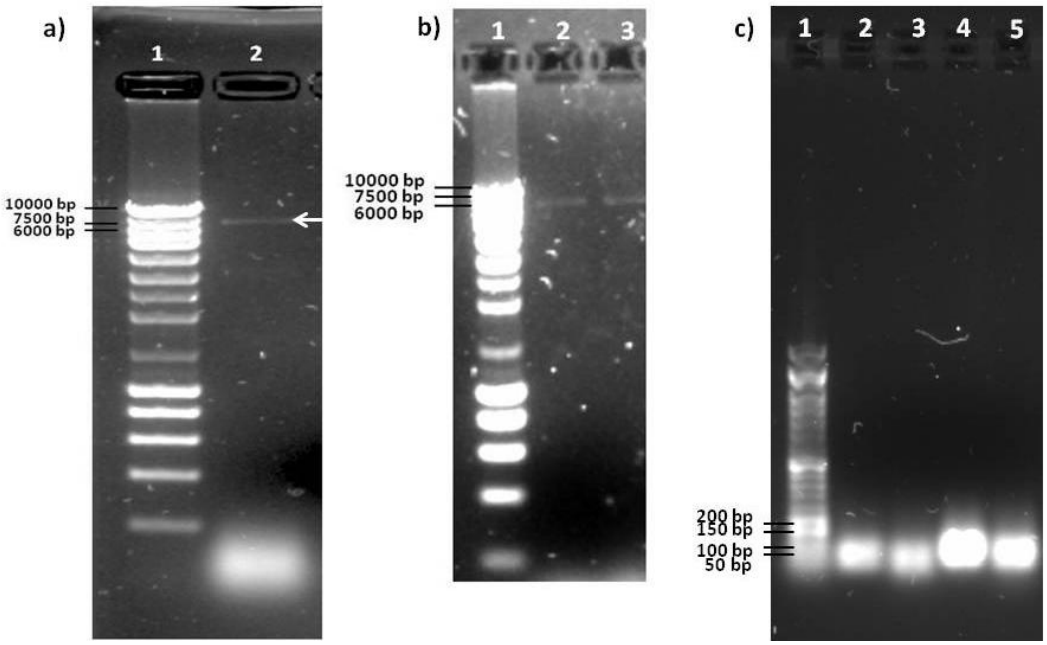

Fig. 23 – M13KE vector and inserts construction. M13KE digestion with HindIII (a), M13KE digestion with HindIII and double digestion with Acc65I and EagI (b) and 7-mer and 12-mer encoding fragments before and after double digestion with Acc65I and EagI (c), on 1% agarose gels. **a)** NZYTech ladder III (1) and M13KE linear vector digested with HindIII (2); **b)** NZYTech ladder III (1), M13KE linear vector digested with HindIII (2) and M13KE double digested vector with Acc65I and EagI (3); **c)** NZYTech ladder VI (1), 7-mer encoding sequence non-digested insert (2), 7-mer encoding sequence insert digested with Acc65I and EagI (3), 12-mer encoding sequence non-digested insert (4), 12-mer encoding sequence insert digested with restriction enzymes Acc65I and EagI (5).

Following extraction and purification of the DNA, it was necessary to confirm that these samples correspond to the M13KE vector without the contamination of bacterial DNA. This is a critical step of the cloning process because the smallest amount of contaminating bacterial DNA could lead to undesirable results (Brown 2010). The simplest way to confirm the presence of M13KE vector is to perform a diagnostic digestion test using a restriction enzyme that is known to cut the vector at a specific site, which can be analyzed through gel electrophoresis. As

explained on CHAPTER 3, point 1.3.2., the HindIII restriction enzyme was used for that purpose. This enzyme cuts the M13KE vector at position 6234 generating a linear fragment, correspondent to the total length of the vector - 7222 bp, Fig. 23 a) (see above). In this figure, it is possible to observe a faint, yet clear band of linearized M13KE right below the band of the ladder that represents the length of 7500 bp (white arrow in the figure). The smear that can be observed at the bottom of the gel corresponds to denatured RNA that was not effectively removed during the plasmid purification.

Next, in Fig. 23 b) it is presented an image of a 1% agarose gel that was used to confirm the correct double digestion of the M13KE vector (CHAPTER 3 point 1.3.3.) in order to proceed to the ligation step. The DNA ladder on lane 1 did not run correctly, as the bands are not clearly discriminated. Smeared bands can be caused by several reasons, one of them being an excess of DNA loaded into the gel, thus using a smaller amount of ladder would solve the problem. Nevertheless, comparing the bands on lanes 2 and 3, it is possible to take some conclusions. M13KE double digestion with Acc65I and EagI generates two DNA fragments, one of them with 7202 bp and the other with 20 bp. The bands on lane 2 and 3 containing M13KE plasmid digested with HindIII (7222 bp) and M13KE plasmid double digested with Acc65I and EagI (7202 bp), respectively, exhibit similar sizes as the gel resolution used did not allow a higher separation. Furthermore, the 20 bp fragment resulting from double digestion was so small that problaby migrate out of the gel during the electrophoresis, therefore not being observable on lane 3.

Random libraries inserts are demonstraded in Fig. 23 c) (see CHAPTER 3 point 1.3.4. for detailed information on the construction of the inserts), where undigested 7-mer and 12-mer encoding sequences are present, with 47 bp (lane 2) and 82 bp (lane 4), respectively. Observing the correspondent lanes on the image it is possible to see that the 47 bp insert is at the same position as the 50 bp band of NZYTech ladder VI, and also that 82 bp insert is slightly above and closer to the 100 bp band of the ladder. Lanes 3 and 5 correspond to the insert samples (7-mer and 12-mer respectively) following double digestion. After double digestion, the annealed duplexes lose approximately 10 bp what would not be perceptible using this type of gel. The noticeable differences between lanes 2-3 and 4-5 are due to a decrease in the amount of DNA that was loaded in the gel.

After this step it was possible to proceed to the ligation of the inserts into M13KE vector.

1.2. M13KE::INSERT(S) LIGATION – PCR CONFIRMATION

Libraries random peptides were constructed using the above products restriction sites. Acc65I and EagI leave complementary sticky ends that are linked on a reaction catalyzed by T4 DNA ligase.

Ligation of complementary sticky ends is very efficient as they can base pair with one another by hydrogen bonding forming a relatively stable structure for the enzyme to work on (Brown 2010). Also, using two restriction enzymes guarantees that the foreign DNA sequence is inserted on the right orientation i.e. in frame.

The main problem with cloning oligonucleotides into a DNA vector is the occurrence of multiple insertions. This is because small amounts (ng) of oligonucleotides inserts represent a large molar excess when compared to the vector amount, due to the small size of the insert as compared to the vector. Generally, to ensure optimal conditions for a single insertion, a molar excess of insert oligonucleotides from 5 to 20 fold comparing to the other fragments in the ligation mixture is recommended (Ausubel et al. 2003). For this cloning procedure, a molar ratio of 3:1 (insert: vector) proved to be efficient. Also, ligations using T4 DNA ligase can be performed with an incubation period of 1 to 16 hours at 12 °C to 30 °C. Ligation of sticky ends is usually carried out at 12 °C to 15 °C to sustain a good balance between annealing of the ends enzyme activity. Higher temperatures make it difficult for the ends to anneal, whereas lower temperatures cause a decrease on enzyme activity (Ausubel et al. 2003). In order to optimize the ligation conditions, two incubation temperatures, as well as two incubation times were tested (CHAPTER 3: point 1.3.6.).

The ligation reaction outcomes were analyzed before proceeding to bacterial transformation to guarantee the correct insertion of the oligonucleotides into the M13KE vector. For that purpose, PCR amplification was performed using specific primers that anneal to the flanking regions of the cloning site as explained on CHAPTER 3: point 4.2.7. PCR amplified fragments were analyzed on a 1% agarose and SGTB gel (CHAPTER 3: point 1.3.8.), regarding the smaller size of the amplified fragments. This time-saving buffer allows the use of high voltages when running the gel providing, at the same time, higher resolution and better separation of bands in the gel (GRiSP Team 2014). Gel electrophoresis results are presented on figure 24.

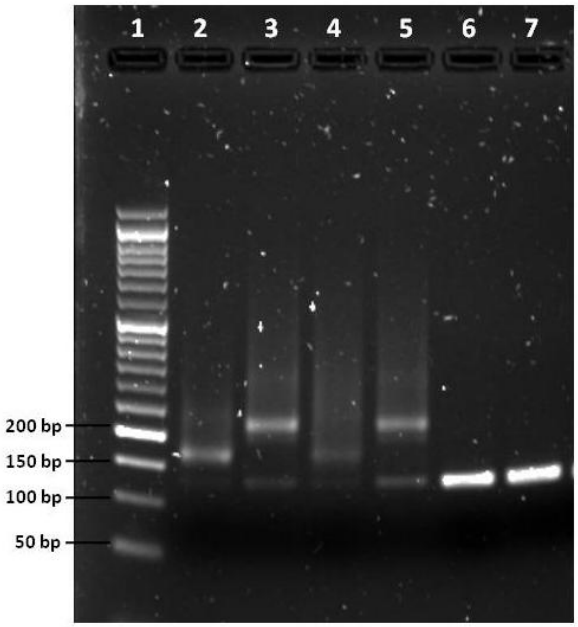

Fig. 24 – Gel electrophoresis results of 7-mer and 12-mer encoding sequences insertion into M13KE vector amplified using PCR, on 1% agarose and SGTB gel. NZYTech ladder VI (1); amplified DNA fragment of M13KE:: 7-mer insert, ligation conditions: 2 hours at 22 °C (2); amplified DNA fragment of M13KE:: 12-mer insert, ligation conditions: 2 hours at 22 °C (3); amplified DNA fragment of M13KE:: 7-mer insert, ligation conditions: overnight at 16 °C (4); amplified DNA fragment of M13KE:: 12-mer insert, ligation conditions: overnight at 16 °C (5); amplified DNA fragment of M13KE without insert, ligation conditions: 2 hours at 22 °C (6); amplified DNA fragment of M13KE without insert, ligation conditions: overnight at 16 °C (7).

The amplification should result in DNA fragments with different sizes, namely 132 bp for M13KE without insert (control); 148 bp for M13KE::7-mer insert and 186 bp for M13KE::12-mer insert.

The bands presented on lane 2 and 4, corresponding to the resulting amplification fragments for the ligation reactions of M13KE::7-mer insert exhibit the expected size being at the same position as the band corresponding to 150 bp of the DNA ladder VI. Similarly, on lane 3 and 5, the bands corresponding to the amplified fragments of the ligation reactions for M13KE::12-mer insert, present the correct length being approximately at the same position as the DNA ladder band corresponding to 200 bp. Finally, the two control bands (lanes 6 and 7) are presented between DNA ladder bands corresponding to 100 bp and 150 bp, as expected. These results prove that the insertions into M13KE vector were well-succeeded. Nevertheless, in lanes 2, 3, 4 and 5 it is possible to observe a smear that could be caused by the presence of free inserts on the ligation mix that linked nonspecifically to the primers. It is also possible to visualize a faint band at the same position of the control band meaning that circular M13KE without insert (undigested M13KE vector) was present on the ligation reaction product. Even though the

recommended double-digestion procedure was followed, the presence of undigested vector suggests that some error occurred that can be caused either by using very low units of enzyme (for the DNA template amount), short incubation times or also the presence of supercoiled DNA (NEW ENGLAND BIOLABS ® n.d.).

Analyzing figure 24 it is also possible to conclude that the overnight incubation at 16 °C (lanes 4 and 5) did not brought any improvements over the 2 hours incubation at 22 °C (lanes 2 and 3), therefore proving that the last conditions presented are optimized for the ligation reaction.

1.3. BACTERIAL CHEMICALLY COMPETENT CELLS TRANSFORMATION

After guaranteeing a successful cloning, it was possible to transform the chemically competent cells as explained on CHAPTER 3, point 1.3.9. After plating a volume of the SOC outgrowth, plates as the one presented on figure 26 were obtained.

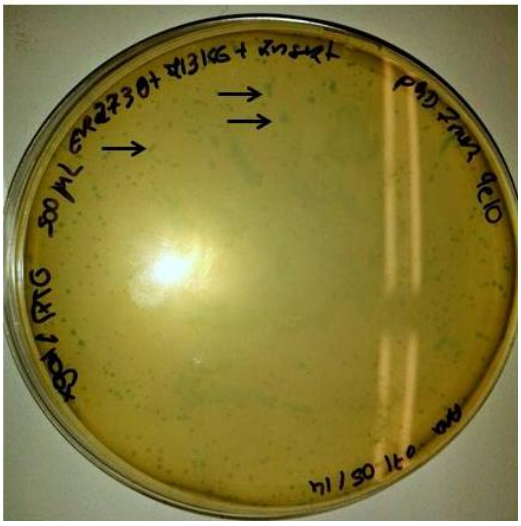

Fig. 25 – Outgrowth SOC plate containing X-Gal/IPTG with ER2738 chemically competent cells transformed with M13KE::7-mer library.

The blue plaques that are visible on Fig. 25 (e.g. pointed with black arrows) correspond to bacterial cells that were successfully transformed. Similar results were obtained for the transformation with M13KE::12-mer inserts (data not shown). Since the transformation was successfully achieved it was possible to proceed to the amplification of the libraries.

1.4. LIBRARIES AMPLIFICATION

The libraries amplification was performed to obtain the necessary titer to proceed to the *in vitro* biopanning. After amplification and purification of the phage particles, the titer assay was carried out and the results are presented on Fig. 26.

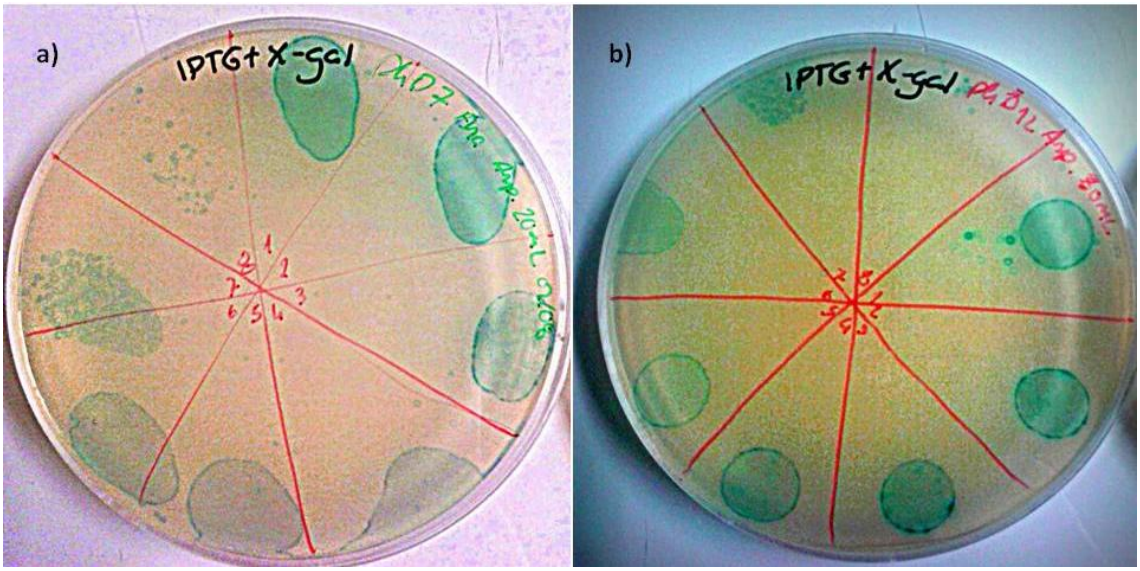

Fig. 26 – Libraries amplification and titer. **a)** Ph.D. 7 and **b)** Ph.D. 12 phage PFUs plate with 8 drops of successive dilutions of the amplified phage libraries.

The plates were used to determine the phage titer by counting the blue plaques. After the amplification, Ph.D. 7 and Ph.D. 12 were obtained with a concentration of $5.7X10^{11}$ pfu/mL and $7.5X10^{10}$ pfu/mL, respectively as it was possible to count 57 blue individual plaques at 10^{-8} for the first (Fig. 26 a), and 75 plaques at 10^{-7} for the second (Fig. 26 b). Phage titer was calculated using equation 5 (CHAPTER 3, point 1.2.3.).

The quality and size or diversity of phage display libraries are important features that guarantee a successful affinity selection. The quality of the phage library, meaning the fraction of the phage particles displaying the recombinant peptides out of the total phage pool, influences the efficiency and the outcome of subsequent biopanning steps as non-recombinant phages present growth advantages over recombinant phages, and so their presence can overwhelm the target-binding clones after a few amplification rounds (Derda et al. 2011; R. Huang et al. 2012). The presented methodology for the construction of phage libraries does not encompass a direct way of assessing if the chemically competent cells were transformed with wild-type M13KE (wtM13KE) or recombinant M13KE, since both form blue plaques in the presence of X-Gal/IPTG. Therefore, it was not possible to directly assess the quality of the library. To indirectly assess the

quality of the libraries it would be necessary to sequence all the clones which would be very time consuming and a huge economic burden. To overcome this limitation, there are currently described some enhanced protocols related to Kunkel mutagenesis method which present alternative approaches that enable the elimination of non-recombinant phages (Huang et al. 2012) or do not allow their assembly (Scholle et al. 2005).

The size of the library is related to the number of phages displaying different peptides. This important feature is closely correlated with the affinity of the isolated mutants since the probability of finding high affinity ligands is related with the number ligands present in the initial library. Also, the complexity of the library is important to have an unbiased screening of the target (Russel et al. 2004).

2. *IN VITRO* BIOPANNING EXPERIMENTS

Biopanning experiments were performed as presented on CHAPTER 3: point 1.4. Conventional and BRASIL biopanning were optimized to the final procedures that are above presented. As mentioned, the main difference between both procedures is the method used to separate the unbound phage from the ones that bound to RKO cells. However, in this thesis the optimization of both protocols led to other differences that should be taken into account, namely the number of RKO cells used as a target for the phage pools; the time and temperature of incubation with phage pools, and also the fact that for the BRASIL biopanning adaptation used there was no amplification of the phage pools between each round.

The amplification of the libraries between the biopanning rounds has been shown to decrease their diversity (Kuzmicheva et al. 2009; Derda et al. 2011), meaning that the selection of specific ligands from phage display libraries is driven by two processes: the panning step enriches the pool with clones that bind to the desired target but also by the amplification step (infection of bacteria by a single phage particle and the secretion of approximately 1,000 copies of phage) that enriches the pool with clones that have a growth advantage (Derda et al. 2011). This implies that ligands with similar or higher affinity to the target can be “lost” between the amplification steps, as they can have growth disadvantages and so, they will be present in lower numbers at the final amplified pool. As the amplified pool is used for the subsequent biopanning steps, the number of these phages will further decrease until they will eventually disappear from the phage pool. Eliminating amplification steps from biopanning experiments is a way of preventing the undesired loss of diversity during amplification, consequently enabling the

identification of a much broader repertoire of binding ligands (Derda et al. 2011). For this reason, the amplification step was eliminated from one of the selection methodologies aiming at the prevention of the loss of useful binding ligands.

2.1. CONVENTIONAL BIOPANNING

The detailed procedure used to perform conventional biopanning was presented on CHAPTER 3: point 1.4.1. For every biopanning round, the concentration of phage input and output was analyzed by counting phage-plate forming units after titer assay using equation 5. In Fig. 27, the concentrations of the phage pools before and after the biopanning step for Ph.D. C7C™ (a), Ph.D. 7 (b) and Ph.D. 12 (c) libraries, are illustrated.

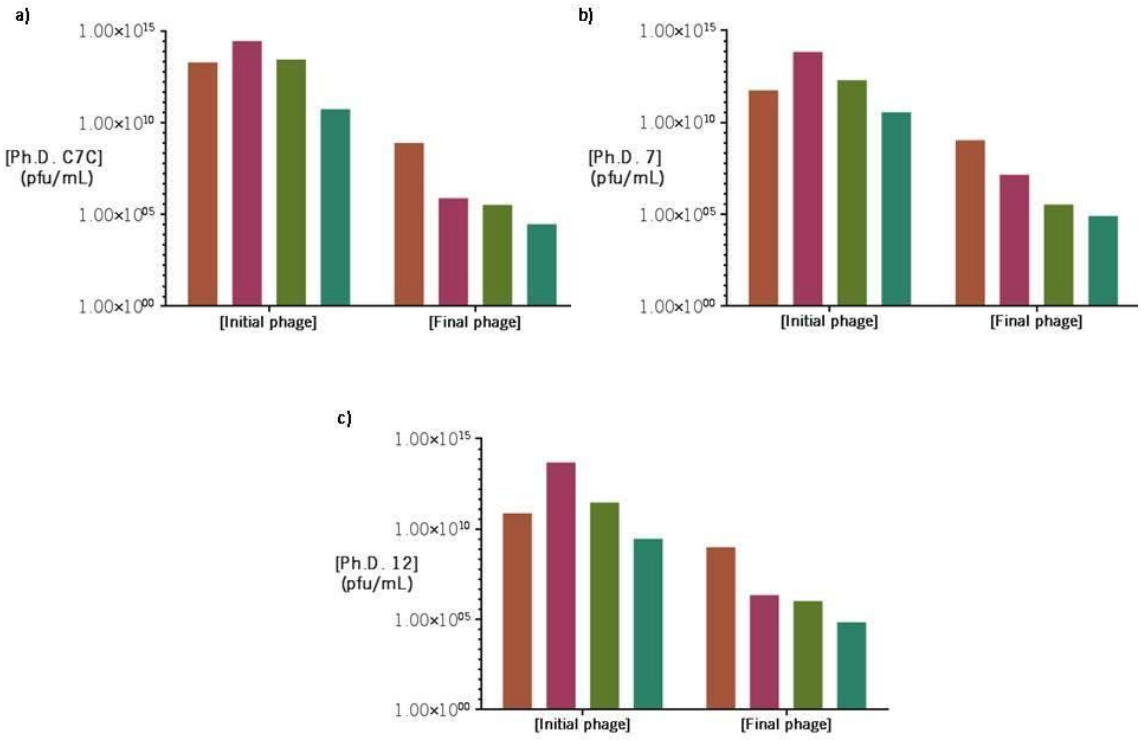

Fig. 27 – Initial phage input and final phage recover concentration, from conventional biopanning rounds, assessed by plaque-counting for **a)** Ph.D. C7C™, **b)** Ph.D. 7 and **c)** Ph.D. 12 libraries. Round 1 ( ); Round 2 ( ); Round 3 ( ); Round 4( ).

On each round of biopanning, the binding clones are enriched over the nonbinding clones. Typically, at least a 10-fold decrease of the phage number is expected between rounds of *in vitro* binding selection, but decreases can vary from 2-fold to more than a 1000-fold as the phage pool is enriched on the binding clones and 10^3 to 10^7 of nonbinding clones are washed out.

This enrichment process is typically evaluated to decide when to stop the selection and start the analysis of the clones (Russel et al. 2004). Usually, two to three (maxim four) rounds of selection against the target are recommended (Huang et al. 2011; Russel et al. 2004; NEW ENGLAND BIOLABS ® n.d.). Indeed, it is known that 3 to 4 rounds of selection provide ligands of the highest diversity and that adding additional rounds can only partially increase the affinity (Derda et al. 2011). From this point ahead, further rounds of amplification and biopanning will result only in the selection of phage that has a growth advantage over the library phage. For example, the smallest levels of contaminating environmental wild-type phage will completely overtake the pool if too many rounds of amplification are carried out, regardless of the strength of the *in vitro* selection (NEW ENGLAND BIOLABS ® n.d.). Concordantly, it is possible to see in Fig. 27 that the final concentration of phages after biopanning rounds 3 and 4 decreases immensely regardless of the amount of phage input. This is also demonstrated as an increase on the binding efficiency of the phage pools that accompanies the stringency of the rounds (increasing number of washes). Binding efficiency represents the ratio between the phage input and output (output/input) of every round. In Fig. 28, it is possible to observe that the binding efficiency increases notably from round 2 to rounds 3 and 4, for all libraries, thus meaning that the phage pools are, at this point, enriched with specific RKO ligands.

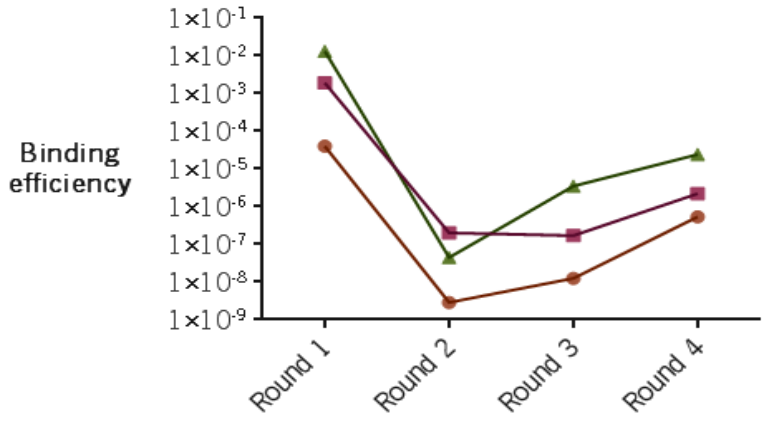

Fig. 28 – Binding efficiency of the conventional biopanning phage pools for the four rounds of affinity selection with Ph.D. C7C™ (—●—); Ph.D. 7 (—■—) and Ph.D. 12 (—▲—) libraries.

2.2. BRASIL BIOPANNING

The detailed procedure used to perform adapted BRASIL biopanning is presented on CHAPTER 3: point 1.4.2. Similarly to conventional biopanning, the concentration of phage input

and output was analyzed by counting phage-plate forming units after titer assay and using equation 5. In Fig. 29, the concentrations of the phage pools before and after the biopanning step for Ph.D. C7C™ (a), for Ph.D. 7 (b) and Ph.D. 12 (c) libraries, are presented. As mentioned, no amplification steps were performed in this methodology, hence meaning that the input of phage particles decreases exponentially since round 1.

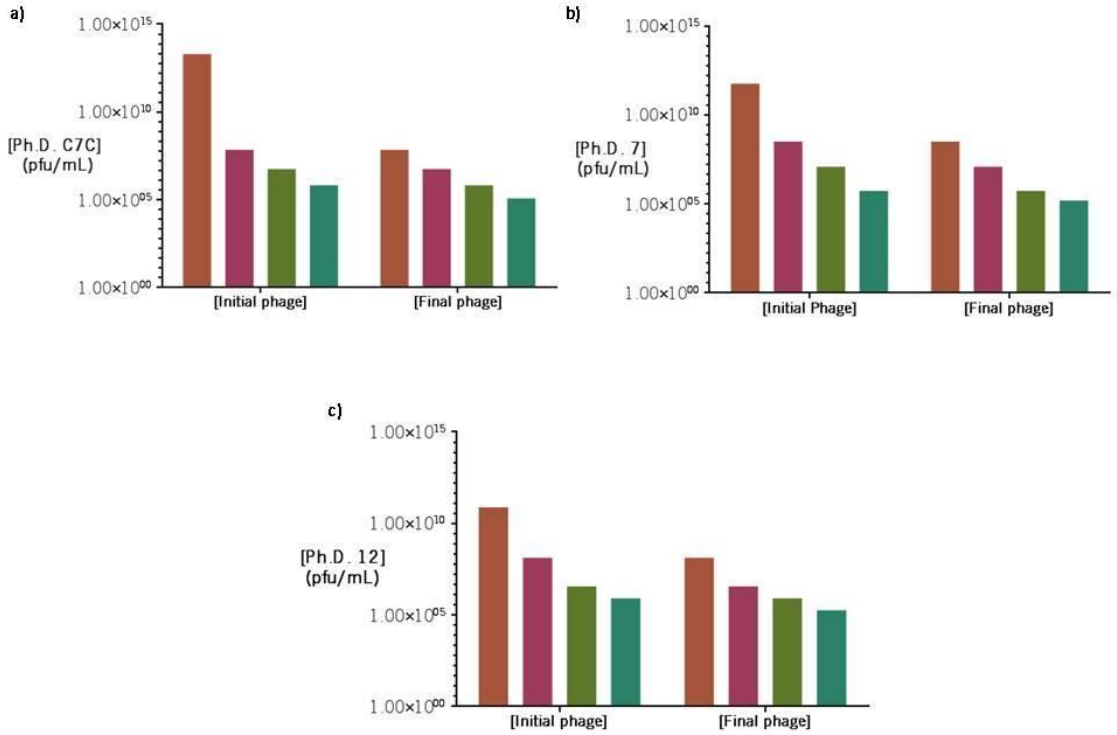

Fig. 29 – Initial phage input and final phage recover concentration, from BRASIL biopanning rounds, assessed by plaque-counting for **a)** Ph.D. C7C™, **b)** Ph.D. 7 and **c)** Ph.D. 12 libraries.
Round 1 ( ); Round 2 ( ); Round 3 ( ); Round 4( ).

From Fig. 29, the same conclusions drawn for conventional biopanning can be applied, i.e. there is a decrease on phage output on every round that accompanies the enrichment of the phage pool in specific RKO ligands. Comparing Fig. 27 and Fig. 29 it is possible to observe that, although the phage input for each round was much smaller the outputs were similar, in terms of number of phage particles. This is in agreement with the conclusions of Giordano and colleagues (2001), namely that BRASIL method allows the recovery of higher number of specific phage particles rather than conventional panning.

Regarding to binding efficiency (Fig. 30), the obtained results are comparable to those obtained for conventional biopanning (Fig. 28). Nevertheless, it is important to notice that there is

not a marked difference between rounds 2 and 3 as in conventional biopanning, instead a linear progression in the binding efficiency is observed. Contrarily to what was done in conventional biopanning in which the stringency of the selection was increased on every subsequent round (namely the increasing number of washes), in BRASIL adaptation, that stringency was maintained from the first to the last round. Therefore, the binding efficiency of the phage pools progressed linearly.

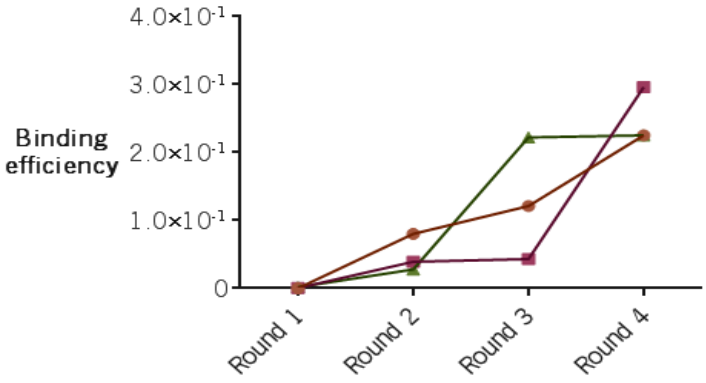

Fig. 30 – Binding efficiency of the BRASIL biopanning phage pools for the four rounds of affinity selection with Ph.D. C7C™ ( ); Ph.D. 7 ( ) and Ph.D. 12 ( ) libraries.

2.3. ELISA

The progress of biopanning through successive rounds is generally reflected in an increasing affinity of individual phage clones or of entire phage pools for the target receptor. The affinity of individual clones or entire phage pools can be assessed quantitatively by ELISA (Smith & Petrenko 1997).

In the present work, ELISA was performed in triplicates for the amplified phage pools recovered in each round of biopanning and using wild-type M13KE, i.e. with no displayed peptide as a baseline control (consult CHAPTER 3 point 1.4.3. for details). The ELISA method used for the detection of specific phage clones present on a given sample can be categorized as an indirect method since the antibody used detects the presence of M13KE phages that, when presenting specific ligands will bind to RKO cells. As it was explained above, the final signal is correspondent to the amount of phage particles that bound to the target giving quantitative information on the affinity of the phages. Thus, this method is a reliable and fast way of testing the phage clone or pools obtained through biopanning experiments.

Nevertheless, it is important to notice that, when biopanning against a polystyrene plate coated with the target, it is possible to unintentionally select peptides that specifically bind the polystyrene surface (Adey et al. 1995), the so-called plastic binders. Therefore, it is important to perform in parallel the ELISA assays with no target (negative control), to distinguish real target-binders from plastic binders, i.e. to remove the background noise.

Furthermore, it is necessary to compare ELISA results for the phage pools with the results for wild-type M13KE. It is known that bacteriophages have no intrinsic tropism for mammalian cell receptors (Yata et al. 2014), thus having to be genetically modified in order to present peptides that can give them that tropism. Accordingly, any signal obtained from ELISA using wild-type M13KE is a result of residual phage clones that were not effectively washed out during the washing steps. For this reason, M13KE ELISA results should serve as a baseline for all the ELISA results.

The results from ELISA assays performed against RKO cells using the amplified phage pools obtained through conventional biopanning and adapted BRASIL method with Ph.D. C7C™ (a), Ph.D. 7 (b) and Ph.D. 12 (c) libraries are presented in Fig. 31 and Fig. 32, respectively. For all, it is presented the raw abs mean values for the triplicate assays of ELISA with RKO cells and phage libraries. The background signal and also the specific signal of the phage pools that was calculated subtracting the background noise to the raw mean values are presented as well.

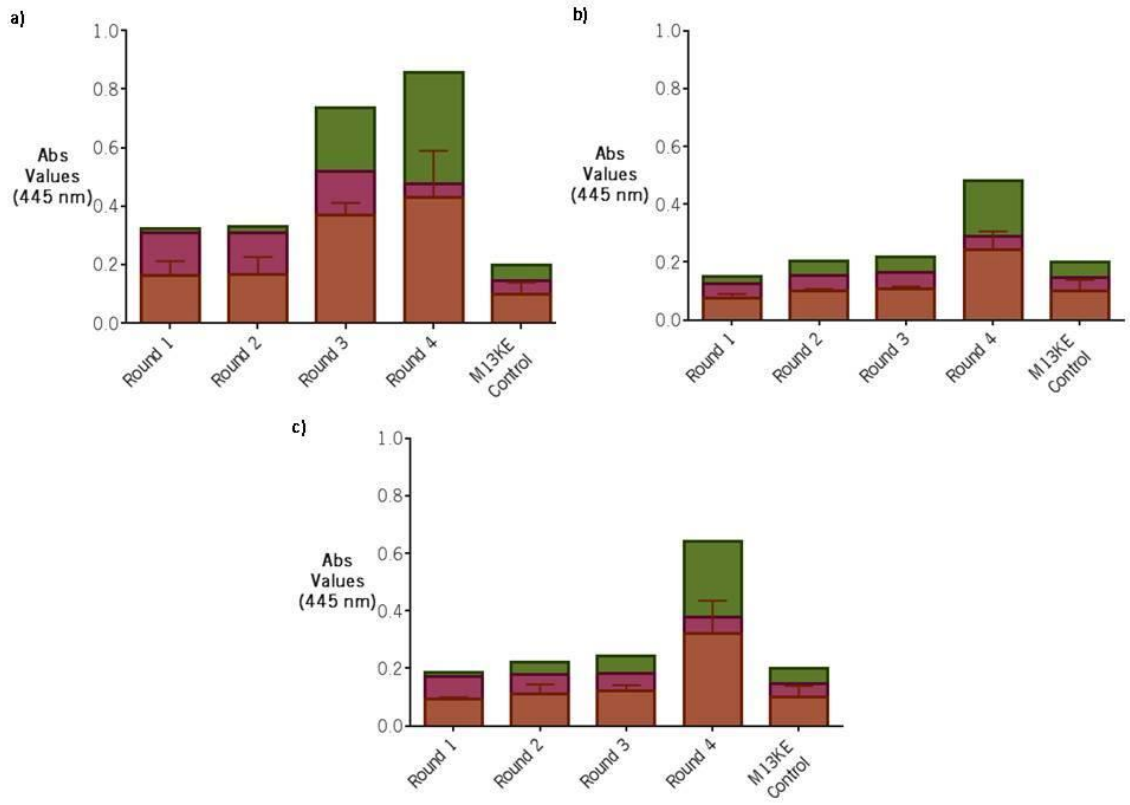

Fig. 31 – ELISA results for amplified phage pools obtained from four rounds of conventional biopanning with **a)** Ph.D. C7C™, **b)** Ph.D. 7 and **c)** Ph.D. 12 libraries. Mean abs raw values with standard deviation ( ); background signal ( ); specific signal of the phage pool ( ).

It is possible to observe, for every round and both selection methods, that from round 1 to round 4, there is an evident increase on the raw mean signal and decrease on the background noise, thus meaning that the overall specific signal that represents the presence of phages displaying RKO-specific peptides increases for each round. These results are concordant to those seen for the binding efficiency, and once again prove that the phage pools obtained after 4 rounds of biopanning are enriched with RKO specific phages.

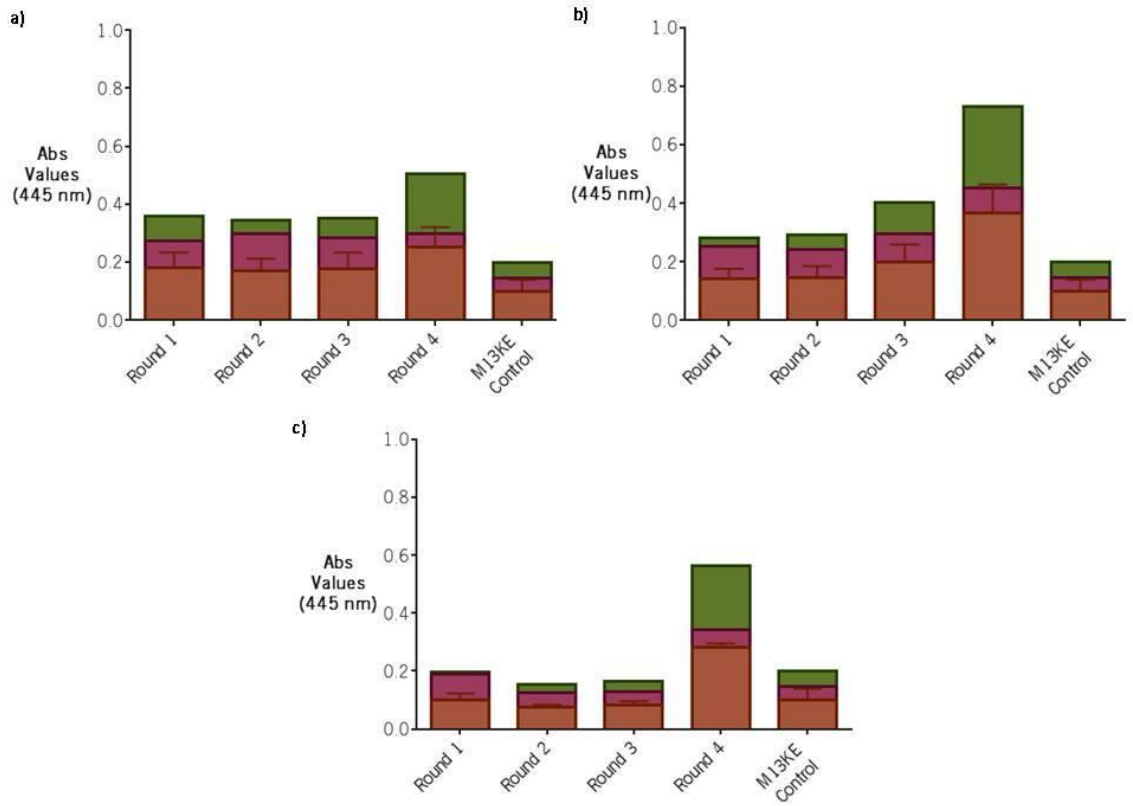

Fig. 32 – ELISA results for amplified phage pools obtained from four rounds of adapted BRASIL biopanning with **a)** Ph.D. C7C™, **b)** Ph.D. 7 and **c)** Ph.D. 12 libraries. Mean abs raw values with standard deviation ( ); background signal ( ); specific signal of the phage pool ( ).

The ELISAs were performed using purified amplified phage pools. The amplification steps result on different phage concentrations that were not normalized before proceeding to the assays. As a result, and in order to be possible to compare the results regardless of the differences on the phage input concentration, an affinity rate was calculated dividing the obtained specific signal value for the concentration of phage input. The results were then normalized using equation 7 presented below.

Normalized affinity rate $= \frac{X - X_{min}}{X_{max} - X_{min}}$, X represents affinity rate values. (7)

This simple method enables rescaling the range of features to make them independent of each other and it aims to scale the range in values between 0 and 1.

Normalized data enable the comparison of the affinity rate of the phage pools that result from each biopanning round. Fig. 33 presents the normalization of affinity rates corresponding to the four rounds, for the three libraries.

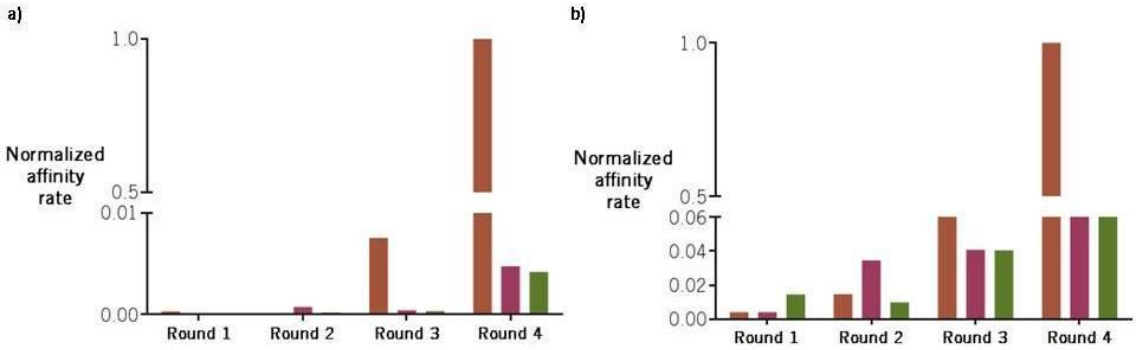

Fig. 33 – Normalized affinity rate of amplified phage libraries for the four rounds of selection with **a)** conventional biopanning and **b)** adapted BRASIL biopanning for Ph.D. C7C™ ( ), Ph.D. 7 ( ) and Ph.D. 12( ) libraries.

Fig. 33 a) that corresponds to the conventional biopanning shows that the best overall affinity rates were obtained for the Ph.D. C7C™ pools from rounds 3 and 4. For the other libraries, the only substantial results were obtained from round 4. Again, it is proved that the four rounds of biopanning enable the selection of RKO specific peptides.

For the BRASIL adaptation, the results are presented of Fig. 33 b). For this selection method it can be observed that the normalized affinity rates are much more homogeneous for the four rounds. This can be due to the fact that no amplification steps were undertaken and therefore, the number of phage particles displaying RKO-specific peptides did not increase exponentially, meaning that the enrichment of libraries was much slower thus enabling to see substantial results for all rounds using this normalization process. Again, the best affinity results were obtained at the final of the fourth round, for every library.

As it was proved that round four phage pools presented the best affinity rates for all the libraries and both selection methodologies, a new normalization was performed taking only these values into account. This normalization enabled the comparison of the three libraries (Fig. 34) and both methodologies (Fig. 35).

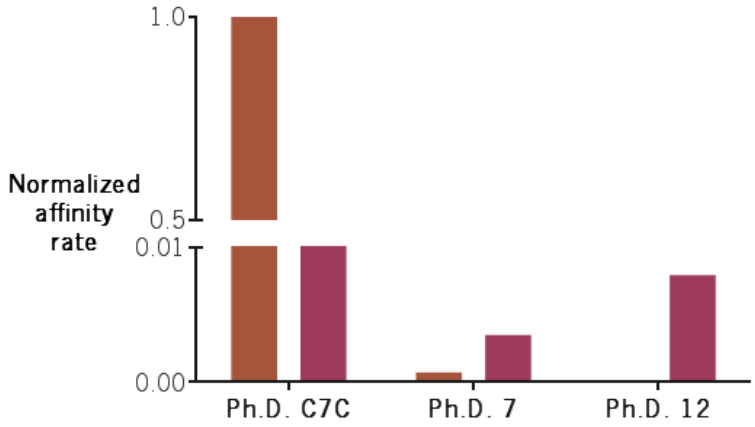

Fig. 34 – Comparison of binding affinity of the three Ph.D. libraries for the fourth round of conventional biopanning ( ) and adapted BRASIL biopanning ( ) using normalized values of affinity rates from the amplified phage pools retrieved from the fourth selection round.

In Fig. 34, we can observe that Ph.D. C7C™ presents the best affinity results for biopanning against RKO cells. In constrained phage libraries, constrains are artificially imposed on the peptide sequences to greatly reduce the range of conformations i.e. three-dimensional structures that are available. Generally a constrained library will represent a lot less three-dimensional shapes than a library of unconstrained (but in all other ways similar) peptides (Smith & Petrenko 1997). This restriction is *per se*, a limiting factor of the library diversity, and as a consequence, it decreases the probability that a clone will possess affinity for a given receptor. On the other hand, a constrained peptide whose accessible conformations happen to overlap extensively with active conformations can present far higher affinity than any unconstrained peptide (Smith & Petrenko 1997) .

Ph.D. 7 and 12 also present measurable affinity rates for BRASIL biopanning. This means that the final pools contained phage particles with displayed peptides of interest, but they could be at low number on the initial pool.

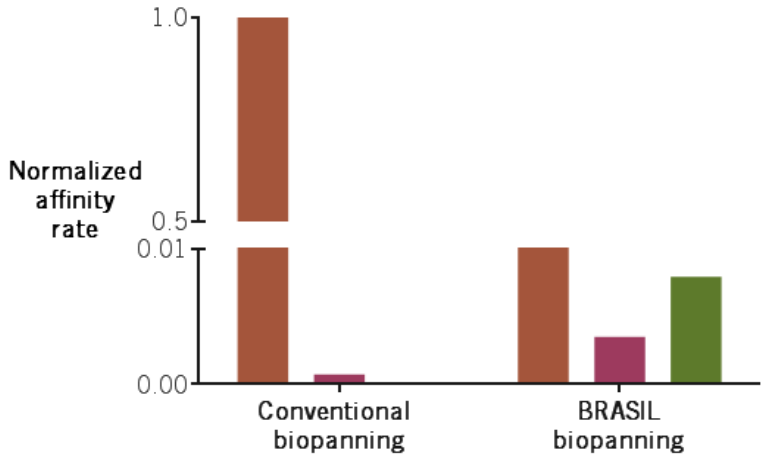

Fig. 35 – Comparison of binding affinity for convention biopanning and adapted BRASIL biopanning for Ph.D. C7C™ ( ), Ph.D. 7 ( ) and Ph.D. 12 ( ) libraries using normalized values of affinity rates from the amplified phage pools retrieved from the fourth selection round.

Fig. 35 enables the comparison of the affinity rates of the phage pools obtained after 4 rounds using conventional biopanning and BRASIL biopanning adaptation. Conventional biopanning presents very low affinity rates for the self-made libraries when comparing to BRASIL. Amplification rounds that were undertaken between each round of conventional biopanning are probably the reason behind this. As seen in CHAPTER 4, point 1.2., the final constructed libraries presented a low fraction of M13KE wild-type (see Fig. 26) that have a growth advantage over the mutated phages. Many biological reasons have been pointed to explain growth advantages between wild-type M13KE and even between phages displaying different peptides. Those include the binding to pili, use of rare codons, interference with packing or infection and rare mutations in the regulatory regions of phage genes (Derda et al. 2011). Amplification steps enriched the library on M13KE wild-type meaning that after four rounds of biopanning the pools of Ph.D. 7 and 12 were mostly composed by these non-mutated phages. This means that even if phages displaying specific peptides were presented on the final phage pool, they were in such a low number that the signal obtained on ELISA was very low and therefore not perceptible.

For Ph.D. C7C ™, the results are much higher for conventional biopanning than for BRASIL. In conventional biopanning, two forces are driving the enrichment of the phage pool on specific ligands, namely the growth advantages of some phages over the others and the stringency of the washes on biopanning rounds. Subsequently, a mutated phage that presents high affinity for the target and also growth advantages over other phages in the pool will be present in a greater number on the final phage pool, therefore retrieving a high signal on ELISA.

It is important to notice that it was not possible to use titer and phage-counting assays to distinguish the wtM13KE from genetically modified phages because both form blue plaques in the presence of X-Gal/IPTG. For this reason, it is not known what fraction of the pools that were used to incubate with RKO cells to perform the ELISA assays actually correspond to the wtM13KE and what is corresponding correspondent to the M13KE with displayed peptides. Nevertheless, the results presented of figures 31 and 32 prove without doubt that the phage pools retrieved from biopanning experiments contained a fraction of phages presenting peptides with affinity for RKO cells.

ELISA results showed higher affinities for the phage pools retrieved from the fourth round of biopanning for both methodologies and the three libraries used. Therefore, twelve clones from those pools were randomly picked for sequencing.

2.4. CLONES SEQUENCING

The final step used to characterize specific ligands is to identify the sequence of the peptides which interact with the target as explained on CHAPTER 3, point 1.4.4.

As previously mentioned, the clones chosen for sequencing were randomly picked from the phage pools retrieved from the fourth round of biopanning. The results from sequencing are presented on table 19.

Table 19 – Sequencing results of 12 random selected clones after 4 rounds of selection with RKO colon cancer cell line (Acc65I restriction site (GGTACC); EagI restriction site (CGGCCG); bold zone corresponds to the peptide codifying sequence)

Selection Method	Ph.D. library	Specific Peptide ID	Sequences (random region) 5' end start
Conventional biopanning	C7C	#1	GGTACCTTTCTATTCTCACTCT**GCTTGTATTGGTAATTCTAATACTCTGTGCG GTGGAGGT**TCGGCCG
		#2	GGTACCTTTCTATTCTCACTCGGCCG
	7	#3	GGTACCTTTCTATTCTCACTCGGCCG
		#4	GGTACCTTTCTATTCTCACTCGGCCG
	12	#5	GTAACCTTTCTATTCTCACTCGGCCG
		#6	GGTACCTTTCTATTCTCACTCGGCCG
BRASIL biopanning	C7C	#7	GGTACCTCTCTACTCTCACTCGGCCG
		#8	GGTACCTCTCTATTCTCACTCGGCCG
	7	#9	GGTACCTTTCTATTCTCACTCGGCCG
		#10	GGTACCTTTCTATTCTCACTCGGCCG
	12	#11	GGTACCTTTCTATTCTCACTCGGCCG
		#12	GGTACCTTTCTATTCTCACTCGGCCG
wtM13KE			GGTACCTTTCTATTCTCACTCGGCCG

Only one of the twelve clones sent for sequencing gave a positive result. The other eleven clones chosen were found to be wild-type M13KE.

For self-made Ph.D. libraries it was already explained that, the amplification of the phage pools could have enhanced the propagation of wild-type M13KE as it has growth advantages over the phages displaying foreign peptides due to N-terminal fusions to pIII (which mediates infectivity by binding to the F-pilus of the recipient bacterium). Therefore, it is possible to conclude that the phage pools presented a fraction of wtM13KE, and consequently lowered the probability of picking a positive clone. Adding to this, performing the biopanning experiments in parallel on the same plate for all libraries could have caused some cross contaminations. Cross contamination explains the appearance of wtM13KE on the phage pools using commercial library. Wild-type M13KE present in less than one part per billion can completely overtake the pool if too many rounds of amplification are carried out (NEW ENGLAND BIOLABS ® n.d.).

Notwithstanding, the clone picked from phage pool recovered from conventional panning with the commercial library was positive, meaning that it was found a peptide with high affinity towards RKO cells (Fig. 34 and Fig. 35).

The high-affinity peptide CIGNSNTLC was run on MimoDB (Huang et al. 2012) and a 55% homology was found to a peptide obtained through biopanning experiments using Ph.D. C7C library against highly tumorigenic colon carcinoma cells HT29 (Kelly & Jones 2003). This homology suggests that the peptides are recognizing the same protein at the surface of both cell lines and consequently that this protein can be a possible cancer biomarker for CRC.

Biopanning experiments resulted on the discovery of a new peptide with high affinity for RKO colon cancer cell line as proved by ELISA assays. However, it is not possible to conclude that the peptide found shows specificity for these cells and that it can be used for diagnostic purposes. In order to prove that it will be necessary to perform biopanning experiments using normal colon endothelial cells to assess if the peptide also displays affinity against them. A peptide exhibiting affinity for cancer cells but not for normal cells holds an enormous potential to be used as a diagnostic tool. Colon cancer–binding peptides can be used as molecular probes for *in vivo* or *in vitro* imaging of early neoplastic disease, leading to early diagnosis of colon cancer and, therefore improved patient survival (Kelly & Jones 2003).

3. NANOLUC™ GENE INSERTION ON M13KE

NanoLuc™ was inserted on M13KE aiming at conferring to the phage the capacity to produce bioluminescence in the presence of a specific substrate, therefore creating a novel reporter phage that when displaying, simultaneously, a specific peptide with high affinity for cancer cells could be used as a diagnostic probe for colon cancer. Hence, using direct cloning, this gene was introduced on the proximity of pVIII gene of M13KE vector; on an attempt to create a peptide displaying ≈2700 copies of NanoLuc™ Luciferase (see CHAPTER 2, point 4). The decision of coupling this protein to pVIII was taken because of the high number of NanoLuc™ Luciferase proteins that a successful cloning would able to present on the surface of the phage; a number that would enable a high enzyme activity and therefore a great response on the presence of the specific substrate. Finally, this would result in a phage capable of producing enough signals that would allow its detection at the naked eye.

In spite of the advantages of having this gene coupled to pVIII protein of M13KE, it is known from the literature that only short peptide sequences, with eight residues at maximum, can be displayed on every copy of pVIII protein as larger peptides interfere with the coat protein function in viral packaging, and even this smaller peptides involve a substantial fraction of the virion mass which can dramatically alter its physical and biological properties (Russel et al. 2004; Pande et al. 2010; Smith et al. 1994). Normally, to overcome this limitation, a hybrid virion is produced, in which the foreign peptide is displayed on only a fraction of the endogenous coat protein. This can be achieved with two types of phage-display systems normally referred as 88 and 8+8 vectors. In a type 88 vector, the phage genome contains two pVIII genes, encoding two different types of pVIII molecule; one displaying the foreign peptide and the other wild-type. The resulting phage is a mosaic that presents both wild-type and recombinant pVIII proteins in general in lower number. The other type, 8+8 vectors, differ from 88 type as non-modified and recombinant pVIII genes are presented on different genomes (Smith & Petrenko 1997). The viral gene encoding the fusion coat protein is carried on a phagemid, which is a plasmid that also has the phage origin of replication and phage packaging signal, whereas the wild-type genome is carried on a phage that provides all other genes required for phage assembly with a deficient phage packaging signal generally referred as helper phage. Co-infection of the bacterial host cell by phagemid and helper phage, yields the production of hybrid phage particle displaying only few copies of the fusion coat protein and carrying the phagemid genome (Pande et al. 2010).

Both options will result on much fewer copies of fusion proteins on the coat of the phage, what would immensely decrease its response to the presence of the specific substrate, hence limiting its use as a reporter phage. For this reason, the direct cloning of NanoLuc™ gene on M13KE genome was attempted, through the methodology explained on CHAPTER 3, point 2.

Firstly, NanoLuc™ gene from pNL1.1 [*Nluc*] vector was amplified to insert the specific restriction sites for the enzymes that were selected for M13KE double digestion, and that will enable the insert-vector ligation. The amplification was verified by gel electrophoresis and the results are presented on Fig. 36. It is possible to see that all amplified samples exhibit the correct length of approximately 512 pb (the nucleotides added for the restriction sizes can be despised).

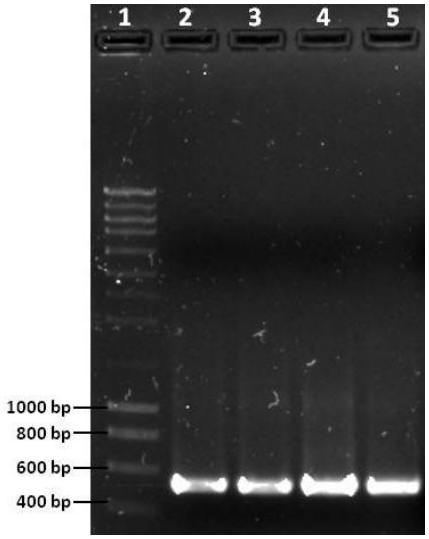

Fig. 36 – Gel electrophoresis results of NanoLuc™ gene amplification through PCR on a 1% agarose gel: NZYTech ladder III (1); amplified NanoLuc™ gene (2-5).

Afterwards, the samples containing the insert were loaded on the agarose gel and were further extracted and cleaned-up as explained in CHAPTER 3, point 2.2. Double digestion of M13KE RF with SnabI and BspHI produces a linear M13KE vector with 7217 bp. In Fig. 37, the results from gel electrophoresis that empowered the confirmation of M13KE digestion and the recovery of the purified insert are presented.

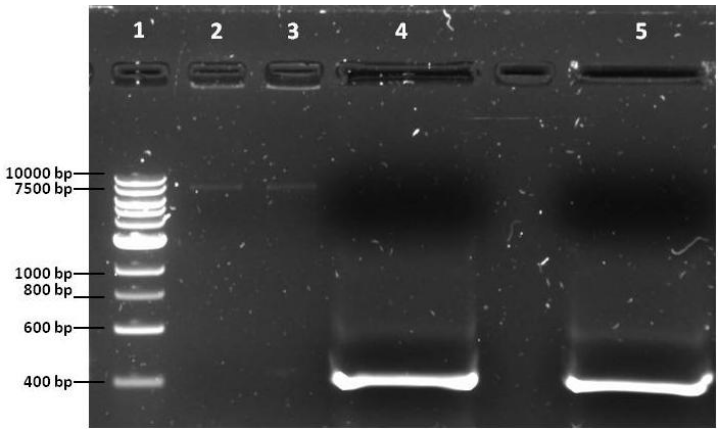

Fig. 37 – Gel electrophoresis results of M13KE digestion with HindIII and double digestion with SnabI and BspHI and amplified NanoLuc™ on a 1% agarose gel: NZYTech ladder III (1); amplified NanoLuc™ gene (2-5). NZYTech ladder III (1); linear M13KE digested with HindIII (2); linear M13KE double digested with SnabI and BspHI (3); amplified NanoLuc™ gene samples (4,5).

Subsequently, it was possible to proceed to the ligation step. On all cloning projects, ligation is critical, but in this particular case, the ligation is especially difficult since SnaBI cleaves the DNA fragments leaving blunt ends. Ligation of DNA molecules with blunt ends is a much less efficient process than sticky-ended ligations (Wilson & Walker 2010). Thus, it is recommended that blunt end ligation reaction to be performed at high DNA concentrations (with a ratio of insert molar excess of 10:1 or higher), to increase the chances of the ends of the molecules coming together in the correct way (Brown 2010). In this work, a ratio of insert molar excess of 3:1 was used. Using higher amounts of insert was not possible as the process of recovering DNA from agarose gels is very ineffective and lead to the loss of a big amount of amplified insert. Despite this, the ligation was performed and the resulting DNA was transformed into chemically competent cells. The resulting SOC outgrowth was plated and 11 clones were selected to confirm the cloning process using DNA electrophoresis (Fig. 38).

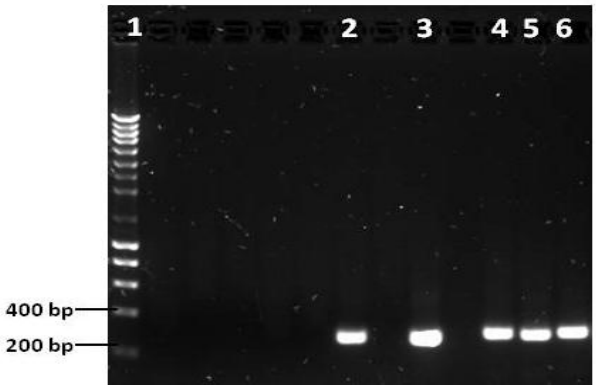

Fig. 38 – Gel electrophoresis results cloning confirmation using PCR amplification on a 1% agarose and SGTB gel: NZYTech ladder III (1); amplified fragments from the picked clones (2-5); amplified fragments from M13KE control (6).

The PCR amplification was not accomplished for all of the clones as it is possible to observe on the figure, probably because the DNA templates were degraded. However, comparing the length of the DNA fragments in lane 2-5, corresponding to four clones resulting from the SOC outgrowth, to the fragment resulting of M13KE without insert PCR amplification (presented on lane 6) it is possible to conclude that none of the clones present the NanoLuc™ gene. There are two possible explanations for this, the ligation was not achieved or contrarily, it was, but there was no phage assembly due to the large length of the insert.

As mentioned in CHAPTER 2, point 4, the construction of reporter phages requires labor-intensive and detailed genetic knowledge of the phage. Besides, it is difficult, as the volume of the phage coat sets a natural limit to the amount of genetic material that can be introduced into the phage genome. In spite of this, it is important to continue the research on this field as the value of the reporter phages that could be used to cancer detection in a fast and reliable way is undeniable.

CHAPTER 5: CONCLUSIONS AND FUTURE PERSPECTIVES

1. CONCLUSIONS

The final goal of this thesis was to construct a phage-based nanoparticle that would serve as an interface on a biosensor, capable of recognizing colon cancer cells and reporting their presence by emitting a bioluminescent signal.

This phage bio-receptor was designed to recognize RKO cell line by displaying a specific peptide on its coat, coupled to pIII minor coat protein. The search for a specific peptide was performed using two phage display approaches and three types of phage displayed random peptide libraries, a constrained commercial library, Ph.D. C7C™ and two self-made linear libraries Ph.D. 7 and Ph.D. 12. The last two were generated by introducing randomly encoding sequences into M13KE cloning vector. Those sequences were synthesized using specific primers containing random zones that when annealed and filled up with nucleotides (in a reaction catalyzed by a DNA polymerase) created a set of double stranded fragments with a variety of encoding sequences. The inserts were successfully constructed, and afterwards linked to Acc65I/EagI digested M13KE vector. The introduction of these encoding sequences into M13KE vector and further transformation of chemically competent ER2738 cells generated a pool of phages displaying a miscellany of linear peptide sequences containing seven (Ph.D. 7) or twelve (Ph.D. 12) peptides. The construction of the libraries was achieved. However, they contained a fraction of wild-type M13KE that resulted from an incomplete digestion of the double stranded M13KE samples.

All three libraries were used to screen the surface of RKO colon cancer cell line in order to find a specific peptide for one RKO membrane protein through two biopanning approaches: conventional and BRASIL adaptation that differ on the methodology used to divide the phage particles that bound to RKO cells from the unbounded. Also, on the first methodology, amplification rounds of the retrieved phage pools were performed between every round of biopanning.

Binding efficiency of the phage pools was evaluated through plaques counting of the phage input and output of every biopanning round. It was proved that BRASIL adaptation allowed a higher recovery of bound phages rather than conventional biopanning, and also that binding

efficiency increased during the subsequent rounds of biopanning meaning that more phages were recovered in every round. These results allowed to infer that more phages displaying effective ligands were present on every subsequent biopanning round, resulting from an increase on the strength of the *in vitro* binding selection.

The affinity of the resulting phage pools retrieved from biopanning rounds was evaluated through ELISA against RKO cells. As a result, it was proved that for both methodologies and for the three libraries, the affinity increased on every successive round proving that the pools were being enriched on phages displaying specific peptides with affinity towards RKO cells. ELISA results also proved the influence of the amplification steps. On conventional biopanning, amplification rounds result on an enrichment of the phage pools on specific phages for the commercial library, as the best affinity rate achieved was referent to this specific library and proved to be greater than the one achieved for the same library on the BRASIL adaptation. Regarding to the self-made libraries, the amplification steps proved to be a drawback once they enriched the pools on wtM13KE which have growth advantage over the mutated phages. Amplification of the phage libraries can cause that the majority of the phage pools after three rounds of biopanning are overtaken by wtM13KE phages, even with residual levels of contamination, as the display of foreign peptides as N-terminal fusions to the infectivity protein pIII, as in the Ph.D. libraries, slightly attenuates infectivity of the library phage relative to wtM13KE. This was the reason for the small affinity rate values presented for the self-made libraries with conventional biopanning – phage pools presented a substantial fraction of wtM13KE that do not have affinity towards RKO cells. For BRASIL biopanning, the effects of the amplification were less notable, as amplification between rounds was not performed. Nevertheless, the phage pools were amplified before proceeding to ELISAs and therefore, the presence of wtM13KE on the phage pools was also noticeable on the affinity values as they were lower than the ones obtained for the commercial library.

The presence of wtM13KE on the phage pools was confirmed since only one of the twelve randomly picked clones sequenced presented an encoding sequence of a high-affinity peptide. As a result, it was found one peptide with high affinity for RKO cells recovered from the fourth round of conventional biopanning. This peptide, CIGNSNTLC, presented a 55% homology with a peptide obtained through biopanning against highly tumorogenic colon cancer cells. Further biopanning rounds against normal colon endothelium cells are needed to assess if the

peptide found exhibits or not affinity towards those cells, therefore proving its value as a diagnostic tool.

The main issue of the current work was the quality of the constructed libraries, given the presence of wtM13KE on the initial library. Thus, using a methodology to enable the elimination of non-recombinant phages would greatly improve biopanning results. Also, it was herein proved that amplification between rounds is not imperative, and could therefore be eliminated in order to maintain the quality of the libraries, and also to prevent the loss of useful ligands with lower growth advantages.

The second goal of this thesis was the insertion of NanoLuc™ gene into the M13KE vector to confer to the phage the capacity of producing a bioluminescent response in the presence of a specific substrate. Firstly, NanoLuc™ gene was successfully amplified from pNL1.1 vector, and the restriction sites that would allow the ligation to the zone of interest on M13KE were inserted on this process. Next, the ligation of the SnabI/BspHI restricted insert and vector was attempted but this was not successfully achieved, what could have been caused by the unfavorable ligation conditions or by the fact that the phage coat could not encompass the addition of 2700 copies of NanoLuc™ luciferase coupled to pVIII protein, directly involved on the phage assembly process. In the future, other approaches to genetically engineer M13KE can be attempted, as well as the construction of a phagemid vector that will enable the display of this enzyme coupled to a fraction of pVIII proteins, what could enable the assembly of the phage. Chemical conjugation of this enzyme to M13 phage could also be attempted, as it showed results for coupling drugs as DOX for target drug delivery (Ghosh et al. 2012) and fluorescent molecules that allowed the use of this phage in cancer cell imaging (Li et al. 2010).

The construction of biosensors using phage-based nanoparticles as an interface to recognize and report the presence of cancer cells or other tumor markers on a given sample can be a major advance on the cancer diagnosis field.

The use of peptides as ligands has many advantages over the use of antibodies, especially for cell-based recognition as the target protein structure is unknown. Also, they are easy and cost-effective to synthesize on a large scale and are not immunogenic, two very important characteristics for its clinical use on general population. Using phage displayed random peptide libraries to screen cells surfaces or known protein biomarkers through the use of biopanning methodology presents as an easy and reliable way of finding high affinity and specific

peptides to a variety of targets. Therefore, the technology herein presented constitutes a noteworthy pathway for finding newly peptides that could be used to target tumor markers and serve as cancer diagnostic probes. In parallel, the construction of a reporter phage capable of emitting visible bioluminescence would enable the use of these viral particles in simple biosensor equipments.

Resuming, there is a need to continue the research on phage display, biopanning and reporter phages construction since the merge of these areas could represent a considerable breakthrough on the use of phages and peptides as molecular recognizing and reporting elements on a biosensor with the prospect of taking cancer diagnosis to a molecular level.

2. FUTURE PERSPECTIVES

During this thesis an assay was conceptualized that could merge the fields of ligands and biomarkers search, but it was not tested due to time limitations. The idea was to use the phages displaying specific peptides towards colon cancer cells obtained with biopanning rounds to incubate them with the plasma membrane proteins fraction of those cells. During this thesis, the fractionation of RKO cells to obtain the plasma membrane fraction was attempted using a methodology presented by Rockstroh and colleagues (method 2 presented on (Rockstroh et al. 2011)), however the protein concentration obtained was not enough to proceed with further experiments.

Those peptides, coupled to phage particles that during biopanning linked to an unknown protein of the cell surface, would now link to the same protein of the plasma membrane fraction. Afterwards, this mixture of phage particles and membrane proteins would be separated using cesium chloride gradient (note that the specific protein for the peptide found would be at this point linked to the phage particles). Equilibrium centrifugation in pre-formed cesium chloride gradients is already a technique used to purify and concentrate phage particles (Bachrach & Friedmann 1971), and in this particular case would allow the separation of the phage-proteins complexes from other phages and proteins.

When purified, the phage-protein complexes would be run through a 2D-PAGE. This method allows comparative studies of different samples, such as normal versus diseased, or treated versus untreated, providing information on the expressional differences at the individual protein or protein group level, which can be assumed as responsible for phenotype changes. In this specific case, 2D-PAGE would be used to compare the profile of the phage proteome to the profile of the phage-protein complex, hopefully to find the spot that is due to the cancer cell protein linked to the phage particle.

Using 2D-PAGE coupled with other downstream processes such as Western blotting and/or mass spectrometry, it is possible to identify protein tumor markers in several types of samples, such as blood, serum tissues, among others (Even-Desrumeaux et al. 2011; Wulfkuhle et al. 2003). As a result, coupling these technologies would eventually enable the identification of the protein that specifically binds to the ligand and that could potentially be considered a cancer biomarker.

REFERENCES

Adams, M.H., 1959. *Bacteriophages*, New York: Interscience Publishers, Inc.

Adey, N.B. et al., 1995. Characterization of phage that bind plastic from phage-displayed random peptide libraries. *Gene*, 156(1), pp.27–31.

Alberici, P., 2007. *The Adenoma-Carcinoma Sequence in Colorectal Cancer: scratching the surface*. Erasmus Universiteit Rotterdam.

Alirezapour, B. et al., 2013. Production and characterization of recombinant scFv against digoxin by phage display technology. *Monoclonal antibodies in immunodiagnosis and immunotherapy*, 32(3), pp.172–179.

American Cancer Society, 2014. Cancer Facts & Figures 2014. *Atlanta: American Cancer Society*.

Arap, 2005. Phage display technology - Applications and innovations. *Genetics and Molecular Biology*, 9, pp.1–9.

Arap, W., Pasqualini, R. & Ruoslahti, E., 1998a. Cancer treatment by targeted drug delivery to tumor vasculature in a mouse model. *Science*, 279(5349), pp.377–380.

Arap, W., Pasqualini, R. & Ruoslahti, E., 1998b. Chemotherapy targeted to tumor vasculature. *Current opinion in oncology*, 10(6), pp.560–565.

Armaghany, T. et al., 2012. Genetic alterations in colorectal cancer. *Gastrointestinal cancer research : GCR*, 5(1), pp.19–27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24603996.

Atcc, 2012. ATCC ® ANIMAL CELL CULTURE GUIDE tips and techniques for continuous cell lines. *Order A Journal On The Theory Of Ordered Sets And Its Applications*.

Ausubel, F.M. et al., 2003. *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.

Bachrach, U. & Friedmann, A., 1971. Practical Procedures for the Purification of Bacterial Viruses. *Applied Microbiology*, 22(4), pp.706–715. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC376390/.

Bakhshinejad, B., Karimi, M. & Sadeghizadeh, M., 2014. Bacteriophages and medical oncology: targeted gene therapy of cancer. *Medical oncology (Northwood, London, England)*, 31(8), p.110. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25012686 [Accessed July 31, 2014].

Barry, M.A., Dower, W.J. & Johnston, S.A., 1996. Toward cell-targeting gene therapy vectors: Selection of cell-binding peptides from random peptide-presenting phage libraries. *Nature Medicine*, 2, pp.299–305.

Bates, J., 2008. Cancer biomarkers: adoption is driving growth. *Insight Pharma Reports, Needham, MA, USA*.

Bélizaire, A.-K. et al., 2003. Identification of a murine ICAM-1-specific peptide by subtractive phage library selection on cells. *Biochemical and biophysical research communications*, 309(3), pp.625–630.

Bendardaf, R., Lamlum, H. & Pyrhönen, S., 2004. Prognostic and predictive molecular markers in colorectal carcinoma. *Anticancer research*, 24(4), pp.2519–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15330208.

Binkowski, B.F., 2012. NanoLuc ™ Luciferase : A Smaller , Brighter & More Versatile Luciferase Reporter. In *Promega Corporation*.

Birnboim, H.C. & Doly, J., 1979. A rapid alkaline ectraction procedure for sreening recombinant plasmid DNA. In *Nuclei Acids Research*. pp. 1513–1523.

Bratkovič, T., 2009. Progress in phage display: evolution of the technique and its applications. *Cellular and Molecular Life Sciences*, 67(5), pp.749–767. Available at: http://link.springer.com/10.1007/s00018-009-0192-2.

Brovko, L.Y., Anany, H. & Griffiths, M.W., 2012. Bacteriophages for Detection and Control of Bacterial Pathogens in Food and Food-Processing Environment. *Advances in food and nutrition research*, 67, p.241.

Brown, K.C., 2000. New approaches for cell-specific targeting: identification of cell-selective peptides from combinatorial libraries. *Current opinion in chemical biology*, 4(1), pp.16–21.

Brown, T.A., 2010. *Gene cloning & DNA analysis: An introduction* Sixth edit., Wiley-Blackwell.

Burch, J. a et al., 2007. Diagnostic accuracy of faecal occult blood tests used in screening for colorectal cancer: a systematic review. *Journal of medical screening*, 14(3), pp.132–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17925085.

Burritt, J.B. et al., 1995. Topological mapping of neutrophil cytochrome b epitopes with phage-display libraries. *Journal of Biological Chemistry*, 270(28), pp.16974–16980.

Cho, W.C.S., 2007. Contribution of oncoproteomics to cancer biomarker discovery. *Molecular cancer*, 6, p.25. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1852117&tool=pmcentrez&rendertype=abstract.

Christensen, D.J. et al., 2001. Phage display for target-based antibacterial drug discovery. *Drug discovery today*, 6(14), pp.721–727.

Clackson, T. & Wells, J.A., 1994. In vitro selection from protein and peptide libraries. *Trends in biotechnology*, 12(5), pp.173–184.

Derda, R. et al., 2011. Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules (Basel, Switzerland)*, 16(2), pp.1776–803. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21339712.

Derda, R. et al., 2010. High-throughput discovery of synthetic surfaces that support proliferation of pluripotent cells. *Journal of the ...*, 282(6), pp.1289–1295. Available at: http://pubs.acs.org/doi/abs/10.1021/ja906089g.

Deschoolmeester, V. et al., 2010. A review of the most promising biomarkers in colorectal cancer: one step closer to targeted therapy. *The oncologist*, 15(7), pp.699–731. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3228001&tool=pmcentrez&rendertype=abstract.

Deutscher, S.L., 2011. Phage Display in Molecular Imaging and Diagnosis of Cancer. , 110(5), pp.3196–3211.

Duffy, M.J. & Crown, J., 2008. A personalized approach to cancer treatment: how biomarkers can help. *Clinical chemistry*, 54(11), pp.1770–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18801934.

Ebrahimizadeh, W. & Rajabibazl, M., 2014. Bacteriophage vehicles for phage display: biology, mechanism, and application. *Current microbiology*, 69(2), pp.109–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24638925.

Even-Desrumeaux, K., Baty, D. & Chames, P., 2011. State of the art in tumor antigen and biomarker discovery. *Cancers*, 3(2), pp.2554–96. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3757432&tool=pmcentrez&rendertype=abstract.

Fearon, E.F. & Vogelstein, B., 1990. A Genetic Model for Colorectal Tumorigenesis. , 61, pp.759–767.

Frangioni, J. V et al., 2000. Minimal activators that bind to the KIX domain of p300/CBP identified by phage display screening. *Nature biotechnology*, 18(10), pp.1080–1085.

García-Bilbao, A. et al., 2012. Identification of a biomarker panel for colorectal cancer diagnosis. *BMC cancer*, 12(1), p.43. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3323359&tool=pmcentrez&rendertype=abstract.

GE Healthcare, 2007. HRP / Anti-M13 Monoclonal Conjugate Product Specification Sheet. Available at: https://www.gelifesciences.com/gehcls_images/GELS/Related Content/Files/1314750913712/litdoc27-9421-01PS Rev D 2007_20110831032452.pdf.

Ghosh, D. et al., 2012. Refactored M13 bacteriophage as a platform for tumor cell imaging and drug delivery. *ACS synthetic biology*, 1(12), pp.576–82. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3905571&tool=pmcentrez&rendertype=abstract.

Gibbs, J. & Kennebunk, M.E., 2001. Effective blocking procedures. *ELISA Technical Bulletin. Kennebunk, ME: Corning Incorporated Life Sciences*.

Giordano, R.J. et al., 2001. Biopanning and rapid analysis of selective interactive ligands. *Nature medicine*, 7(11), pp.1249–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11689892.

Green, M.R. & Sambrook, J., 2012. *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press New York.

GRiSP Team, 2014. GRiSP Research Solutions. Available at: http://www.grisp.pt/sites/default/files/GRiSP Catalogue 2014_web.pdf.

Hazbo, M.H. et al., 2003. Photographic and Luminometric Detection of Luciferase Reporter Phages for Drug Susceptibility Testing of Clinical Mycobacterium tuberculosis Isolates. , 41(10), pp.4865–4869.

Holten-Andersen, M.N. et al., 2004. Plasma TIMP-1 in patients with colorectal adenomas: a prospective study. *European Journal of Cancer*, 40(14), pp.2159–2164.

Hornbeck, P., 1991. Enzyme-Linked Immunosorbent Assays. *Current protocols in immunology*, pp.1–2.

Huang, J. et al., 2012. MimoDB 2.0: a mimotope database and beyond. *Nucleic Acids Research* , 40 (D1), pp.D271–D277. Available at: http://nar.oxfordjournals.org/content/40/D1/D271.abstract.

Huang, R. et al., 2011. Phage-Displayed Combinatorial Peptides. In A. B. Hughes, ed. *Protection Reactions, Medicinal Chemistry, Combinatorial Synthesis*. WILEY-VCH Verlag GmbH & Co. KGaA, pp. 451–471.

Huang, R., Fang, P. & Kay, B.K., 2012. Improvements to the Kunkel mutagenesis protocol for constructing primary and secondary phage-display libraries. *Methods (San Diego, Calif.)*, 58(1), pp.10–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3491174&tool=pmcentrez&rendertype=abstract.

Hundt, S., Haug, U. & Brenner, H., 2007. Blood markers for early detection of colorectal cancer: a systematic review. *Cancer Epidemiology Biomarkers & Prevention*, 16(10), pp.1935–1953.

Imperiale, T.F. et al., 2004. Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *New England Journal of Medicine*, 351(26), pp.2704–2714.

Karasseva, N.G. et al., 2002. Identification and characterization of peptides that bind human ErbB-2 selected from a bacteriophage display library. *Journal of protein chemistry*, 21(4), pp.287–296.

Kay, B.K., Winter, J. & McCafferty, J., 1996. *Phage display of peptides and proteins: a laboratory manual*, Academic Press.

Kelly, K. a & Jones, D. a, 2003. Isolation of a colon tumor specific binding peptide using phage display selection. *Neoplasia (New York, N.Y.)*, 5(5), pp.437–44. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1550331&tool=pmcentrez&rendertype=abstract.

Kim, H. et al., 2008. Noninvasive molecular biomarkers for the detection of colorectal cancer. *BMB reports*, pp.685–692.

Kropinski, A.M. et al., 2009. Bacteriophages M. R. J. Clokie & A. M. Kropinski, eds. , 501(1), pp.2–6. Available at: http://link.springer.com/10.1007/978-1-60327-164-6.

Kulasingam, V. & Diamandis, E.P., 2008. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nature clinical practice. Oncology*, 5(10), pp.588–99. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18695711.

Kuzmicheva, G.A. et al., 2009. Diversity and censoring of landscape phage libraries. *Protein Engineering Design and Selection*, 22(1), pp.9–18.

Lane, D.P. & Stephen, C.W., 1993. Epitope mapping using bacteriophage peptide libraries. *Current opinion in immunology*, 5(2), pp.268–271.

Li, K. et al., 2010. Chemical modification of M13 bacteriophage and its application in cancer cell imaging. *Bioconjugate chemistry*, 21(7), pp.1369–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20499838.

Lin, C.-H. et al., 2011. Molecular profile and copy number analysis of sporadic colorectal cancer in Taiwan. *Journal of biomedical science*, 18(1), p.36. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3123622&tool=pmcentrez&rendertype=abstract.

Liu, G., Chen, X. & Ai, H., 2012. Multifunctional Probes for Multimodality Imaging of Cancer. In *Molecular Imaging Probes for Cancer Research*. pp. 863–905.

Loktionov, A. et al., 1998. Quantitation of DNA from exfoliated colonocytes isolated from human stool surface as a novel noninvasive screening test for colorectal cancer. *Clinical cancer research*, 4(2), pp.337–342.

Losi, L. et al., 1996. K-ras and p53 Mutations in Human Colorectal Aberrant Crypt Foci. *The Journal of pathology*, 178(3), pp.259–263.

Ludwig, J. a & Weinstein, J.N., 2005. Biomarkers in cancer staging, prognosis and treatment selection. *Nature reviews. Cancer*, 5(11), pp.845–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16239904.

Marisa, L. et al., 2013. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS medicine*, 10(5), p.e1001453. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3660251&tool=pmcentrez&rendertype=abstract.

McConnell, S.I. et al., 1996. Construction and screening of M13 phage libraries displaying long random peptides. *Molecular diversity*, 1(3), pp.165–176.

McGuire, M.J. et al., 2004. In vitro Selection of a Peptide with High Selectivity for Cardiomyocytes In vivo. *Journal of molecular biology*, 342(1), pp.171–182.

Mcguire, M.J., Li, S. & Brown, K.C., 2009. *Biosensors and Biodetection* A. Rasooly & K. E. Herold, eds., Totowa, NJ: Humana Press. Available at: http://link.springer.com/10.1007/978-1-60327-569-9 .

Mohr, S., 2002. Microarrays as Cancer Keys: An Array of Possibilities. *Journal of Clinical Oncology*, 20(14), pp.3165–3175. Available at: http://www.jco.org/cgi/doi/10.1200/JCO.2002.12.073.

NEW ENGLAND BIOLABS ®, Electroporation Tips. Available at: https://www.neb.com/tools-and-resources/usage-guidelines/electroporation-tips?device=pdf.

NEW ENGLAND BIOLABS ®, Ph . D . ™ Phage Display Libraries Intruction Manual. Available at: http://www.neb.com/~/media/Catalog/All-Products/BDA9A6DB00DC42E8B93A8D8FBD08C49B/Datacards or Manuals/manualE8100.pdf.

NEW ENGLAND BIOLABS ®, Restriction Enzyme Troubleshooting Guide. Available at: https://www.neb.com/tools-and-resources/troubleshooting-guides/restriction-enzyme-troubleshooting-guide.

Newton, K.F., Newman, W. & Hill, J., 2011. Review of biomarkers in colorectal cancer. *Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland*, 14(1), pp.3–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21040359.

O'Neil, K.T. et al., 1992. Identification of novel peptide antagonists for GPIIb/IIIa from a conformationally constrained phage peptide library. *Proteins: Structure, Function, and Bioinformatics*, 14(4), pp.509–515.

Oyama, T. et al., 2003. Isolation of lung tumor specific peptides from a random peptide library: generation of diagnostic and cell-targeting reagents. *Cancer Letters*, 202(2), pp.219–230. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0304383503005494.

Pande, J., Szewczyk, M.M. & Grover, A.K., 2010. Phage display: concept, innovations, applications and future. *Biotechnology advances*, 28(6), pp.849–58. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20659548.

Paschke, M., 2006. Phage display systems and their applications. *Applied microbiology and biotechnology*, 70(1), pp.2–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16365766.

Pasqualini, R. & Ruoslahti, E., 1996. Organ targeting in vivo using phage display peptide libraries. *Nature*, 380(6572), pp.364–366.

Petrenko, V., 2008. Evolution of phage display: from bioactive peptides to bioselective nanomaterials. *Expert opinion on drug delivery*, 5(8), pp.825–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18712993.

Rasmussen, U.B. et al., 2002. Tumor cell-targeting by phage-displayed peptides. *Cancer gene therapy*, 9(7), pp.606–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12082461 [Accessed August 29, 2014].

Ray, P., Wu, A.M. & Gambhir, S.S., 2003. Advances in Brief Optical Bioluminescence and Positron Emission Tomography Imaging of a Novel Fusion Reporter Gene in Tumor Xenografts of Living Mice. *Cancer research*, 2(12), pp.1160–1165.

Rockstroh, M. et al., 2011. Cell fractionation - an important tool for compartment proteomics. *Journal of Integrated OMICS*, 1(1), pp.135–143. Available at: http://www.jiomics.com/index.php/jio/article/view/52 [Accessed May 25, 2014].

Roth®, C., 2009. Cell Passage and Use of Trypsin. , pp.1–2.

Russel, M., Lowman, H.B. & Clarkson, T., 2004. Chapter 1 Introduction to phage biology and phage display. In T. Clarkson & B. H. Lowman, eds. *Phage Display A Practical Approach*. Oxford, University Press, pp. 1–26.

Schatzlein, A.G. et al., 2001. Phage derived peptides for targeting of doxorubicin conjugates to solid tumours. *Journal of controlled release*, 74(1), pp.357–362.

Schmelcher, M. & Loessner, M.J., 2014. Application of bacteriophages for detection of foodborne pathogens. *Bacteriophage*, 4(1), p.e28137. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3919822&tool=pmcentrez&rendertype=abstract.

Scholle, M.D., Kehoe, J.W. & Kay, B.K., 2005. Efficient construction of a large collection of phage-displayed combinatorial peptide libraries. *Combinatorial chemistry & high throughput screening*, 8(6), pp.545–551.

Shao, Q., Yang, Y.M. & Xing, B.G., 2012. Chemistry of Optical Imaging Probes. In *Molecular Imaging Probes for Cancer Research*. pp. 419–444.

Shen, L. et al., 2007. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proceedings of the National Academy of Sciences*, 104(47), pp.18654–18659.

Shinya, H. & Wolff, W.I., 1979. Morphology, Anatomic Distribution and Cancer Potential of Colonic Polyps. *Annals of Surgery*, 190(6), pp.679–683. Available at: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00000658-197912000-00001.

Shivapurkar, N. et al., 1997. K-ras and p53 mutations in aberrant crypt foci and colonic tumors from colon cancer patients. *Cancer letters*, 115(1), pp.39–46.

Sidhu, S.S., 2001. Engineering M13 for phage display. *Biomolecular engineering*, 18(2), pp.57–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23422437.

Simon, S., 2014. Stool DNA Testing for Colon Cancer. *American Cancer Society*. Available at: http://www.cancer.org/cancer/news/features/stool-dna-testing-for-colon-cancer.

Singh, A., Poshtiban, S. & Evoy, S., 2013. Recent advances in bacteriophage based biosensors for food-borne pathogen detection. *Sensors (Basel, Switzerland)*, 13(2), pp.1763–86. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3649382&tool=pmcentrez&rendertype=abstract.

Smith, A.J. et al., 1994. Somatic APC and K-ras codon 12 mutations in aberrant crypt foci from human colons. *Cancer research*, 54(21), pp.5527–5530.

Smith, G.P., 1985. Filamentous Fusion Phage : Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. , 508(1984), pp.4–6.

Smith, G.P. & Petrenko, V. a., 1997. Phage Display. *Chemical reviews*, 97(2), pp.391–410. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12538914.

Soper, S. a et al., 2006. Point-of-care biosensor systems for cancer diagnostics/prognostics. *Biosensors & bioelectronics*, 21(10), pp.1932–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16473506.

Srivastava, S., Verma, M. & Henson, D.E., 2001. Biomarkers for Early Detection of Colon Cancer Biomarkers for Early Detection of Colon Cancer. , pp.1118–1126.

Strober, W., 2001. Trypan blue exclusion test of cell viability. *Current protocols in immunology / edited by John E. Coligan ... [et al.]*, Appendix 3, p.Appendix 3B. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18432654.

Tanaka, T. et al., 2010. Biomarkers for colorectal cancer. *International journal of molecular sciences*, 11(9), pp.3209–25. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2956090&tool=pmcentrez&rendertype=abstract.

Tothill, I.E., 2009. Biosensors for cancer markers diagnosis. *Seminars in cell & developmental biology*, 20(1), pp.55–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19429492.

Ulitzur, S. & Kuhn, J., 1987. Introduction of lux genes into bacteria, a new approach for specific determination of bacteria and their antibiotic susceptibility. *Bioluminescence and chemiluminescence new perspectives. Wiley, New York*, pp.463–472.

Voss, S.D. et al., 2002. An integrated vector system for cellular studies of phage display-derived peptides. *Elsevier Science*, 308, pp.364–372.

Walther, A. et al., 2009. Genetic prognostic and predictive markers in colorectal cancer. *Nature reviews. Cancer*, 9(7), pp.489–99. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19536109 [Accessed May 23, 2014].

Willats, W.G.T., 2002. Phage display: practicalities and prospects. *Plant molecular biology*, 50(6), pp.837–854.

Wilson, K. & Walker, J., 2010. *Principles and Techniques of Biochemistry and Molecular Biology* 7th ed., Cambridge University Press.

Winawer, S. et al., 2003. Colorectal cancer screening and surveillance: clinical guidelines and rationale-Update based on new evidence. *Gastroenterology*, 124(2), pp.544–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12557158.

Wulfkuhle, J.D., Liotta, L. a & Petricoin, E.F., 2003. Proteomic applications for the early detection of cancer. *Nature reviews. Cancer*, 3(4), pp.267–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12671665.

Xiao, Y. & Isaacs, S.N., 2013. Enzyme-Linked Imunnosorbent Assay (ELISA) and Blocking with Bovine Serum Albumin (BSA) - Not all BSAs are alike. , 384, pp.148–151.

Yata, T. et al., 2014. Hybrid Nanomaterial Complexes for Advanced Phage-guided Gene Delivery. *Molecular Therapy—Nucleic Acids*, 3(8), p.e185.

Zavoral, M. et al., 2009. Colorectal cancer screening in Europe. *World Journal of Gastroenterology*, 15(47), p.5907. Available at: http://www.wjgnet.com/1007-9327/15/5907.asp.

Zhang, Y. et al., 2007. Panning and identification of a colon tumor binding peptide from a phage display peptide library. *Journal of biomolecular screening*, 12(3), pp.429–35. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17332091.

APPENDIX 1 - BUFFERS AND SOLUTIONS

PBS buffer (1X) – 137 mM Sodium Chloride [Panreac], 10 mM Sodium Phospate Dibasic [Scharlau], 2.7 mM Potassium Chloride [AppliChem] and Potassium Phosphate Monobasic [Riedel de Haën] in deionized water.

TSS buffer – LB medium supplemented with 10% (wt/vol) PEG 8000 [Fisher Scientific], 5% (vol/vol) DMSO, and 6 g/L of $MgCl_2.6\ H_2O$.

1M Tris-HCl Buffer – 121.14 g/L of Tris-base [AnalaR, 103156X] in deionized water; the quantity of $HCl_{(aq)}$ (chloridric acid) added should be adjusted for the desired pH.

SM buffer – 100 mM sodium chloride (NaCl) [Panreac, 1316591211], 8 mM magnesium sulfate heptahydrate ($MgSO_4.7H^2O$) [Panreac, 1058860500] and 50 mM Tris-HCl (pH 8) in deionized water.

P1 Solution – 50 mM glucose ($C_6H_{12}O_6$) [Liofilchem®], 10 mM EDTA [Sigma-Aldrich Co. LLC, Fluka Analytical] and 1 mM Tris-HCl (pH 8) in deionized water.

P2 Solution – 0.2 M sodium hydroxide (NaOH) [Sigma -Aldrich Co. LLC] and 1 % sodium dodecyl sulfate (SDS) [Fisher Scientific, Portugal] in deionized water.

P3 Solution – 3 M potassium acetate (KOAc) [Fisher Scientific] and 11 % acetic acid (CH_3COOH) [Fisher Scientific] in deionized water.

TE buffer – 0.1 M Tris-base and 10 mM EDTA in deionized water.

TAE buffer (1X) – 40 mM Tris-base, 20 mM acetic acid and 0.05 M EDTA in deionized water.

TBST buffer – 50 mM Tris-HCl (pH 7.5), 150 mM NaCl and 0.1% (v/v) Tween-20 [Fisher Scientific], pH adjusted to 7.6 adding HCl, in deionized water.

Iodide buffer – 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 4 M sodium iodide (NaI) in deionized water.