Cláudia Sofia Mendes Fernandes

Licenciada em Biologia Celular e Molecular

Magnetic nanoparticles as a versatile solid-support for fusion protein purification and antimicrobial assays

Dissertação para obtenção do Grau de Mestre em Biotecnologia

Orientadora: Prof.ª Doutora Ana Cecília Afonso Roque
Co-orientador: Doutor Ricardo Jorge Flores Branco

Júri:

Presidente: Prof. Doutor Pedro Miguel Ribeiro Viana Baptista
Arguente: Prof.ª Doutora Maria Raquel Murias dos Santos Aires Barros
Vogal: Prof.ª Doutora Ana Cecília Afonso Roque

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro 2013

Cláudia Sofia Mendes Fernandes

Licenciada em Biologia Celular e Molecular

Magnetic nanoparticles as a versatile solid-support for fusion protein purification and antimicrobial assays

Dissertação para obtenção do Grau de Mestre em Biotecnologia

Orientadora: Prof.ª Doutora Ana Cecília Afonso Roque
Co-orientador: Doutor Ricardo Jorge Flores Branco

Júri:

Presidente: Prof. Doutor Pedro Miguel Ribeiro Viana Baptista
Arguente: Prof.ª Doutora Maria Raquel Murias dos Santos Aires Barros
Vogal: Prof.ª Doutora Ana Cecília Afonso Roque

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro 2013

MAGNETIC NANOPARTICLES AS A VERSATILE SOLID-SUPPORT FOR FUSION PROTEIN PURIFICATION AND ANTIMICROBIAL ASSAYS

AGRADECIMENTOS

Agradeço a todos aqueles que me ajudaram a concluir a minha tese de mestrado neste último ano.

Em primeiro lugar, à Professora Cecília Roque, que me deu a oportunidade de trabalhar no seu laboratório. Obrigada por todo o apoio e orientação e por todo o tempo que me dedicou e conversas encorajadoras, que me ajudaram a decidir os meus próximos passos.

Um grande agradecimento ao meu co-orientador, Doutor Ricardo Branco, que me apresentou a um novo e abrangente mundo da ciência e me acompanhou em todos os passos. Obrigada por todo o apoio oferecido, dentro e fora do meu trabalho.

A todos os meus colegas do Biomolecular Engineering Group, pelo bom ambiente dentro e fora do laboratório, pelos sorrisos nos tempos livres e pela preciosa ajuda que tiveram na realização do meu trabalho. Em particular, à Doutora Ana Pina pelas valiosas instruções e discussões, carinho e amizade e por me dar força para os meus primeiros passos no mundo da investigação. Não podia deixar de referir a Bianca Gonçalves, por me ajudar a passar os tempos de espera e por me ter dado companhia nas horas mais tardias.

A todos os meus amigos espalhados pelos laboratórios que contribuíram para as minhas horas de almoço e outras horas afins, que me deram força e me mostraram toda a sorte que tive durante todo este ano. Um obrigado a todos os meus amigos fora dos laboratórios por todos os bons momentos e companhia fora das horas de trabalho e que contribuíram para os momentos de lazer. Incluindo todos aqueles que se encontram longe demais mas cujo apoio esteve sempre presente.

Um muito obrigado à minha família, por todo o interesse e por todo o apoio que me dão em todas as etapas da minha vida. Esta foi mais uma que consegui finalizar com sucesso com a vossa ajuda.

Em último lugar, um agradecimento especial ao Tiago Monteiro, um amigo, um companheiro, que me acompanhou em todos estes passos. Tudo teria sido mais complicado sem a sua ajuda e sem o seu inabalável apoio, nos bons e maus momentos.

RESUMO

Partículas magnéticas constituem plataformas com propriedades únicas para a purificação de proteínas e ensaios antimicrobianos.

Neste trabalho partículas magnéticas foram funcionalizadas com novos ligandos que foram utilizados para a purificação de proteínas. Devido às suas características inovadoras, estes ligandos foram estudados por técnicas de modelação e simulação molecular, de modo a explicar a interação entre dois pares de afinidade: GFP/LA-A4C7 e RK-GFP/LR-A7C1. Os resultados obtidos apontam para uma interação entre GFP e LA-A4C7 de natureza hidrofóbica, enquanto a interação entre RK-GFP e LR-A7C1 é assegurada por pontes de hidrogénio. Uma segunda biblioteca de ligandos de afinidade para a purificação da GFP foi idealizada com base na maximização da energia livre de ligação estimada e correspondente constante de afinidade.

Ao imobilizar os ligandos biomiméticos nas partículas magnéticas foi possível eluir a proteína de interesse e recuperá-la com elevada pureza. As condições ideais de eluição para GFP foram 0.1mM glicina-NaOH pH9 50% (v/v) etileno glicol e as melhores condições de eluição para RK-GFP foram PBS pH7.4, 500mM arginina. As constantes de ligação para os sistemas estudados ($Ka=0.83\times10^{5}M^{-1}$ e $Q_{max}=4mg/g$ para GFP/LA-A4C7, $Ka=3.21\times10^{5}M^{-1}$ e $Q_{max}=2mg/g$ para RK-GFP/LR-A7C1) são promissoras para um sistema de purificação de proteínas por afinidade. Os resultados experimentais estão de acordo com os resultados teóricos.

Finalmente, MNPs foram funcionalizadas com o $(RW)_3$, um péptido com propriedades antimicrobianas, por diferentes métodos de imobilização. Foi possível observar o potencial antimicrobiano contra *Escherichia coli* e *Bacillus subtilis* de um novo dispositivo baseado no método de imobilização por "EDC-coupling".

Palavras-chave: nanopartículas magnéticas; ligandos de afinidade; modelação e simulação molecular; purificação de proteínas; péptidos antimicrobianos

ABSTRACT

Magnetic nano- and microparticles are unique platforms for the development of bioseparation and antimicrobial devices.

This work explored the application of magnetic particles for the purification of fusion proteins through the use of magnetic adsorbents coupled to novel affinity ligands towards peptidic and proteic tags. Furthermore, and in view of the novelty of these ligands, molecular modeling and simulation techniques were employed to explain the key structural features involved in the binding of two affinity pairs: GFP/LA-A4C7 and RK-GFP/LR-A7C1. The results showed that the interaction between GFP and LA-A4C7 is mainly hydrophobic while the interaction between RK-GFP and LR-A7C1 is mostly driven by hydrogen bonds. Moreover, the same modeling techniques have been used to idealize a theoretical second generation library with view of maximizing the estimated free energy of binding and the correspondent affinity constant.

When immobilizing the biomimetic ligands LA-A4C7 and LR-A7C1 onto magnetic nanoparticles, it was possible to bind the protein of interest and recover pure elution fractions. The best elution condition for GFP elution was 0.1mM glycine-NaOH pH9 50% (v/v) ethylene glycol and the best elution condition for RK-GFP elution was PBS pH 7.4, 500mM arginine, which are in accordance with the theoretical results described previously. Final binding constants for the studied systems ($Ka=0.83\times10^{5}M^{-1}$ and $Q_{max}=4mg/g$ for GFP/LA-A4C7, $Ka=3.21\times10^{5}M^{-1}$ and $Q_{max}=2mg/g$ for RK-GFP/LR-A7C1) show promising results for an affinity-based protein purification system.

Magnetic particles were also functionalized with $(RW)_3$, an peptide with antimicrobial properties, by different routes. We were able to develop a novel antimicrobial nanodevice based on the EDC-coupling of $(RW)_3$ that has shown antimicrobial activity against *Escherichia coli* and *Bacillus subtilis*.

Keywords: magnetic nanoparticles; affinity ligands; molecular modeling and simulation; protein purification; antimicrobial peptides

CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

LIST OF ABBREVIATIONS

AMP – antimicrobial peptides

ANOLEA – atomic non-local environment assessment

APS – ammmonia persulfate

APTES – (3-aminopropyl)triethoxysilane

BCA – bicinchoninic acid assay

BSA – bovine serum albumin

c-myc – protein tag composed by the EQKLISEEDL peptide sequence, which is derived the c-myc gene product

Cys-$(RW)_3$ – peptide composed by C-PP-RWRWRW

Cys-Tet213 – antimicrobial peptide composed by CRRWWKWWRW

DLS – dynamic light scattering

EA – elution condition A

EB – elution condition B

EDTA – ethylenediamine tetraacetic acid

EDC – N-(3-dimethylaminopropyl)-N'-ethyl-carbodiimide

IMAC – immobilized metal ion affinity chromatography

IPTG – isopropyl β-D-1-thiogalactopyranoside

Fe_2O_3 – maghemite

Fe_3O_4 – magnetite

FLAG – protein tag composed by DYKDDDDK peptide sequence

FT – flow-through

G3R6TAT – cell penetrating antimicrobial peptide composed by GGRRRRRRYGRKKRRQRRR peptide sequence

GFP – green fluorescent protein

GFP_IT – GFP homology model obtained by the I-TASSER webserver

GFP_R – GFP homology model obtained by the Rosetta software

GFP_s1M – GFP homology model obtained by the Modeller software

GFP_SM – GFP homology model obtained by SWISS-MODEL webserver

HA-tag – protein tag composed by YDVPDYA peptide sequence, which is derived from the human influenza hemagglutinin

His-tag – protein tag composed by 5-15 repeats of histidine

HST – high-throughput screening

Ka – affinity constant

Ki – estimated inhibition constant

LB – Luria Broth

LA-A4C7 – Ugi-based biomimetic affinity lead ligand for the purification of GFP

LL-37 – antimicrobial peptide composed by the LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES peptide sequence

LR-A7C1 – Ugi-based biomimetic lead affinity ligand for the purification of RK-GFP

MD – molecular dynamics

MIC – minimum inhibitory concentration

MNP – magnetic nanoparticle

MNP_DEX_NH_2 – aminated dextran-coated magnetic nanoparticles

MNP_PEG_NH_2 – aminated PEG-coated magnetic nanoparticles

MTT – 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, tetrazolium dye that is used for measuring the activity of cellular enzymes

NA-A7C1 – Ugi-based biomimetic ligand used as negative control for the modeling study of GFP

NHS – N-Hydroxysuccinimide

NR-A5C2 – Ugi-based biomimetic ligand used as negative control for the modeling study of RK-GFP

OD_{600nm} – optical density measured at 600nm

pAP001 – expression vector containing the RK-GFP gene

pAP006 – expression vector containing the GFP gene

PDB – RCSB Protein Data Bank

PBS – buffer containing 10mM sodium phosphate, 150mM NaCl, pH7.4

pDNA – plasmid DNA

PEG – poly(ethylene glycol)

QMEAN – qualitative model energy analysis

RK-GFP – tagged-GFP that comprises a protein tag at N-terminus composed by the “RKRKRK-PPP-DDDKG-TGS-“ amino acid sequence

RK-GFP_R – RK-GFP homology model obtained by the Rosetta software

RK-GFP_s1M – RK-GFP homology model obtained by the Modeller software

RMSD – root mean square deviation

RT – room temperature

$(RW)_3$ – peptide composed by RWRWRW

SDS – sodium dodecyl sulphate

S-tag – protein tag composed by LETAAALFERQHMDS amino acid sequence, which is derived from the pancreatic ribonuclease A

Streptag II – protein tag composed by WSHPQFEL

Sulfo-SMCC – sulfosuccinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate

TAE – buffer containing 40mM Tris-base, 20mM glacial acetic acid, 1mM EDTA, pH8.4

TEMED – tetramethylethylenediamine

1. LITERATURE REVIEW

1.1. Magnetic core materials

1.1.1. Synthesis, coating and functionalization

Magnetic nanonoparticles (MNPs) are usually formed by maghemite (Fe_2O_3) or magnetite (Fe_3O_4), usually in the nano- to micrometer range. Usually MNP have single domains of about 5-20nm in diameter [1]. MNP exhibit superparamagnetic behavior and become easily magnetized upon exposure to a magnetic field [2].

Many synthesis techniques have been applied to obtain high quality iron oxide dispersed MNP, as co-precipitation, thermal decomposition, microemulsion, hydrothermal, electrochemical or biological synthesis. These methods have been reviewed and compared elsewhere (e.g. [3–5]).

The most commonly used method is the co-precipitation, an efficient and easy method to synthesize high amounts of magnetite MNPs [3]. This reaction occurs at a defined stoichiometric Fe^{2+}/Fe^{3+} ratio by the addition of a strong base under inert atmosphere [5]. The overall reaction may be written as:

$$Fe^{2+} + 2Fe^{3+} + 8OH^- \rightarrow Fe_3O_4 + 4H_2O \qquad (Equation\ 1.1)$$

The main disadvantage of this method is that it allows a low control of shape, size distribution and leads to particle aggregation [5]. However these parameters can be tuned by changing the type of salts used, the Fe^{2+}/Fe^{3+} ratio, temperature, pH and ionic strength of the media [6].

Due to their small size and large surface area, non-coated MNPs tend to aggregate [5,7]. Steric or electrostatic stabilization through coating is required to ensure stable aqueous dispersions [4], as well as increased biocompatibility, decreased non-specific interactions and facilitated functionalization [8]. The presence of hydroxyl groups at the MNPs surface allows a versatile functionalization [4].

Biopolymers are widely used for MNP stabilization [4], as they are renewable, generally non-toxic and biodegradable. Dextran is a R-D-glucopyranosyl polyssaccharide [3] and it is the mostly used polymer for MNP coating [2]. It confers optimal polar interactions with iron oxide surfaces, stability in physiological fluids [4], biocompatibility [5] and provides a chemical

handle for further chemical modification with proteins [9] and affinity ligands for purification purposes [10].

Synthetic polymers are also employed for MNP coating (e.g. poly(ehtyleneglycol) (PEG), poly(vinyl alcohol), poly(lactide acid), alginate, polyacrylic acid), with different advantages. For example, PEG-coated MNP are hydrophilic, water-soluble and biocompatible; inorganic coating can decrease MNP oxidation, and silica is the most commonly used by providing stability, biocompatibility and hydrophilicity [5].

MNPs have been used in a wide range of applications, as sealants, damping agents, drug and gene delivery vehicles, MRI contrast agents and materials for tissue engineering. These applications have been thoroughly described elsewhere [3–5]. In this work the potential of MNPs for bioseparation and as antimicrobial devices has been explored.

1.1.2. MNP as solid-support for affinity-based bioseparation

Recombinant proteins are used in numerous areas of biosciences and biotechnology [11]. In 2010 the market for recombinant proteins exceeded USD100 billion and it is expected to rise up to USD169 billion by 2014 [12]. However, protein purification accounts as much as 60-90% of the total production costs [7]. This step is usually performed using a combination of chromatography, ultrafiltration, precipitation, among others techniques [11].

Currently affinity-based separations are the most powerful tools available for downstream processing [11]. These include affinity chromatography [13], which is based on a specific, reversible and non-covalent interaction between a pair of complementar molecules (e.g. a ligand and a tag) [14]. Due to this high specificity, the use of affinity methods reduces non-specific interactions, increases yields and facilitates the elimination of contaminants [14].

To confer selectively to the solid-support, the separation matrix must be functionalized with specific affinity ligands. Currently there are three main categories of affinity ligands: biological, structural and synthetic ligands.

Biological ligands are biological molecules that recognize their specific and natural binding targets. These ligands present high selectivity and are well-established and optimized. However they present some disadvantages, as the high purification cost, instability with consequent leaching and product contamination, and high cost for large-scale production

[15,16]. The most well-known example is Protein A that binds selectively to the Fc domain of antibodies [13].

Structural ligands are based on easy chemistry, which can produce at affordable prices and are stable to harsh conditions as sterilization and cleaning-in place procedures [17]. The most used strategy is the immobilized metal ion affinity chromatography (IMAC), that is based on interaction between a metal ion (Co^{2+}, Ni^{2+}, $Cu^{2+,}$ Zn^{+2}) immobilized on a matrix and specific aminoacid side chains [17]. However these ligands are not suitable for metal-containing protein purification, and are prone to metal leakage and unspecific binding [17].

Synthetic ligands are tailor-made molecules that mimic natural biological interactions [14]. These are derived from a combination of modeling studies, *in silico* approaches and high throughput combinatorial chemistry techniques [15]. These methods allow the discovery of robust alternatives for known biological ligands [16]. Synthetic ligands allow to overcome the disadvantages of biological ligands, while maintaining affinity and specificity and improving their characteristics [14]. Triazine-based ligand 22/8 [18] and ligand 8/7 [19] and Ugi-based ligands [15] have shown to display affinity towards human IgG.

Instead of developing individualized purification protocols for each protein, the purification of recombinant proteins can be simplified by genetically fusing the protein to an affinity tag [17]. Tags are defined as exogenous, preferentially short amino acid sequences that present a high affinity for a specific ligand. Tags should allow the purification of tagged-proteins with high yields and purity, have minimal effect on the tertiary structure and biological activity of the protein, be non-immunogenic and non-toxic [13,17,20]. Depending on the final application, tags do not need to be removed after protein purification [13].

On the other hand, many authors have explored the use of non-chromatographic tags. These methods allow protein purification due to the properties conferred by the tags, reducing the cost of using affinity resins [13]. These include elastin-like polypeptide or annexin-B tagged-proteins that allow the selective and reversible precipitation of the fusion protein in response to temperature and ionic strength changes [13]; phasin-tagged proteins are co-expressed with polyhydroxybutyrate or polyhydroxyalkanoates granules, binding selectively to the granules which are recovered by centrifugation [13,17]; other tags promote the formation of inclusion bodies to facilitate the purification steps [21].

Tags can be used due to other properties besides protein purification.

Self-cleaving tags (e.g. inteins) can be used to avoid external tag-cleavage methods. These tags have inducible proteolytic activity [13,21], which can be triggered by changes of solution pH or the addition of a thiol reagent [17,21].

Other properties can be conferred by a tag such as improved protein yield, proteolysis prevention, protection of the host against antigenicity of the fusion protein, facilitated protein refolding and increased solubility [17,21].

A typical affinity separation method based on affinity tags is simple, starting with the expression of a target protein fused with a tag. Afterwards the protein crude extract is incubated with the matrix containing the immobilized ligand, cell contaminants are washed off and the target protein is eluted from the matrix by a sudden change in conditions (e.g. pH, ionic strength, use of solutions with chaotropic salts or with polarity reducing agents) [11,13].

However the traditional affinity chromatography presents several technical disadvantages, as fouling when using viscous samples and limitations by pore diffusion. To overcome these disadvantages, other affinity techniques with novel solid supports have been investigated. A well-studied example are magnetic core materials functionalized in order to selectively capture the desired molecule.

MNPs have several advantages when compared to standard bioseparations procedures. The target compounds can be isolated directly from crude samples (e.g. blood, cultivation media) [1,22]. Their magnetic properties also allow to easily and selectively remove the functionalized MNPs from viscous sample solutions [1]. MNPs can be used in batch processes which allows to work with various sample volumes [1]. MNPs have a highly accessible surface area and are non-porous supports, therefore bioseparations are not limited by pore diffusion [7,22].

MNPs have been explored as a solid support for magnetic separation of various biomolecules [7].

To be suitable for protein purification, MNPs should (1) be stable under synthesis, screening and elution conditions, (2) hydrophilic and inert, (3) amenable to further functionalization [22], (4) responsive to an external magnetic field, (5) possess large surface areas, (7) reasonable price, and (9) easy production methods [8].

Similar to chromatographic affinity separations, various biological molecules have been used to functionalize the coated-surface of MNPs [1,8] (Table 1.1). These can be based on general non-specific (electrostatic, hydrogen bonding, hydrophobic), or group-specific interactions (chelating, IMAC [23], and specific affinity interactions (antibody-antigen pair [24], steptavidin-biotin [25], avidin-biotin, enzyme-inihibitor, other ligands [26,27]).

Table 1.1 - Examples of affinity ligands used to functionalize magnetic carriers for protein purification purposes.

Purified protein	Affinity ligand	Reference
Luciferase (His-tagged)	Ni^{2+}	[23]
Haemoglobin	Cu^{2+}	[28]
Biotinylated peptides	Streptavidin	[29]
Nisin Z	Anti-nisin antibody	[30]
Trypsin	Soybean trypsin inhibitor	[31]
Human serum albumin	Cibacron Blue F3GA	[27]
Lysozyme	Cibacron Blue F3GA	[26]
IgG	Ligand 22/8	[10]
IgG	Ligand 8/7	[22]

1.1.3. MNP as solid-support for antimicrobial peptides tethering

The use of antibiotics has led to the appearance of multi-drug resistant microorganisms, specially in hospital environments. AntiMicrobial Peptides (AMPs) are seen as promising alternatives [32], capable of eliminating a broad spectrum of microorganisms, including Gram-positive [33] and Gram-negative [34] bacteria and fungi [32]. AMPs are a promising class of molecules which have shown to be able to reduce infections without easily incurring pathogen resistance [33,35], or cross-resistance [32].

Most known AMPs share the same characteristics: (a) positively charged residues, and (b) residues with a hydrophobic side chain [36]. High contents of phosphatidyl serine and phosphatidyl glycerol (negatively charged at physiological pH) at the outer layer of most bacterial membranes allow the preferential activity of AMPs [32,37]. Upon contact with the bacterial membrane, AMPs adopt a particular secondary conformation, allowing the insertion of the hydrophobic components into the membrane lipid domains and the disruption of its structure [37]. Two mechanisms for AMP action have been proposed - pore-forming and non-pore forming – and have been reviewed thoroughly elsewhere [37]). Both rely on the physical disruption of the bacterial membrane structure [37].

Nowadays the choice of antimicrobial molecules is large and tends to increase. Hybrid antimicrobial materials have two or more functional fragments combined [35] – some literature examples are described in Table 1.2.

Table 1.2 - Examples of AMP-functionalized solid supports.

Solid support	AMP	Microbes tested	Antimicrobial assay	Reference
Supermagnetic iron oxide particles	Bacitracin	*E. coli, S. aureus, B. subtilis*	Agar plating; OD measurement	[38]
Silica magnetic microparticles	Magainin I	*Listeria ivanovii*	Fluorescencent LIVE/DEAD Baclight assay	[33]
Nickel magnetic nanoparticles	LL-37	*E.coli*	Disk diffusion test	[34]
Silver nanoparticles	G_3R_6TAT	*B. subtilis, E. coli, Candida albicans*	Broth microdilution method	[39]
Multi-wall carbon nanotubes	Nisin	*E. coli, P. aeruginosa, S. aureus, B. subtilis*	Agar plating	[40]
Titanium (Ti) surface	Cys-Tet213	*P. aeruginosa*	Fluorescence assay	[41]
Stainless steel surface	Magainin I and nisin	*L. ivanovii*	Agar plating	[42]
Quartz surface	Cys-Tet213	*P. aeruginosa*	Fluorescence assay	[43]
Discs of various materials	apoE derivates	*S. aureus, P. aeroginosa*	Neutral red and MTT assay; disk diffusion test	[44]

Antibacterial coatings are based on the immobilization of AMPs onto planar surfaces while keeping its accessibility and antimicrobial activity. Recent examples include functionalization of titanium [41,43,45], quartz [43] and steel surfaces [42]. Although functionalization of planar surfaces yields protection against biofilm formation, it is inefficient against bacteria present in the surrounding medium [33].

Nanostructured antimicrobials allow an increased local charge density and general enhanced antimicrobial activity [37].

Although magnetic silver nanomaterials are the most studied as antimicrobial agent [46], metal oxide nanoparticles functionalized with AMPs have been recently described as antimicrobial agents and gather great interest [38]. Magnetic iron oxide [38], silica [33] and

nickel [34] particles have been functionalized with various AMPs in order to display antimicrobial activity against Gram-positive and Gram-negative bacteria. Properly functionalized particles are able to disinfect a solution containing bacteria, while their magnetic properties allow easy removal from solution which would avoid environmental dissemination [33].

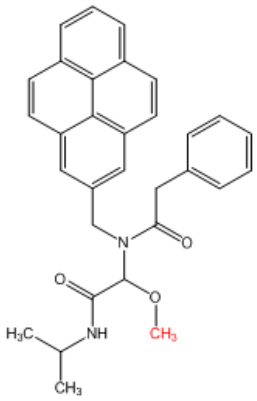
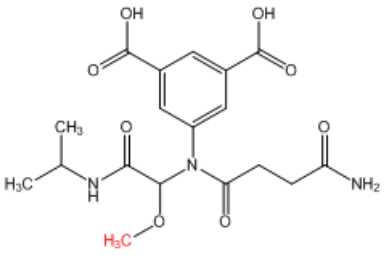
2. BACKGROUND AND AIM OF THE WORK

Magnetic particles have been used in a wide range of applications, due to their versatile properties, as large surface area to volume ratio and superparamagnetic properties. In this work the potential of MNPs for bioseparation and as antimicrobial devices has been explored.

New affinity pairs ligand-tag were previously developed by the Biomolecular Engineering Group@REQUIMTE [47]. In this work, interesting Ugi-based lead ligands specific for the purification of green fluorescent protein (GFP) and GFP tagged-systems were developed. The ligands were obtained by a combined methodology of synthesis of libraries of ligands immobilized on solid support (agarose) and subsequent high-throughput screening (HTS).

The tagged-GFP studied in this work comprised a protein tag at N-terminus composed by the "RKRKRK-PPP-DDDKG-TGS-" amino acid sequence (RK-GFP). Lead ligands for GFP and RK-GFP are shown in Table 2.1.

Table 2.1 – The affinity pairs studied in this work. The chemical group used to immobilize the ligand onto the solid support is highlighted in red.

Target protein	**GFP**	**RK-GFP**
Lead Affinity Ligand	LA-A4C7 	LR-A7C1 

Previous work has also included the study of others tags with versatile properties, as antimicrobial properties (RWRWRW-tag) or induction of the formation of inclusion bodies (WFWFWF-tag) [47].

Based on these previous results, and in view of the importance of magnetic supports for several biotechnological and biomedical areas, this dissertation had three main goals:

(1) To study the potential key structural features of the interaction between the recombinant proteins GFP and RK-GFP with the respective lead ligand, employing homology modeling, automated docking and molecular dynamics simulation protocols.

(2) To study the potential of the novel biomimetic ligands for recombinant protein (GFP and RK-GFP) purification purposes when coupled to magnetic supports.

(3) To study the potential of MNPs as antimicrobial devices through the coupling of $(RW)_3$, an antimicrobial peptide, against Gram-positive and Gram-negative bacteria.

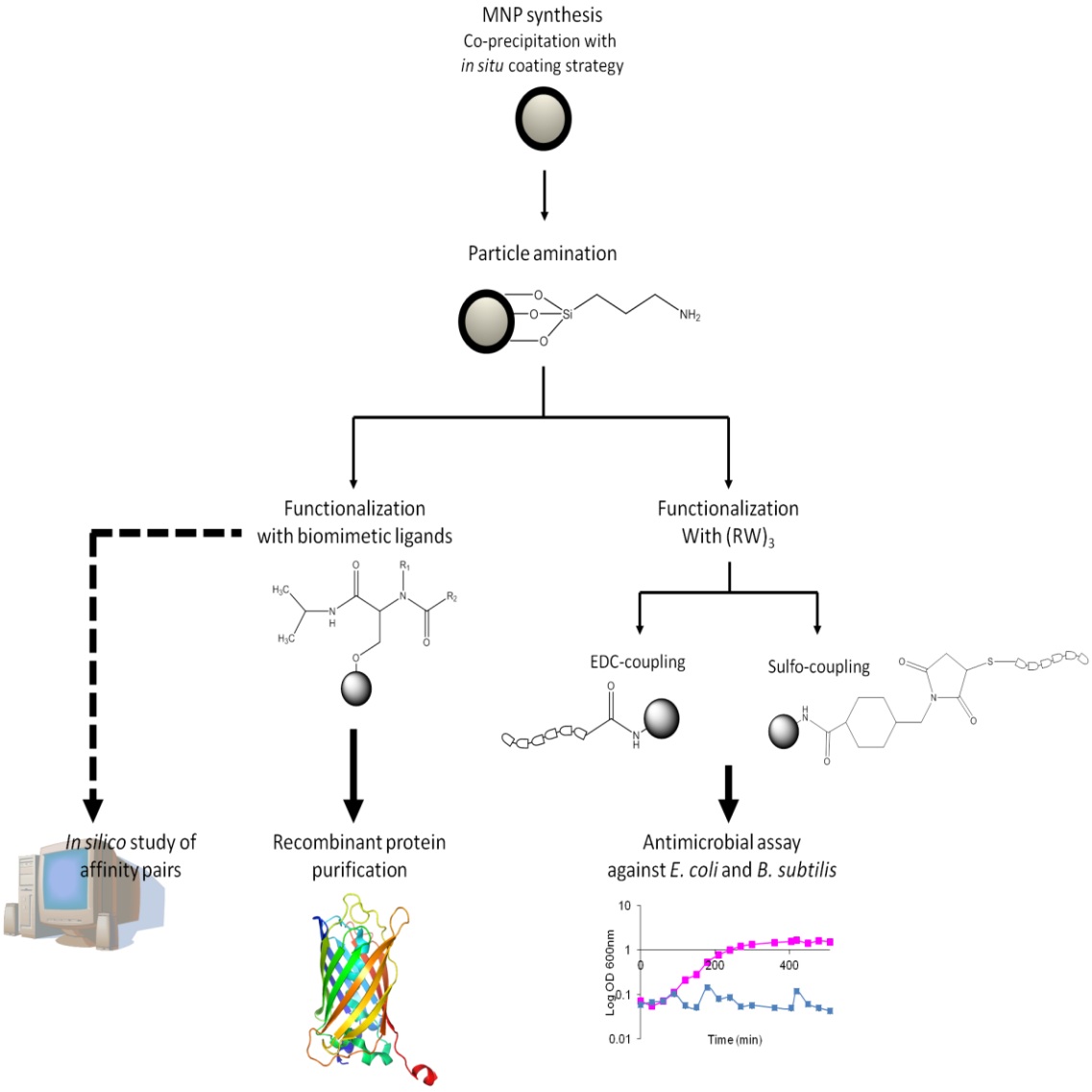


Figure 2.1 – Schematic representation of the research work followed in this work.

3. MATERIALS AND METHODS

3.1. Materials

The reagents utilized were of the highest grade available.

For the MNP synthesis and surface modification we used dextran (40000g/mol; A2249.100) from Biochemica, poly(ethylene glycol) (10000g/mol; 81280), ammonium hydroxide 25% (986030501), ninhydrin (33437) from Fluka, iron(III) chloride (44944), iron(II) chloride (44939), phenol (185450) and potassium cyanide (207810) from Sigma, ethanol absolute (1210851212) from Panreac, (3-aminopropyl)trithoxysilane (APTES) (44014-0), pyridine (00633BE-137) from Aldrich and glycine (120072500) from Acros. A sample of silica/dextran- and silica/PEG-coated MNPs has been kindly provided by Iris Batalha from the Biomolecular Engineering Group (FCT-UNL).

For ligand attachment onto MNP we used glutaric dialdehyde (340855), N,N-dimethylformamide (33437) from Sigma, sodium hydroxide (1316871211) from Panreac, methanol (8388.5) from Roth, isopropyl isocyanide (553344) from Aldrich. For ligand synthesis, the chemical compounds were aminoisophtalic acid (33437) from Sigma, succinamic acid (134376), 4-aminobenzamide (284572), 3-(4-hydroxyphenyl)propionic acid (H52406), 1-pyrenemethylamine hydrochloride (410705) and phenylacetic acid (P16621) from Aldrich. Further chemicals used were ethylene glycol (31000012131623), sodium-di-hydrogen phophate 1-hydrate (1319651211), (Di) sodium-hydrogen phosphate 2-hydrate (12507.1211) and sodium chloride (1316591211) from Panreac and L-arginine (W381918) from Sigma.

For protein quantification using Green Fluorescent Protein rTurboGFP (FP552) from Evrogen and bovine serum albumin (A7906) were used as standards. The BCA kit (BCA1) from Sigma was employed as the colorimetric test for total protein quantification. Fluorescence intensity measurements ($\lambda_{excitation}$=485nm and $\lambda_{emission}$=535nm) were used for GFP quantification. SDS-PAGE was performed using β-mercaptoethanol (M6250), glycerol (G9012) from Aldrich, Tris (MB01601) from Nzytech, 30% acrylamide and bis-acrylamide solution 19:1 (161-0158), sodium dodecyl sulphate (SDS) (161-0416) from Biorad, ammmonia persulfate (APS) (9592.2), tetramethylethylenediamine (TEMED) (2867.3), Coomassie Blue R-250 from Roth, 2-butanol (1310891611) from Panreac. For gel staining, silver stain plus (161-0449) from Bio-Rad was used.

Peptide immobilization was done with sulfosuccinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate (sulfo-SMCC) (22322) from Thermo-scientific,

ethylenediamine tetraacetic acid (EDTA), N-(3-dimethylaminopropyl)-N'-ethyl-carbodiimide (EDC) (39391) and N-Hydroxysuccinimide (NHS) (130672) from Aldrich, L-cystein (16931) from Roth. The peptides were synthesized by GeneCust: Cys-PP-RWRWRW (PO#P110192) and RWRWRW (PO#P120175_5).

Bacterial growths were performed using Luria Broth (LB) (MB02802), ampicilin (MB0260) from Nzytech, agar powder (RM026) from Himedia. *Escherichia coli* K12 was kindly provided by Doutor Pedro Vidinha (FCT-UNL) and *Bacillus subtilis* 168 was kindly provided by Professor Isabel Sá Nogueira (FCT-UNL).

Protein production was performed using NZY5α competent cells (MB00401) and BL21(DE3) competent cells (MB00601) from Nzytech. The DNA fragments were cloned in the pET21-c expression vector by Geneart™. Protein induction was done using isopropyl β-D-1-thiogalactopyranoside (IPTG) (MB0210) from Nzytech. DNA purification was achieved using NZYMiniprep (MB01001) from NZYTech. Agarose gels were done using agarose (MB02702), Tris Base (MB01601) from Nzytech and bromophenol blue (A512.1) from Roth. Green Safe (MB08801) was used for agarose gel staining and NZYDNA Ladder III (MB0440) was used as molecular weight marker from Nzytech.

3.1.1. Equipment

For MNP synthesis and functionalization a mechanic stirrer, a Sonicator SilverCrest and an Incubator KS4000ic from IKA were used. The DLS measurements were performed in a Dynamic Zetasizer NanoZS from Malvern (ITQB, UNL). A Water-bath SHC 2000 from Scanvac was used for the Kaiser test. The screening assays were performed using a Centrifuge Scanspeed from Scanlab. Mini-Protean Tetra System from BIO-RAD was utilized for the electrophoresis SDS-PAGE gels.

Temperature control during large scale production of protein and microbial pre-inoculum and growth were performed using the IKA Incubator described above. A Hearaeus multifuge X39 Centrifuge from Thermo Scientific and a Beackman Couter Optima LE-80K Ultracentrifuga were used for crude extract purification. DNA concentration and purity ratios were calculated using a Nanodrop ND-1000 spectrophotometer (courtesy of Professor Alice Pereira, FCT-UNL).

The Microplate Reader Tecan Infinite F200 from Tecan was used to perform all the spectrophotometric and spectrofluorometric measurements. Colometric assays were analyzed using a Microplate 96-well Flat bottom Transparent Polystyrene (Starstedt) and in the fluorescence studies, BRAND Plates-Imunograde Tech Scientific were used.

Fluorescence microscopy assays were performed using an Olympus BX51 microscope (400X magnification), a U-MWB filter (λ_{exc}=460–490nm; λ_{em}=515–700nm) and an Olympus U-RFL-T lamp.

3.2. Methods

3.2.1. Molecular modeling methods

1. Structure template searching

The ExPASy translate tool (http://web.expasy.org/translate/) was used to translate the nucleotide sequences (5´to3') into the protein sequences. Two query sequences have been used: GFP and RK-GFP. The nucleotide sequence and corresponding amino acid sequence are shown in Figure 3.1.

Protein sequence identity was searched through alignment tools such as the NCBI BLASTp (http://blast.ncbi.nlm.nih.gov/Blast.cgi), ExPASy SIM (http://web.expasy.org/sim/) and EMBL-ebi ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/) [48].

a)
```
>GFP nucleotide sequence
ATGAGCAAAGGTGAAGAACTGTTTACCGGTGTTGTTCCGATTCTGGTTGAACTGGATGGT
GATGTTAACGGTCACAAATTTTCAGTTAGCGGTGAAGGTGAAGGTGATGCAACCTATGGT
AAACTGACCCTGAAATTTATCTGCACCACCGGTAAACTGCCGGTTCCGTGGCCGACCCTG
GTTACCACCTTTAGCTATGGTGTTCAGTGTTTTAGCCGTTATCCGGATCATATGAAACGC
CACGATTTTTTCAAAAGCGCAATGCCGGAAGGTTATGTTCAAGAACGTACCATCAGCTTT
AAAGATGACGGCAACTATAAAACCCGTGCCGAAGTTAAATTTGAAGGTGATACCCTGGTG
AATCGCATTGAACTGAAAGGCATCGATTTTAAAGAGGATGGTAACATTCTGGGCCACAAA
CTGGAATATAACTATAATAGCCATAATGTGTATATCACCGCAGATAAACAGAAAAACGGC
ATTAAAGCCAACTTCAAAATCCGCCATAACATTGAAGATGGTAGCGTTCAGCTGGCAGAT
CATTATCAGCAGAATACCCCGATTGGTGATGGTCCGGTTCTGCTGCCGGATAATCATTAT
CTGAGCACCCAGAGCGCACTGAGCAAAGATCCGAATGAAAAACGTGACCATATGGTGCTG
CTGGAATTTGTTACCGCAGCAGGTATTACCCATGGTATGGATGAACTGTATAAATGA

>GFP amino acid sequence
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL
VTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLV
NRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGSVQLAD
HYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK
```

b)
```
>RK-GFP nucleotide sequence
CGTAAACGTAAACGCAAACCGCCTCCGGATGATGATGATAAAGGTACCGGATCCATGAGC
AAAGGTGAAGAACTGTTTACCGGTGTTGTTCCGATTCTGGTTGAACTGGATGGTGATGTT
AACGGTCACAAATTTTCAGTTAGCGGTGAAGGTGAAGGTGATGCAACCTATGGTAAACTG
ACCCTGAAATTTATCTGCACCACCGGTAAACTGCCGGTTCCGTGGCCGACCCTGGTTACC
ACCTTTAGCTATGGTGTTCAGTGTTTTAGCCGTTATCCGGATCATATGAAACGCCACGAT
TTTTTCAAAAGCGCAATGCCGGAAGGTTATGTTCAAGAACGTACCATCAGCTTTAAAGAT
GACGGCAACTATAAAACCCGTGCCGAAGTTAAATTTGAAGGTGATACCCTGGTGAATCGC
ATTGAACTGAAAGGCATCGATTTTAAAGAGGATGGTAACATTCTGGGCCACAAACTGGAA
TATAACTATAATAGCCATAATGTGTATATCACCGCAGATAAACAGAAAAACGGCATTAAA
GCCAACTTCAAAATCCGCCATAACATTGAAGATGGTAGCGTTCAGCTGGCAGATCATTAT
CAGCAGAATACCCCGATTGGTGATGGTCCGGTTCTGCTGCCGGATAATCATTATCTGAGC
ACCCAGAGCGCACTGAGCAAAGATCCGAATGAAAAACGTGACCATATGGTGCTGCTGGAA
TTTGTTACCGCAGCAGGTATTACCCATGGTATGGATGAACTGTATAAATGA

>RK-GFP amino acid sequence
RKRKRKPPPDDDDKGTGSMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKL
TLKFICTTGKLPVPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKD
DGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIK
ANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLE
FVTAAGITHGMDELYK
```

Figure 3.1 - Nucleotide and translated protein sequence in fasta format of the proteins (a) GFP and (b) RK-GFP.

Identification of sequence associated homologous structures were retrieved from the RCSB Protein Data Bank (PDB; http://www.rcsb.org/pdb/).

2. Homology Modeling

Due to the lack of a crystallographic structure available for the protein sequence used in this work, we employed homology modeling protocols to obtain an accurate model. Most software require the amino acid sequence of the protein of interest and the PDB ID code of the chosen template to proceed with the modeling task. Four different modeling tools were used: the Automated Mode of SWISS-MODEL (Swiss Institute of Bioinformatics) (swissmodel.expasy.org/) [49]; Modeller (Laboratory of Andrej Sali, University of California) module of Chimera interface [50], with resource to the MultiAlign Viewer; Rosetta software package (Baker's Lab, University of Washington) [51]; I-TASSER (ZhangLab, University of Michigan) (zhanglab.ccmb.med.umich.edu/I-TASSER/).

Rosetta input files comprise the Rosetta fragment database (http://robetta.bakerlab.org/), the structure prediction file calculated by PSIPRED v3.0 [52] (http://bioinf.cs.ucl.ac.uk/psipred/), and the multi pair-wise alignment by HHpred (http://toolkit.tuebingen.mpg.de/hhpred) [53]. In the Robetta options no homologues were excluded and due to lack of experimental information, no NMR information was provided (e.g. chemical shifts, NOE constraints, dipolar constraints). The filtering options included masking low complexity regions and the multiple sequence alignment options were marked as to find new close and distant relatives. The HHpred options were used as default.

I-TASSER webserver was used with no additional restraints defined and no templates were chosen to be excluded from the I-TASSER template library.

2.1. Homology models evaluation

The inaccuracy of homology models might result from wrong packing of side chains, or assignment of backbone conformation in structurally undefined regions [54]. Evaluation of the homology modeling accuracy was monitored by the criteria described below. The homology models that were chosen to proceed with the modeling task were the ones that had a better global evaluation.

- **ANOLEA** (*A*tomic *NO*n-*L*ocal *E*nvironment *A*ssessment) (http://swissmodel.expasy.org/workspace/index.php?func=tools_structureassessment1) It is a packing quality score and gives information on the non-local environment energy for each residue. Negative values represent favorable and positive values represent unfavorable energy environment for a given protein residue [55].

- **GROMOS**

  (http://swissmodel.expasy.org/workspace/index.php?func=tools_structureassessment1)

  Evaluates each residue based on empirical force field energy and is applied to the analysis of obtained conformations [49]. Negative values represent favorable and positive values represent unfavorable energy environment for a given aminoacid [55].

- **QMEAN** (*Q*ualitative *M*odel *E*nergy *AN*alysis)

  (http://swissmodel.expasy.org/qmean/cgi/index.cgi)

  Scores homology models according to the agreement of the predicted and observed secondary structure and solvent accessibility [56]. It is a global indicator of quality of the model and it is given as the standard deviation from the mean value. The smaller the value, the better is the model evaluated [57].

- **SOLVX**

  (http://ekhidna.biocenter.helsinki.fi/solvx/start)

  Computes the solvation profile for a protein structure [58]. The solvation free energy of each residue is determined considering its intrinsic solvation free energy and the fraction of its non-polar and polar surfaces that are solvent-accessible [58]. An overall solvation preference should be below zero, indicating a well-packed structure [59]. The more negative the value, the better is the model.

- **RMSD** (*R*oot *M*ean *S*quare *D*eviation) corresponds to the deviation of each aligned pair of the backbone atoms between each homology model and the template structure. [60]. It was calculated using the superimposition function implemented in PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC.

- **VERIFY3D**

  (http://nihserver.mbi.ucla.edu/Verify_3D/)

  Analyzes the compatibility of an three-dimensional model with its aminoacid sequence [55]. Each residue is assigned a structural class based on its characteristics (buried area of the residue, the fraction of side-chain that is covered by polar atoms, and the secondary structure [61]). Each residue is scored according to a protein database, obtaining a 3D-1D profile score, ranging from -1 (bad score) to +1 (good score) [55].

3. Molecular Dynamics

Molecular dynamics (MD) simulations were performed with the Gromacs v.4.5 simulation package [62] on a Sun Grid Engine mounted on a high performing computer cluster with 188 cores. The best evaluated homology models were taken as starting structures. Each system was solvated in a truncated octahedral box filled with explicit water molecules, keeping a distance between the protein and the box edges of 12Å. The system electroneutrality was guaranteed by adding the appropriate number of counter ions. The system was simulated using the united-atom force field GROMOS53A6 in an isothermal-isobaric ensemble (NPT), coupled to the Berendsen barostat with a reference pressure of 1.0bar and a coupling time constant of 0.6ps [63], as well as to the V-rescale thermostat with a reference temperature 300K and a coupling time constant of 0.1ps [64]. The LINKS algorithm was applied to constrain all H-bonds [65] and the electrostatic term was described by using the particle mesh Ewald algorithm, as implemented in GROMACS software. An integration time step of 2fs was used and the resulting coordinates were recorded on the trajectory every 2ps.

The MD protocol comprised three main steps. First, the energy of the system was minimized through a steepest descent followed by a conjugate gradient algorithm of 2000 and 1000 steps, respectively. Second, each individual system was equilibrated in three subsequent steps for 100ps each, and a decreasing force constant for positional restraint of 1000, 100 and 10kJ/mol was applied to all protein heavy atoms, in order to relax the system gradually. Third, the production phase comprises a fully free trajectory with a certain length. Remaining parameters details are described elsewhere [16]. All jobs ran in parallel in the SGE cluster in with 8cores per job.

1.2. Ligand Topology Parameterization

New ligand topologies have to be generated, ensuring a full consistency and compatibility with the GROMOS53a6 force field. For this purpose, the online server PRODRG (http://davapc1.bioch.dundee.ac.uk/prodrg/) [66] was used to generate new ligand topology files.

1.3. Interactions/distance analysis criteria

Analyzing the resulting MD trajectories allows the characterization of interactions between key atom pairs. For hydrogen-bond interactions, the threshold criteria assumed an angle cutoff of 20° and a donor-acceptor distance between 2.0-3.0Å [67]. For hydrophobic interactions, a threshold of 6.4Å between adjacent aromatic rings [68] was considered. A particular interaction was considered significant if it persists 10% or more of the simulated MD time.

4. Molecular Docking

Ligands' coordinates were sketched on ChemBioOffice 12.0 (CambridgeSoft®) software and were docked into the target GFP and tagged-GFP proteins, using the Autodock v.4.2 software package [69]. The Gasteiger partial charges and AutoDock atom types were automatically assigned to the receptor and ligand pdb coordinate files through the python scripts distributed with the software version. A blind docking approach was followed [70]. A Lamarckian genetic algorithm was used with an initial population of 150 conformations, a maximum of 2500000 energy evaluations, a maximum number of 27000 generations, a 0.02 mutation rate, and a 0.8 crossover rate, a total of 256 independent solutions, a grid size of 130x100x100 (x,y,z) and a spacing of 0.5Å. The center of the grid was defined automatically by the software as the system center mass.

Non-specified settings were assumed by default. A RMSD cut-off value of 2.0Å was used in the automated cluster analysis of docking results.

4.1. Docking solution analysis criteria

Docking results were assessed taking into account four criteria: (1) energy criteria – the top-scoring docking solutions with the best estimated binding free energy were selected; (2) geometry criteria – docking solutions leading to ligand interactions in the protein inner cavities were discarded; (3) Ka criteria - the estimated affinity constant (Ka) might be comparable to the Ka experimentally determined [47]; (4) anchoring point criteria – only solutions where the anchoring point to the solid support was exposed to solvent were selected.

3.2.2. Experimental methods

1. Synthesis and characterization of the magnetic supports

1.1. Synthesis of coated-MNP

Dextran- and PEG-coated MNP were synthesized by the co-precipitation method, using $FeCl_3$ and $FeCl_2$ salts with a molar ratio of Fe^{2+}/Fe^{3+} of 0.5 in a basic and inert environment. A solution of 25% ammonium hydroxide in deionized water was purged with N_2 gas for 30minutes. A freshly prepared iron solution (25ml) and the coating solution (80mg/ml, 25ml) were added in a dropwise manner, to a total volume of 200ml. The reaction continued under an inert atmosphere during 2h. The resulting MNPs were washed with distilled water several times with the aid magnetic separation.

1.2. Amination of the magnetic supports

The supernatant of the prepared solution of MNPs is removed and the particles were washed twice with 200ml 50%ethanol:50%water and sonicated during 5min. The MNPs (10mg/ml) were ressuspended in 20ml APTES (9.4g/ml), 1ml of acetic acid and 179ml ethanol:water (50:50). The mixture was incubated for 1h at 70°C with constant shaking. Afterwards the solution was washed first with ethanol and then with water.

1.3. Quantification of amines by Kaiser Test

The Kaiser test is a colorimetric test based on the reaction of ninhydrin with primary amines. 50µl of each reagent were added to 1ml of the aminated-supports: 80%phenol in ethanol (w/v), 2% 0.001M aqueous solution of potassium cyanide in pyride (v/v) and 5%ninhydrin in ethanol (w/v). Samples were incubated for 5min at 100°C, and the absorbance was measured at 560nm. Calibration curves were obtained using standard solutions of glycine (0-5µmol/ml). In this work we have achieved amination contents from 140µmol NH_2/g support to 428 µmol NH_2/g support.

1.4. Characterization of magnetic supports

To estimate the concentration of the MNP solution containers were weighted before and after adding 0.25ml of the MNP solution. The solutions were dried overnight at 60°C and the concentration (mg/ml) was determined.

All magnetic supports have been characterized by dynamic light scattering (DLS) and by zeta potential, using 0.05mg/ml solution in deionized water (pH5.80).

2. Production of crude extracts containing GFP and RK-GFP

2.1. Preparation of Luria Broth (LB) liquid and LB agar medium

LB medium (25g LB/l distilled water) and LB agar medium (25g LB/l and 15g agar/l) were autoclavated at 120°C for 20min. For LB agar containing ampicilin, the liquid was cooled down at room temperature (RT) and a final concentration of 100µg/ml of ampicilin was added under sterile conditions.

2.2. Transformation of the plasmids pAP001 (GFP) and pAP006 (RK-GFP) in NZY5α competent cells

NZY5α competent cells were used to replicate the plasmids. The fragments DNA shown in Figure 3.1a and Figure 3.1b were cloned into pET-21c, resulting in pAP001 (expressing RK-GFP)

and pAP006 (GFP). 60µl of cells and 2µl of plasmid were mixed and incubated in ice for 30min. Afterwards the solution was subjected to a heat shock during 30s at 42°C, and then incubated in ice for 2min. 940µl of LB medium were added to the mixture and it was incubated at 37°C for 1h with constant shaking (210rpm). Afterwards, 50µl of the transformed cells were spread on LB agar ampicilin plates, which were incubated at 37°C overnight. The negative control was performed by not adding any plasmid and the positive control was performed by adding 1µl of pNZY28 plasmid.

2.3. Isolation and purification of plasmid DNA (pDNA)

This procedure was carried out with the DNA isolation NZYMiniprep kit. On the previous day, the pre-inoculum was performed by using test tubes containing 6ml of LB medium, 0.1µg/ml of ampicilin and one colony of the corresponding freshly transformed plate (obtained in 2.2), which was incubated overnight at 37°C at 210rpm. The resulting cultures were centrifuged for 2min at 11000×*g*, and the supernatant was removed. The cell pellet was ressuspended in 500µl of buffer A1/RNAse by vortexing. 500µl of buffer A2 was added and the solution was mixed and it was incubated at RT for 4min. 600µl of buffer A3 was added and gently mixed. Afterwards the solution was centrifuged for 10min at 11000×*g* at RT. The supernatant was loaded into a NZYTech spin column placed inside a 2ml collecting tube. It was centrifuged 2min at 11000×*g* and the flow-through was discarded. The spin column was washed by adding 500µl of buffer AY, centrifuged 2min at 11000×*g*, added 600µl of buffer A4 and centrifuged again. The flow-through was discarded, and the column is centrifuged again for 3min at 11000×rpm. For the pDNA elution, the column was placed inside a new collecting tube and 30µl of deionized warm water was added to the column and incubated at 65°C for 1min. The column was centrifuged for 2min at 11000×*g* and the flow-through corresponded to the first elution of pure pDNA. A second elution with 50µl of deionized warm water was performed in the same conditions. Finally the plasmid concentration was determined.

2.4. Evaluation of pDNA through agarose gel electrophoresis

0.8% agarose gel was prepared by dissolving the agarose in 100ml of TAE pH8.5 (40mM Tris-base, 20mM glacial acetic acid, 1mM EDTA) and the solution was poured into the casting frame with the comb inserted. The agarose gel was polymerized for 1h and transferred to the running module. The tank was filled with TAE and the samples were applied in each well. The running conditions were set as 100V for 1h. Afterwards the gel was immersed in the staining solution (100ml TAE, 10µl Green Safe) for 30min and the gel was photographed.

The solutions for loading were prepared as followed: 2µl of sample or 5µl of the molecular weight marker and 5µl of loading buffer (65% sucrose, 10mM Tris-HCl, 10mM EDTA, 0.3% bromophenol blue).

2.5. Transformation of BL21(DE3) competent cells

BL21(DE3) competent cells were used for protein expression of T7 RNA polymerase-based systems. The same protocol as in 2.2 was followed, using the pDNA isolated in 2.3.

2.6. Protein expression in BL21(DE3) competent cells

The previous day, a pre-inoculum was performed as described in 2.2 with the resulting transformant colonies from 2.5. After 8-9h of incubation at 37°C (210rpm), 1ml of the culture was used to inoculate 50ml of LB medium containing 0.1µg/ml ampicilin. The culture was incubated overnight at 37°C (210rpm). The next day, 10ml of the resulting culture was used to inoculate 1L of LB medium containing 0.1µg/ml ampicilin, and it was incubated at 37°C with constant shaking. The culture growth was monitorized by optical density at 600nm. Once the cultures reached OD_{600nm} 0.6-0.8 the protein expression is induced by adding 1ml of IPTG (1M). The protein expression was monitorized by GFP fluorescence, OD_{600nm} and SDS-PAGE analysis.

2.7. Cellular fractionation

After 22h of protein expression, the cell culture was centrifuged (20min, 18000×*g* at 4°C). The ressuspended cells in PBS were subjected to three freeze/thaw cycles. Cell lysis was performed by a mechanical process (French Press), with a maximum of 20000psi of pressure applied. Afterwards, DNAseI was added, the sample was incubated for 30min in ice and centrifuged (15min, 18000×*g* at 4°C). The resulting pellet was ressuspended in 15ml of PBS and stored at -20°C. The supernatant was further ultracentrifuged at 180000×*g* for 1h30 at 4°C. The protein expression was monitorized by GFP fluorescence, OD_{600nm} and SDS-PAGE analysis.

2.8. Sample preparation for SDS-PAGE analysis

In order to evaluate the amount of protein expression, sample volumes were normalized.

For the time course samples the normalization was performed according to the ratio $1.2/OD_{600nm}$ for each sample. The corresponding volumes were centrifuged 5min at 6800×*g* and the supernatant was discarded. All samples were ressuspended in 20µl of sample buffer and boiled for 2min. 15µl of each sample was used to apply in the 12.5% acrylamide/bisacrylamide gel (see section 4.1).

For the fractionation samples, the normalization was performed relatively to one initial volume of one of the samples. 5µl of sample buffer is added to each sample and it was boiled. 10µl were used to apply in the gel (see section 4.1).

3. Immobilization of specific affinity ligands onto magnetic supports

The solid phase synthesis of affinity ligands was based on the Ugi mechanism [15] which is summarized in Figure 3.2. The Ugi mechanism is a one-pot four-component reaction, with the combination of an aldehyde, an amine, an isonitrile group and a carboxylic acid [15]. In this work, the aldehyde coumpound used was glutaraldeyde and the isonitrile group was provided by isopropyl cyanide.

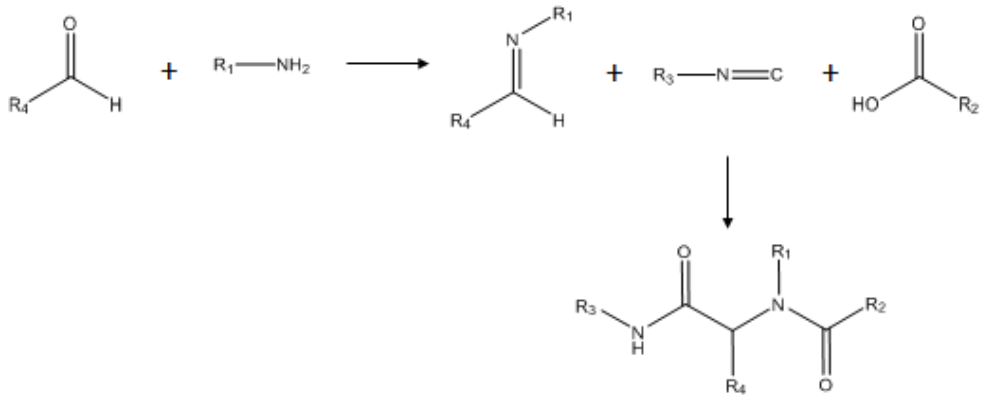
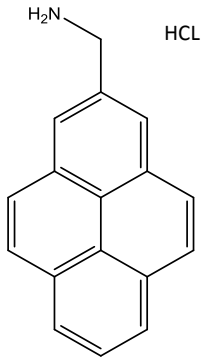
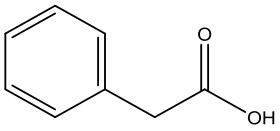
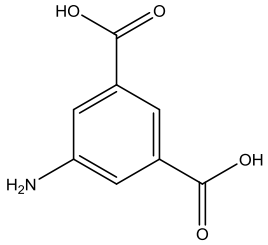
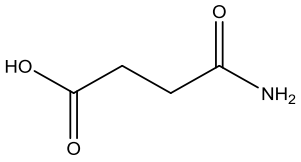
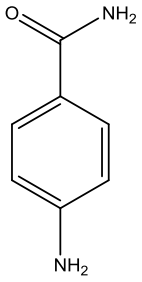
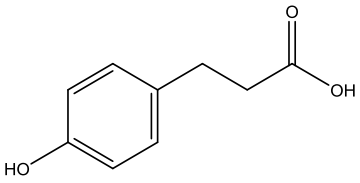


Figure 3.2 - Ugi reaction mechanism.

The amine and carboxylic acids used for this reaction are shown in Table 3.1 and correspond to the compounds necessary to synthesize the lead ligands for GFP and RK-GFP, as well as the ligands showing the worst binding affinity (employed here as negative controls) [47].

Table 3.1 - Amine (on the left) and carboxylic acid (on the right) compounds used for the Ugi reaction, for ligands LA-A4C7, NA/LR-A7C1 and NR-A5C2, respectively.

Amine compounds	Carboxylic acid compounds
A4 1-pyrenemethylamine hydrochloride  	C7 phenylacetic acid 
A7 aminoisophtalic acid  	C1 succinamic acid 
A5 4-aminobenzamide  	C2 3-(4-hydroxyphenyl)propionic acid 

The aminated magnetic supports were firstly fucntionalized with aldehyde groups. After removing the water, 30ml of 10mg/ml aminated dextran-coated MNPs (MNP_DEX_NH_2) were ressuspended in 5% (v/v) glutaraldeyde in a basic environment (5.4ml NaOH 1M). The mixture was incubated 1h at 30°C. Afterwards the solution was washed five times with 30ml of deionized water.

The next step was to add the corresponding amine compound dissolved in 10ml methanol in a 5-molar excess to the amines of the support. Amines with carboxylic functionalities (e.g.

A7) or protected by HCl (e.g. A4) were pre-treated with NaOH, in the same molar amount as the species in solution. The mixture was incubated for 2h at 60°C. Afterwards 5-molar excess of the correspondent carboxylic acid in 10ml methanol and 5-molar excess of isopropyl cyanide were added to the solution and the mixture was incubated for 48h at 60°C.

The particle slurry was washed twice in 100% methanol, twice in 50%DMF:50%methanol, once with water, twice with 0.2M NaOH in 50% isopropanol and finally twice in distilled water.

3.1. Screening crude extracts with the functionalized magnetic supports

The methodology is summarized in Figure 3.3. In each assay, non-functionalized MNP_DEX_NH_2, MNP_DEX_NH_2 functionalized with the best ligand and MNP_DEX_NH_2 functionalized with the worst ligand were tested. Previously the supports are washed with 1ml of the following solutions: 0.2M NaOH in 50% isopropanol, deionized water, PBS (10mM sodium phosphate, 150mM NaCl, pH7.4). All tests were performed in duplicates.

The loading solution was incubated for 15min at 4°C with each magnetic support. The supernatants were collected (fall through - FT) and the supports were washed with PBS by means of centrifugation (5min, 1850×*g*) and an external permanent magnet.

Each magnetic support was then divided into two aliquots, on which two different elution conditions were tested. The best two elution conditions were determined in a previous study [47]: for GFP purification, 0.1mM glycine-NaOH pH9 and 0.1mM glycine-NaOH pH9 50% (v/v) ethylene glycol; for RK-GFP purification, PBS pH 7.4, 500mM arginine and 0.1M glycine-NaOH pH 11, 150mM NaCl. Each sample was washed five times with 1ml of the elution buffer.

All supernatants were recovered and stored at 4°C and protected from light until quantification. Each sample was quantified through GFP fluorescence and by BCA colorimetric assay. An example of each duplicate was also analyzed by SDS-PAGE.

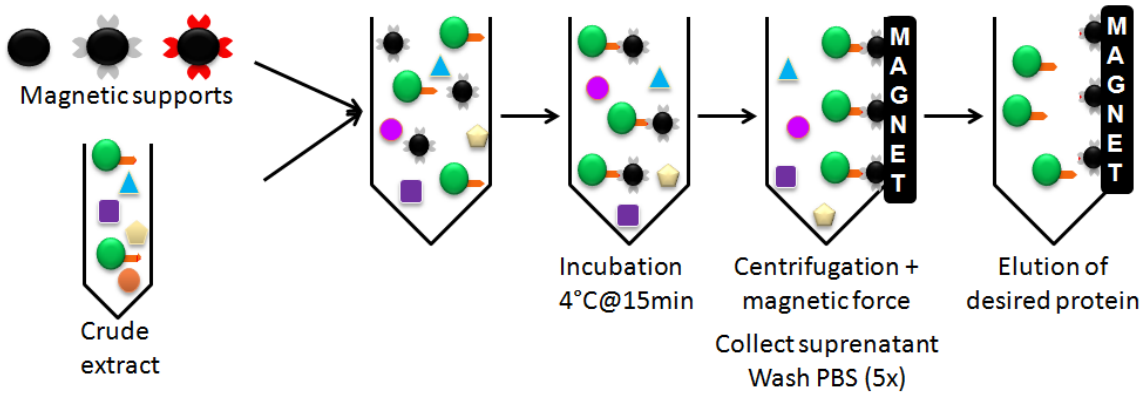


Figure 3.3 - Schematic representation of the methodology followed to test the magnetic supports with the different crude extracts. Each assay was performed using MNP_DEX_NH_2, MNP_DEX_NH_2 functionalized with the best ligand (in grey) and MNP_DEX_NH_2 functionalized with the worst ligand (in red).

3.2. Static partition equilibrium studies

Previously the needed volume of MNP solution is washed twice in 0.2M NaOH in 50% isopropanol and twice in PBS. A volume of 500µl from different concentrations of crude extract containing GFP or tagged-GFP in PBS were incubated with the 500µl of MNPs (40mg/ml) functionalized with the corresponding affinity ligand. The interaction was promoted over 12h at 4°C to achieve chemical equilibrium. Afterwards the supernatant was collected and the fluorescence intensity was measured. The adsorption phenomenon followed a Langmuir isotherm and the experimental data was fitted accordingly with OriginPro (v8.5.1).

3.3. Sample preparation for SDS-PAGE analysis

5µl of sample and 5µl of sample buffer were boiled for 2min and all sample was applied to the previously polymerized 12.5% acrylamide gel (see section 4.1).

3.4. Data analysis: selectivity, percentage of recovery, yield and purity

The results obtained for both systems (LA-A4C7/GFP and LR-A7C1/RK-GFP) have been analyzed according to selectivity (Equation 3.1), percentage of recovery (Equation 3.2) and yield (Equation 3.3). The purity of the eluted fraction has been calculated by the ImageLab4.1. software.

$$Selectivity\ (\%) = \frac{amount\ of\ GFP\ bound\ (mg)}{amount\ of\ total\ protein\ bound\ (mg)} \times 100 \qquad (Equation\ 3.1)$$

$$Recovery\ (\%) = \frac{amount\ of\ GFP\ eluted\ (mg)}{amount\ of\ GFP\ bound\ (mg)} \times 100 \qquad (Equation\ 3.2)$$

$$Yield\ (\%) = \frac{amount\ of\ GFP\ eluted\ (mg)}{amount\ of\ GFP\ loaded\ (mg)} \times 100 \qquad (Equation\ 3.3)$$

3.5. Fluorescence studies with ligand LA-A4C7

The amine of LA-A4C7 is 1-pyrenemethylamine which has a fluorescence signal in the same wavelength as GFP ($\lambda_{excitation}$=485nm and $\lambda_{emission}$=535nm). This study had two objectives: (1) confirm the presence of the pyrene molecule in the MNP functionalized samples, and (2) confirm that there is no pyrene (ligand) leaching during the testing of the functionalized magnetic supports with crude extracts.

For the first goal, fluorescence microscopy was applied for a qualitative analysis of the magnetic beads before and after functionalization with the Ugi reaction employing 1-pyrenemethylamine as amine. Briefly, the MNP solution (10mg/ml) was analyzed by fluorescence microscopy by placing the solution in a glass microscope slide for observation.

For the second goal, the protocol described in Figure 3.3 was used, using the same volume of PBS instead of the crude extract and 40mg/ml of magnetic support. The fluorescence intensity of the collected samples was analyzed by spectrofluorimetry on 96-well plates.

4. Protein quantification methods

4.1. SDS-PAGE preparation and staining

12.5% acrylamide/bisacrylamide SDS-PAGE gels were used in this work. The volumes of each solution used to prepare the gels are shown in Table 3.2. The running gel was prepared first and the solution was transferred to the glass plates of the casting frame. 1ml of 2-butanol was added on the top of the gel solution, and the gel was polymerized for 30min. Afterwards the 2-butanol was removed and the gel was washed with distilled water. The 5% acrylamide stacking gel was prepared and the mixture was added on top of the previous one. The comb was inserted and the gel was polymerized for 30min. The electrophoresis buffer (0.25M Tris-Base, 1.92M glycine, 0.1% SDS pH8.3) was added to the tank and the polymerized gel was introduced in the running module. The samples prepared previously were applied in each well. The running conditions were set as 150V for 1h.

Table 3.2 - Volumes necessary to prepare a 12.5% acrylamide gel for SDS-PAGE.

Stacking gel	
	V (ml)
Solution I (3M Tris Base pH 8.8)	0.75
Solution III (30% acrylamide and bis-acrylamide solution 19:1)	2.08
10% SDS	0.05
Distilled water	2.1
APS	0.038
TEMED	0.0025
Running gel	
Solution II (0.5M Tris Base pH 7.0)	0.450
Solution III	0.3
10% SDS	0.018
Distilled water	0.94
APS	0.0135
TEMED	0.002

4.1.1. Blue-Coomassie staining

The SDS-PAGE gel was transferred into the staining solution (1g Coomassie Blue R-250, 15ml glacial acetic acid, 90ml methanol and 95ml distilled water) for 30 min. Afterwards the gel was destained overnight with the destaining solution (75ml glacial acetic acid, 450ml methanol and 475ml distilled water).

4.1.2. Silver staining

The gel was transferred into the fixative enhancer solution (50%methanol, 10%acetic acid, 10%fixative enhancer concentrate, 30%distilled water), and incubated for 20min with gentle agitation. Afterwards the gel was rinsed twice in 400ml of distilled water for 10min. The staining solution (5ml silver complex soluion, 5ml reduction moderator solution and 5ml image development reagent, 50ml development accelerator solution) was prepared within 5min of use, and the gel incubated for 20min. The final step involved the incubation of the gel in a 5% acetic acid solution to stop the previous reaction. The gels were rinsed in deionized water.

4.2. Quantification of GFP and total protein

The GFP fluorescence allows the quantification of GFP and tagged-GFP by pipetting 200µL of each samples in a 96-well microplate and measuring the fluorescence intensity ($\lambda_{excitation}$=485nm and $\lambda_{emission}$=535nm). A calibration curve was obtained using a pure GFP solution (10^{-6}-10^{-1}mg/ml). PBS was used as blank.

The total protein was quantified through the colorimetric BCA assay. The BCA reagent was prepared by adding 50 parts of reagentA (bicinchoninic acid, sodium carbonate and sodium bicarbonate in 0.1N NaOH pH11.25) with 1 part of reagentB (4% (w/v) copper(II) sulfate pentahydrate). 25µl of each samples were added to each well in a transparent 96-well microplate, followed by the addition of 200µl the freshly prepared BCA reagent. The microplates were incubated at 37°C during 30min and the absorbance at 560nm was measured. A calibration curve was obtained using a solution of BSA (0-1mg/ml).

5. Preparation and testing of magnetic antimicrobial devices

5.1. Immobilization of Cys-$(RW)_3$ antimicrobial peptide by sulfo-coupling

The sulfo-coupling strategy (Figure 3.4) was used to immobilize the peptide Cys-PP-RWRWRW onto the magnetic supports (dextran-, silica-dextran- and PEG-coated MNP) using sulfo-SMCC as an amine-to-sulfhydrl crosslinker. Each support was previously washed five times with conjugation buffer (10mM phosphate, 150mM NaCl, 1mM EDTA, pH7.2 purged with N_2) after which the supernatant was discarded. Five-molar excess of sulfo-SMCC to the amines of the support in 1ml distilled water was added to the support and incubated for 30min at RT with constant shaking (200rpm). One-molar excess of peptide to the amines was added, the mixture was incubated for 1h at RT and it was washed 10 times with 1ml of conjugation buffer. 100mM of cysteine is added to the particle slurry. The mixture was incubated 1h at RT with constant shaking. Between reactions the mixture was always washed several times with conjugation buffer and the supernatants were collect and further analyzed.

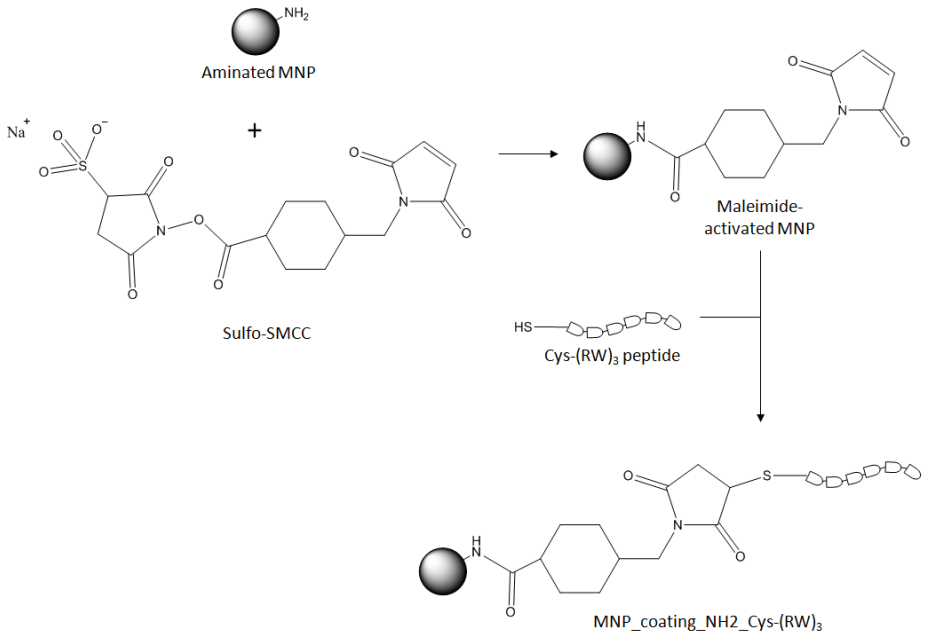


Figure 3.4 - Schematic representation of the functionalization of aminated coated-MNP with the peptide Cys-PP-RWRWRW by Sulfo coupling.

5.2. Immobilization of $(RW)_3$ antimicrobial peptide by EDC-coupling

The EDC-coupling (Figure 3.5) was used to immobilize the peptide RWRWRW onto the magnetic supports (PEG-coated MNP) via carbodiimide activation. Each support was previously washed with PBS buffer five times and the supernatant was discarded. 1ml of distilled water was added to the magnetic support after which all reagents (peptide, EDC and NHS) were added simultaneously with a 1:1 molar excess to the amines of the support. The mixture was incubated for 2h at RT with constant shaking. Afterwards, the support was washed with PBS buffer and the supernatants were collect to quantify the immobilization yield.

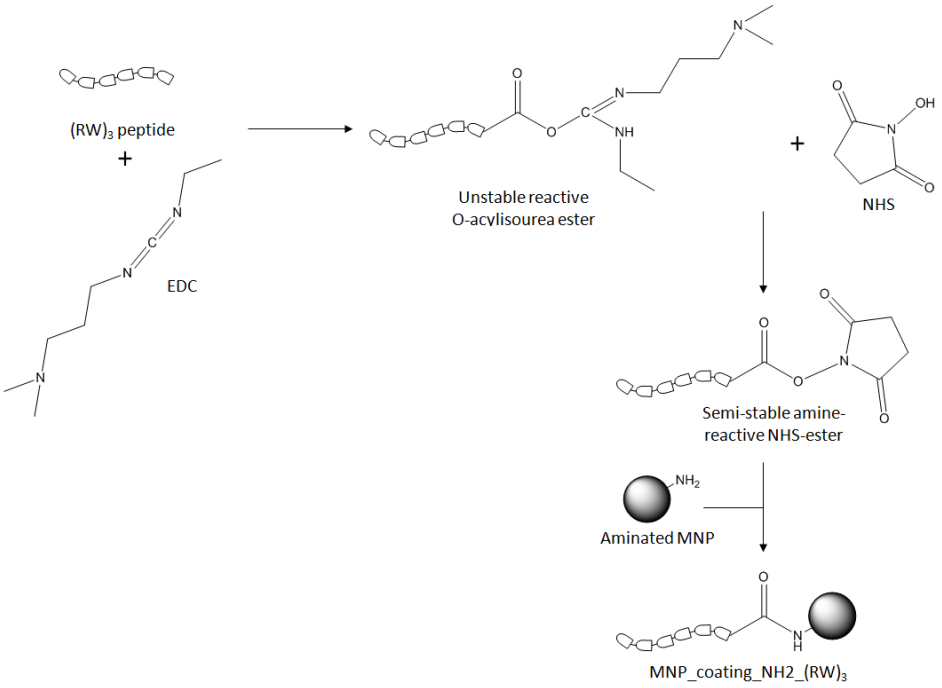


Figure 3.5 - Schematic representation of the functionalization of aminated coated-MNP with the peptide RWRWRW by EDC coupling.

5.3. Quantification of peptide immobilization and peptide characterization

The percentage of immobilization was calculated by Equation 3.4, by taking into account the measurement of the fluorescence of the initial peptide solutions (total protein added) and in the collected supernatants during the washes (total protein washed). The tryptophan present in the peptide can be measured through fluorescence intensity using $\lambda_{excitation}$=280nm and $\lambda_{emission}$=340nm. Calibration curves were obtained with the respective pure peptide (0-1mg/ml).

$$\%\ immobilization = \frac{Total\ peptide\ added\ (mg) - Total\ peptide\ washed(mg)}{Total\ peptide\ added\ (mg)} \times 100 \quad (Equation\ 3.4)$$

Furthermore, both peptides described above have been characterized by the zeta potential (0.05mg peptide/ml).

5.4. Antimicrobial assays with the antimicrobial magnetic devices

The strains used to test the antimicrobial magnetic devices were *E. coli* K12 and *B. subtilis* 168. Two methods have been tested (Figure 3.6): (a) shaking flasks for a total of 5mL of culture, and (b) 96-deep well titer plate for a total of 2mL of culture. The pre-inoculum of the bacteria were prepared by growing strains in LB medium at 37°C with constant shaking (300pm) overnight. The next day the cultures were diluted until reaching OD_{600nm} of 0.045. The

cultures were then incubated (37°C, 300rpm) and when the cultures reached an OD_{600nm} between 0.1-0.15, 10% volume/volume of magnetic support and/or antimicrobial agent was added to the culture. The OD_{600nm} was recorded each 30min until the end of the assay (the MNPs were removed from solution using an external permanent magnet recording the absorbance). All assays were carried out in triplicates.

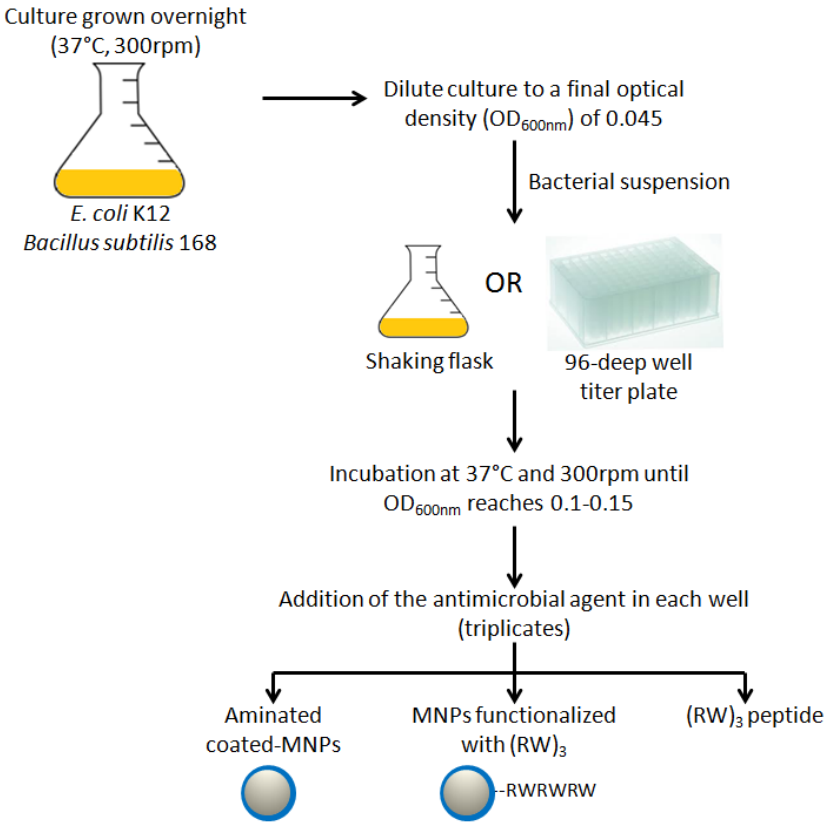


Figure 3.6 - Schematic representation of the antimicrobial assay.

4. *IN SILICO* STUDY OF AFFINITY PAIRS FOR PROTEIN PURIFICATION

4.1. Objective

Experimental techniques for the determination of three-dimensional structure proteins have contributed for the deposition of over 94000 protein structures in the PDB [71]. However the determination of the structure of all protein-ligand complexes experimentally at high resolution is still an impossible task [71]. Hence, reliable computational methods are of increasing importance.

In this work, the aim was to unveil the possible binding location and main interactions of the affinity pairs previously found [47], namely LA-A4C7 with the protein GFP and ligand LR-A7C1 with the protein RK-GFP. For that purpose we have combined homology modeling, molecular dynamics and automated docking to predict the docking location and main interactions between the affinity pairs.

4.2. *In silico* study of affinity ligands for the purification of GFP

The first step to structurally characterize the system developed in [47], was to have an accurate structure of the GFP used in this work.

GFP was first isolated from the jellyfish *Aequorea victoria*. The fluorescent properties of this protein make it a perfect candidate to be used as a reporter for recombinant gene expression [72] and it has been used to follow the expression of recombinant proteins in *E. coli* [47]. The solution's fluorescence is directly proportional to the amount of protein being expressed and it is a reliable method for protein quantification [73].

The GFP is a 238 aminoacid polypeptide [74] constituted by a β-barrel fold of eleven strands forming an almost perfect cylinder of about 25Å in diameter and 40Å tall [75]. The protein is capped at both ends, isolating the central chromophore from solvent [74]. A graphical representation of the GFP secondary structure is shown in Figure 4.1.

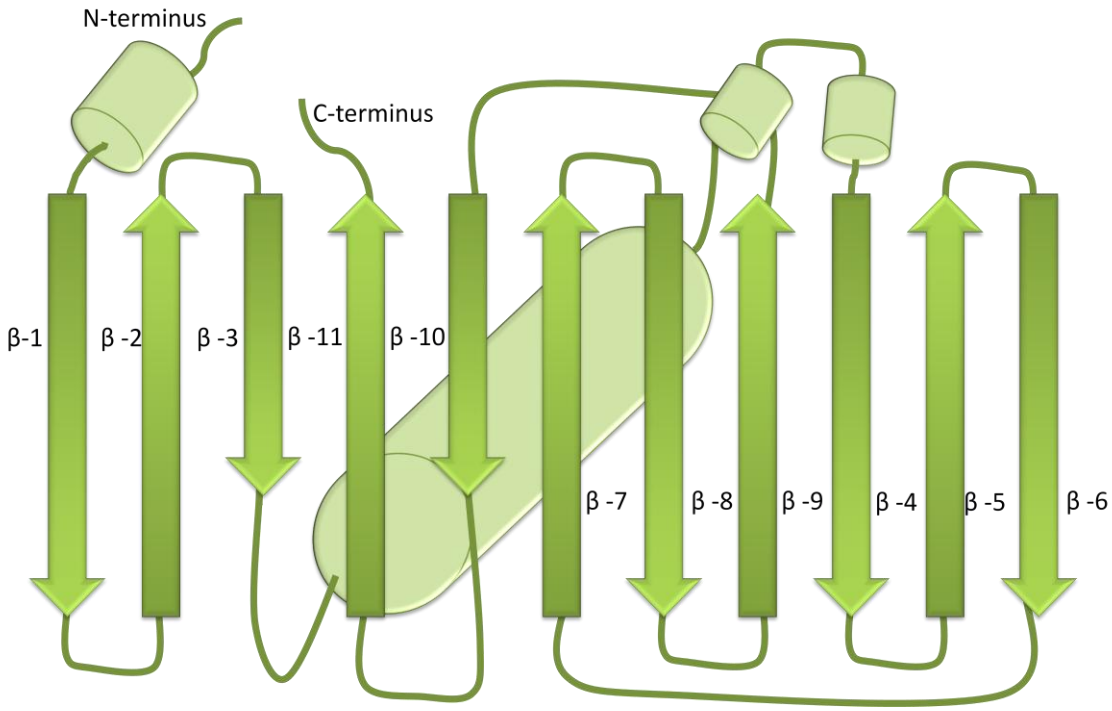


Figure 4.1 – Schematic **r**epresentation of *Aequorea victoria* GFP secundary structure. The β-sheets are numbered and represented as dark green cilinders with an arrow pointing towards the C-terminus. The α-helices are represented as light-green cilinders. The loops that link the β-sheets are represented as green full lines.

There has been an effort to develop engineered variants of GFP with various optimized experimental properties (see Shaner *et al* and references within [76]). Until 2011, there were 250 different structures of GFP deposited in the PDB [77].

4.2.1. Homology modeling

The first system modeled was GFP (non-tagged system). The nucleotide and aminoacid sequences corresponding to the GFP used in this work are shown in Figure 3.1a, in the Methods section.

The first step was to identify high identity-sequences for template selection. NCBI BLASTp was used to search for homologous templates. ExPASy SIM and EMBL-ebi ClustalW were used to perform the alignments of protein sequences. The parameters were used as default. The results obtained are shown in Figure 4.2 and summarized in Table 4.1.

The GFP aminoacid sequence used was confirmed within the GFP superfamily. Despite the variety of GFP structures deposited in the PDB, none had a 100% protein sequence identity with the target sequence. These differences are based on some point mutations (see Figure 4.2).

Table 4.1 - Top ranking sequence alignment, using the GFP sequence of this work as query in the search.

PDB ID	Resolution (Å)	Sequence Identity (%)
4GES	1,23	99,2
1GFL	1,90	98,7
1YHI	1,90	98,3
1QYO	1,80	98,3
1EMG	2,00	98,3
1EMA	1,90	98,3
2WUR	0,90	97,9

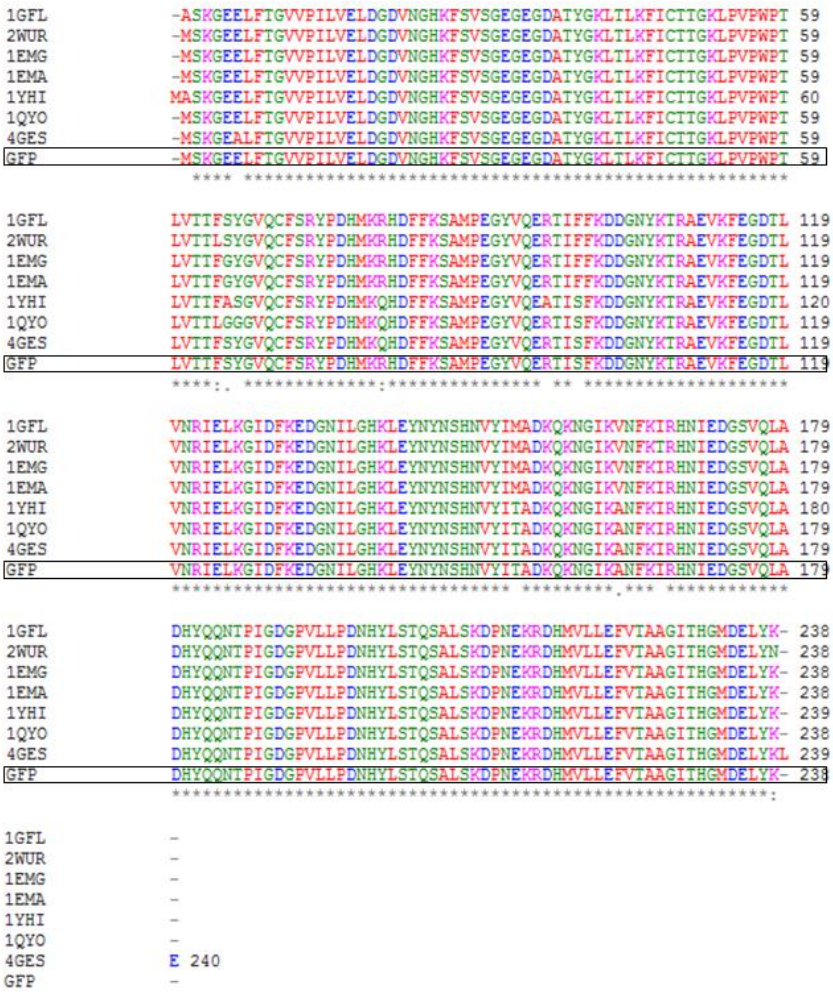


Figure 4.2 - Alignment results obtained by the webserver tool ClustalW (EMBL-ebi) using the protein sequence of GFP as query. * (asterisk) indicates positions which have a conserved residue; : (colon) indicates coservation between groups of strongly similar properties; . (period) indicates conservation between groups of weakly similar properties. Gap indicates no conservation between residues.

As there was no available structure for the target protein sequence, other strategy had to be used. Homology modeling refers to the construction of a three-dimensional model of a protein structure (target) based on the sequence similarity with family-related ones, from which the structure is known (template) [49]. This strategy has been used successfully by other authors to generate models of protein structures that are not available (e.g. [78,79]).

Protein homology modeling is characterized by four general steps: (1) identifying suitable templates, (2) pairwise sequence alignment between the amino-acid sequences of the target and the putative templates, (3) building the model, and (4) evaluating the quality of the results [80].

Poor target-template alignments, structural flexibility, low template quality or inaccuracies introduced by the modeling program can influence significantly the accuracy of the excepted model [49]. If the resulting model is not satisfactory, some or all of the steps can be repeated iteratively in order to obtain a better one [81].

Thus, homology modeling can only generate a model as good as the chosen template, which makes the availability of homologous templates a critical bottleneck of this approach [49]. Insertions, deletions, among other genetic modifications might cause significant structural changes and should be avoided by a careful choice of the template [49]. Homology modeling is known to give quite accurate results for target-template pairs sharing more than 40% sequence identity [55] and all results from the alignment tools (Table 4.1) resulted in high sequence identities (>97%).

The algorithm and associated energy function used for the homology modeling task depend on the software chosen. In general, most software packages start to assign the coordinates of aligned amino acid positions from the backbone atoms retrieved from the template structure or by using this information to generate spatial restraints [80]. Modeling unaligned regions requires different tactics, which differ from software to software [80]. Four different homology modeling software were used, namely: (1) SWISS-MODEL, (2) Chimera-Modeller, (3) Rosetta and (4) I-TASSER. All predicted models that have resulted from this task are shown below, in Figure 4.4. The models were chosen to proceed with the modeling task according to the evaluation scores (crystallographic resolution of the chosen template, available information about the template's structure, RMSD from the template structure, ANOLEA, GROMOS, QMEAN, SOLVX and VERIFY3D). The evaluation results of the homology modeling task are shown in Table 4.3, at the end of this section.

SWISS-MODEL

The methodology used by this software is modeling by rigid-body assembly [82], which constructs the model from a few core regions, loops and side chains retrieved from accessing information on related structures [81]. The main limitation of this method is the dependency of model accuracy on a good alignment – gaps in the target-template pairwise alignment result in poorer models.

The server could only recognize 1EMA and 1QYO as template structures. As 1EMA shared less percentage of identity and did not have an associated structure for the last nine aminoacids at C-terminal, 1QYO was selected as template.

The structure of 1QYO (Figure 4.3) is one of the few GFP structures deposited in the PDB that has an associated secondary structure for the C-terminal residues (Figure 4.3, in red). The result of the homology modeling task (GFP_SM) is shown in Figure 4.4a superimposed to the 1QYO structure. A good indicator of homology modeling accuracy is given by the estimated RMSD between the template structure and the homology model. The higher is the RMSD value the worse is the homology model [55]. The rigid body assembly modeling method used by SWISS-MODEL [82] results in a small RMSD of 0.050Å, between the template backbone heavy atoms of the structure and the predicted homology model structure.

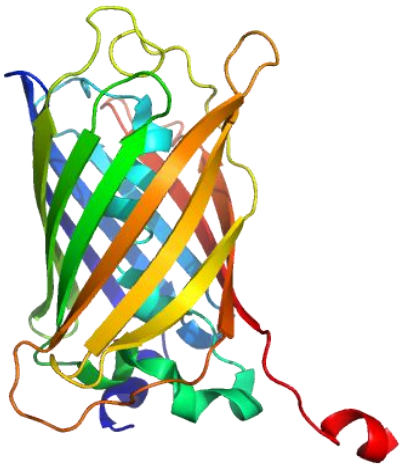


Figure 4.3 - Crystallographic structure of the S65G Y66G GFP variant (PDB ID: 1QYO). The backbone structure is colored blue beginning at the N-terminal to red at the C-terminal of the structure. Displayed using cartoon representation in PyMol.

This homology model scored the best ANOLEA profile and a good score by GROMOS accuracy indicator. The SOLVX scored indicates this model as a well packed structure, with a score of -133.7. The SWISS-MODEL model for GFP had the best QMEAN score regarding all-atom interaction, solvation and torsion terms. However this was the model that had the worst

VERIFY3D result, being the model with the least agreement between its aminoacid sequence and its 3D atomic model. The residues scored as “very bad” (scores below 0) are concentrated mainly at the C-terminus. This is in agreement with a poor score for the secondary structure agreement term in QMEAN. For further information, consult Table 4.3. As the evaluation criteria were not as good as the ones obtained with other modeling software, this model was discarded.

Chimera Modeller

Modeller software package implements an automated approach to homology modeling by satisfaction of spatial restraints [82].

Modeller generated five models for each template, ordered by RMSD (Table 4.2). It is clear that the best solution (GFP_s1M, Figure 4.4b) was obtained for the 1QYO template which had the most completed sequence information. This is a consequence of the homology modeling module implemented by Modeller [83]. Accordingly, the worst model in terms of RMSD was obtained when using 1YHI, the template with more unstructured terminals and missing residues.

Table 4.2 - Best solution obtained by Chimera’s Modeller tool for each template used for GFP. The solutions are ordered by the corresponding RMSD and some extra remarks are given.

Template	RMDS (Å)	Aminoacids missing
1QYO [1]	0.909	1
4GES	1.021	7
1GFL	1.288	8
1YHI [2]	3.664	10

[1] The best for the given template.
[2] Worst solution for the given template.

The resulting best model (Figure 4.4b) has a secondary structure associated with the last residue in C-terminal, according to the used template.

The obtained RMSD between the template and the resulting model was higher (0.909Å) than for the SWISS-MODEL model, which is a direct result of the different homology modeling algorithm employed by Modeller. The consideration of extra spatial restraints by the modeling software [82] results in a homology model that differs more from its starting template. Although less residues are in a favorable energy environment when scored by ANOLEA

(93.70%), this is even more evident when analyzing the GROMOS scores (only 49.22% spread along the sequence are in a favorable energy environment). However the VERIFY3D results are overall better than for the SWISS-MODEL homology model. The Modeller homology model was the model that was best scored in terms of the overall QMEAN score - for further information, consult Table 4.3. Therefore, this model GFP_s1M (Figure 4.4c b) was selected for the following modeling steps.

Rosetta

Rosetta's homology modeling approach is based on three steps: (1) built an incomplete model structure based on the alignment between the target sequence and a chosen template, (2) complete the missing structure using loop modeling and (3) rank or evaluate the energy of the resulting structural models [84].

1QYO was chosen as the homologous template and it was possible to obtain one homology model for GFP (GFP_R, Figure 4.4c). When compared to the 1QYO structure, a very low RMSD value of 0.003Å was obtained, the lowest for the evaluated homology models.

This model had the best SOLVX score (-138.7) and was well evaluated according ANOLEA and GROMOS. The VERIFY3D score showed a low percentage for residues with no agreement with their expected 3D atomic model, mainly concentrated in the C-terminus. The VERIFY3D profile is very similar to the one obtained for the homology model for GFP predicted by Modeller. This model GFP_R (Figure 4.4c) was also took into account in the following modeling steps.

I-TASSER

I-TASSER is a webserver that allows homology modeling. This software combines various techniques as threading, *ab initio* and atomic-level refinement to built an accurate homology model [85].

The best model obtained by the I-TASSER server was GFP_IT (Figure 4.4d). As shown, there is no associated secondary structure to the C-terminal residues, differing from the previously described homology models of GFP. This can be explained since 1QYO was not selected as template by the I-TASSER software.

I-TASSER does not allow the user to chose the template that is selected [85] and it selects multiple homologous templates simultaneously, all corresponding to PDB entries for fluorescent protein structures or GFP-fusion proteins. This set is identified by the LOMETS

threading programs and correspond to a representative PDB subset with a pair-wise sequence identity cutoff of 70% [85]. The protein structures chosen as templates by the software (consult Table 4.3) were not identified previously as good alignment results (see Table 4.1).

This homology model had the worst scores regarding ANOLEA and GROMOS, with unfavorable energy residues spread along the sequence. However this homology model had the best overall VERIFY3D results, although the worst values were still focused on the C-terminal residues. This is in agreement with the best secondary structure agreement term of QMEAN - for further information, consult Table 4.3. As it was the only model with an unstructured C-terminal this homology modeling solution was discarded.

Table 4.3 reports to the homology modeling evaluation results for the models described above.

Selected homology models for GFP

All predicted homology models are shown in Figure 4.4. As described above, only two of the obtained homology models for GFP were chosen to proceed with the modeling task: GFP_s1M (Figure 4.4b) and GFP_R (Figure 4.4c), shown with the same orientation for better comparison.

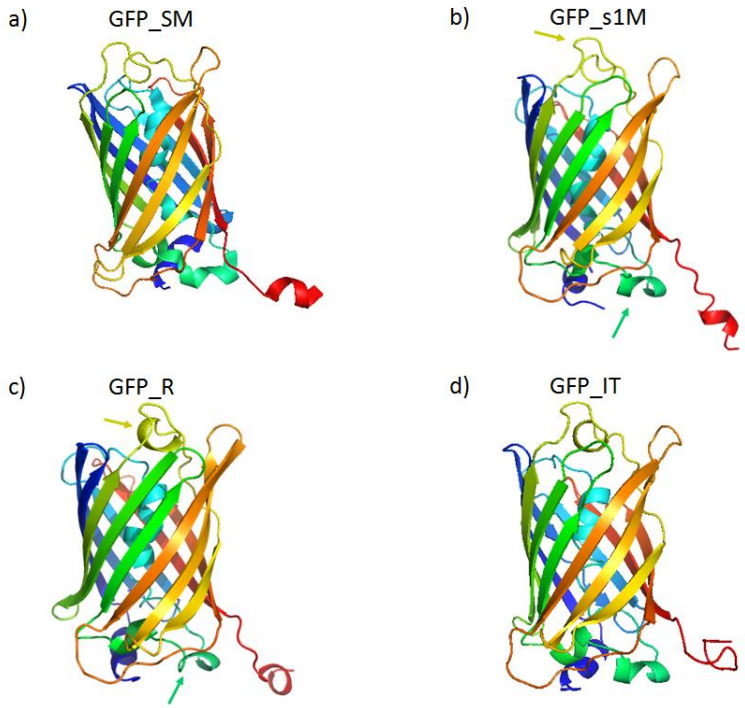


Figure 4.4 - Best Homology models obtained for GFP using (a) SWISS-MODEL, (b) Modeller, (c) Rosetta, and (d) I_TASSER. Only the models (a) and (c) were chosen to carry out the modeling task. The backbone structure is colored in gradient from blue (at N-terminal) to red (at C-terminal) of the structure. Two helix structures are highlighted with arrows. Displayed using cartoon representation in PyMol.

Table 4.3 - Summary of the comparative homology modeling evaluation for the homology models obtained for the GFP amino acid sequence. The best result for each score is highlighted in green

	Template	**Sequece identity**	**ANOLEA [1]**	**GROMOS [2]**	**RMSD (Å)**	**SOLVX**	**VERIFY3D [3]**	**QMEAN**						
								Overall Score	Cβ interaction	All-atom interaction	Solvation	Torsion	Secondary structure agreement	Solvent accessibility agreement
SWISS-MODEL	1QYO	98.31%	99.58%	81.00%	0.05	-133.7	5.04% 11.34%	-0.54	-1.47	-1.51	0.3	-0.39	0.96	1.29
Modeller	1QYO	98.31%	93.70%	49.22%	0.909	-127.9	6.30% 4.62%	0.36	-1.58	-1.68	-0.12	-1.46	0.71	1.06
Rosetta	1QYO	98.31%	100%	73.95%	0.003	-138.7	4.62% 10.50%	0.7	-1.45	-1.56	0.34	-0.8	0.78	1.21
I-TASSER	1B9C, 3NED, 1KYP, 2JAD, 3U8P, 2VZX, 3CGL	**[4]**	50.43%	22.48%	**[5]**	-124.2	2.10% 5.04%	-0.96	-2.19	-2.48	-0.31	-4.42	0.06	0.98

[1] Percentage of residues that were scored positively (>0) by the ANOLEA score.
[2] Percentage of residues that were scored positively (>0) by the GROMOS score.
[3] Percentage of residues that were scored as bad (0-1.1; first line) and very bad (<0; second line) by the VERIFY3D score.
[4] It was not possible to calculate the sequence identity to the group of templates chosen by I-TASSER.
[5] RMSD value was not calculated due to template selection by the I-TASSER webserver.

As shown in Figure 4.4, the two selected homology models have a very similar structure and highly conserved folding and packing. The main differences are (1) the orientation of the C-terminal residues (in red); (2) the length of the helix structure in the N-terminal; (3) the subtle difference in orientation and length of some of the loops between the beta sheets; (4) the length of the helix highlighted by the green arrow; and (5) the presence of a helix structure in the yellow-colored loop in GFP_R.

The interest of proceeding with both models in parallel concerns a systematic search for eventual structural inconsistency and to avoid any computational bias.

Despite extensive optimization of known homology modeling software, there is still a considerable uncertainty and source of error in the automated estimation of protein three-dimensional structure and its quality evaluation. Though, whenever it is possible, the final evaluation should be experimentally validated through mutagenesis experiments, affinity labelling, NMR dipolar coupling, cryo-electron microscopy [85].

4.2.2. Structural relaxation of homology models

Selected protein models have to be relaxed through MD methodologies, in order to take into account the dynamical aspects of recognition and binding mechanisms observed in Nature and to avoid any structural clash or unphysical structural constraint due to computational artifacts.

In classical MD simulations it is possible to derive the time-dependent random movement of a protein system along its characteristic potential energy landscape for a given temperature and pressure by integrating the Newton's equations of motion step-by-step [86]. This asset of differential equations [87] describe the net force applied on each particle by the atomic environment and the particles' positions and correspondent velocities [88].

MD idealizes the system as spherically symmetric separate point particles [88] and the kinetic energy is randomly distributed by all the system's particles, in order to obtain an initial velocity. However particles are interconnected through covalent bonds with fixed lengths and stiffness [86,87] and interacting with non-bonded neighbors through van der Walls and Coulomb-type electrostatic interaction terms that sum up into the energy function [86]. The energy function is based on a force field, which described the mechanical behavior of each pair of interacting atoms in the system. The potential energy is described in terms of cumulative

energy terms [88]. The commonly used force fields for proteins are CHARMM [89], AMBER [90], and GROMOS [91].

The GROMOS96 53a6 force field, the one used in this work, has been parameterized and optimized to work with biological systems, specifically proteins in solution [92]. A simulation of 10ns length was analyzed and the results are summarized in Figure 4.5 for GFP_s1M and in Figure 4.6 for GFP_R.

By Figure 4.5a, one can see that the helix structure which was associated with the C-terminal residues of GFP_s1M was replaced by an unfolded structure, remaining exposed to the solvent. The same phenomenon occurred to the helix highlighted by the blue arrow in Figure 4.4a. The GFP protein has a very stable β-barrel core structure [75] and stable systems in MD usually display RMSD values between 1.0-2.0Å [78]. Figure 4.5c shows that the major flexibility is conferred by the C-terminal residues (232-238) which have the highest RMSF deviation (RMSF values between 0.5-2.0Å are quite common in MD studies for proteins around equilibrium [78]). This contribution is reinforced through the representation of the high B-factors corresponding to the C-terminal residues (Figure 4.5d, in red) – high B-factors correspond to high flexibility [93]. This is due to the unfolding of the modeled C-terminal helix structure.

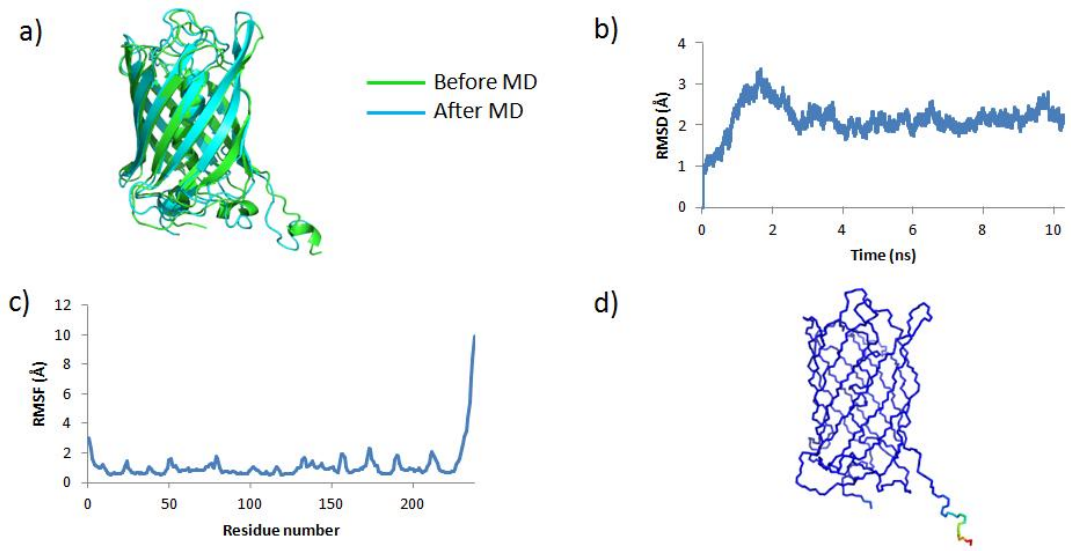


Figure 4.5 - Results obtained after the first MD simulation for the GFP_s1M. a) Superimposition of the model obtained by Modeller before MD simulation (in green) and after MD simulation (in blue). b) Variation of the RMSD values along the MD simulation trajectory. c) RMSF values per residue. d) B-factor representation of the structure flexibility colored in gradient. Blue represents residues with less flexibility and red represents residues with more flexibility average over the entire trajectory.

GFP_R had a very similar MD profile to the model from Modeller (Figure 4.6). The helix structure of the N-terminal residues did not persist and the C-terminal residues remain with no associated secondary structure. Contrary in GFP_s1M, the C-terminal residues stopped being so solvent-exposed. Both helix structures, highlighted by the yellow and blue arrows in Figure 4.4b, did not show any resilient secondary structure. In Figure 4.6b one can see that the RMSD profile is similar to the GFP_s1M model on the first 2ns of simulations, correspondent to the equilibration phase. The conformation of C-terminal tends to interact with the protein surface and the protein continued to search for a more stable conformation, reaching a higher RMSD plateau. As described before, the major contribution is due to the flexibility of the C-terminal residues, resulting in the high corresponding RMSF values and B-factors (Figure 4.6c and Figure 4.6d, respectively).

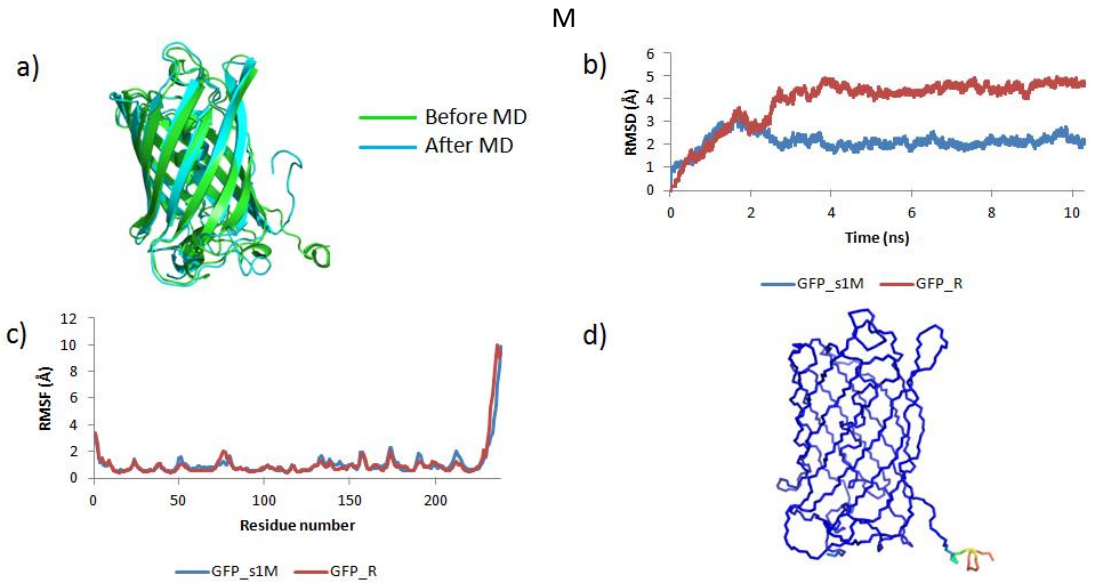


Figure 4.6 - Results obtained after the first MD simulation for the GFP homology model obtained by Rosetta. a) Superimposition of the model obtained by Rosetta before MD simulation (in green) and after MD simulation (in blue). b) Variation of the RMSD values along the MD simulation trajectory. c) RMSF values per residue. d) B-factor representation of the structure flexibility colored in gradient. Blue represents residues with less flexibility and red represents residues with more flexibility average over the entire trajectory.

Obtained results confirmed the need for a pre-equilibration MD step before docking.

4.2.3. Automated molecular docking

Protein docking predicts the ideal structural complementarity between two molecular surfaces [71]. The main goal of protein docking is to predict *how* a pair of molecules interact [71], predicting accurate ligand poses based on the maximization of the binding energy [16]. It explores the conformational available space and calculates the free energy of binding of each conformation to identify the minimum energy conformation [94].

To successfully predict a suitable target/ligand complex there are three steps: (1) have accurate structures of the molecules involved in the interaction, (2) conformational sampling to find the location of the binding site, and (3) determination of the binding mode and energy evaluation [70]. The success of a docking software depends on two components: the search algorithm, and the scoring function [95]. The combination of these two components will dictate the overall results of the docking task.

The relaxed structures of GFP (GFP_s1M and GFP_R) were used to study the potential binding modes of the lead ligand LA-A4C7 and the negative control NA-A7C1. This study was performed using automated molecular docking tools. Protein docking requires the structures of the elements that form the complex and aims to predict correctly the binding site on the target and the orientation of the ligand. At the end, a rank of possible docking poses based on estimated binding affinities is given [96].

Regarding GFP, the coordinates of the lead ligand LA-A4C7 (Figure 4.7a) and the negative control ligand NA-A7C1 (Figure 4.7b) were set up based on previous experimental solid phase affinity results of the complex [47]. The value of the constant affinity (Ka) that was experimentally determined for the affinity system GFP/LA-A4C7 was $2.38\times10^{5}M^{-1}$.

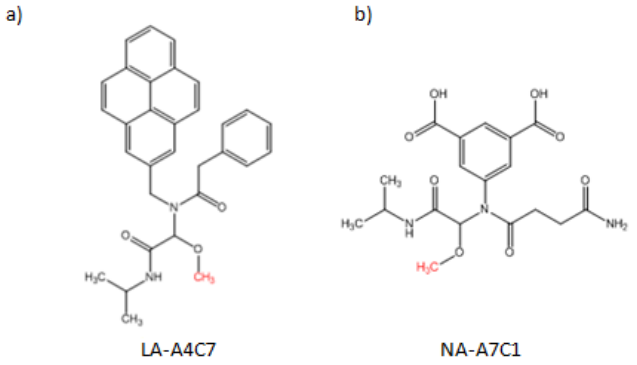


Figure 4.7 - Molecular structures of the solid phase screened ligands for GFP system [47]. a) Lead ligand with the best affinity, and b) negative control ligand with the worst affinity for GFP system. The chemical group used to immobilize the ligand into the solid support is highlighted in red.

Autodock [69] was the docking program package used for the docking task in this work as it allows a robust and accurate docking procedure at a reasonable computational demand [69].

The Lamarckian genetic algorithm is used by default in the Autodock software package. Based on the analogy with the concept of natural genetics and biological evolution [69], its goal is to “evolve” previous conformations into new low energy conformations. Each docking pair is represented by a conformation that is described as a “gene” that has a particular energy and the entire “genome” is a representation of the energy landscape which will be explored.

Similar to biological evolution, random pairs of individuals are “mated” using a process of crossover and there is also the possibility of a random mutation in the “offspring” [69]. During each iteration, high-scoring features in the current “generation” are preserved in the next cycle [96]. This approach allows exploring large conformational spaces in a more efficient manner.

The AutoDock scoring function [95] is a semi-empirical free energy force field scoring function that evaluates conformations and calculates the ligand-receptor binding affinity [97] [98]. The calculation of the estimated free energy of binding is based on its components: 1. final intramolecular energy, which is divided into (1.1.) van der Waals, hydrogen bond and desolvatation energy, and (1.2) electrostatic energy; 2. final total internal energy; 3. torsional free energy; and 4. unbound system’s energy [95,98]. The estimated free energy of binding is estimated as: 1+2+3-4 according to the energy function implemented in Autodock.

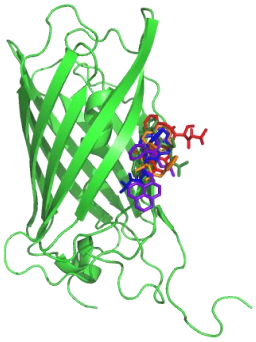
Some settings were altered as described in the Methods Section. Changed parameters were (1) the number of runs (256), (2) the number of points of the grid (130 100 100) and (3) the grid spacing (0.5). The number of runs corresponds to the number of docking solutions. Higher population values result in a better sampling of the conformational spacing, increasing the probability of finding a better docking pose, on the expenses of allocated computational time. The size of the grid defines the three dimensional area to be explored by the ligand and its value can be tuned by altering the number of points and/or the spacing between the grid nodes. Both these variables correlate with the accuracy of the search and on the final prediction of the best binding solutions. Although a finer grid (a larger number of points and /or less spacing between grid nodes) would be advantageous for the quality of the docking results, it would increase the memory requirements and processing time [70]. At this stage, it was decided to do a blind docking – the search of conformational space is not biased into a particular location of the protein structure where there are evidences of a specific binding. For that purpose, the docking and grid parameters file were optimized to include all the protein surface in the searching area.

The criteria followed for choosing the best results have been described in the Methods section.

Docking results of the lead candidate (LA-A4C7)

The results for the affinity system GFP/LA-A4C7 are summarized in Table 4.4.

Table 4.4 - Summary of the docking results for the affinity system GFP/LA-A4C7.

Target protein	Cluster Rank	Number of solutions per cluster	Model			
			Solution	Binding Energy (kcal/mol) [1]	Ka ($x10^5 M^{-1}$)	Ligand binding site
GFP_s1M	1	10	12	-7.03 -9.78/-0.02	1.43	
	2	6	159	-6.98 -9.75/-0.02	1.31	
	3	5	69	-6.97 -10.08/0.08	1.29	
	4	16	239	-6.81 -8.89/-0.07	0.98	
	5	9	231	-6.79 -9.23/-0.15	0.95	
GFP_R	1	5	71	-6.85 -9.25/-0.07	1.06	
	2	7	14	-6.62 -9.1/-0.04	0.7	
	3	1	74	-6.33 -7.77/-0.13	0.11	
	4	2	206	-6.28 -6.85/-0.06	0.02	
	5	3	235	-6.23 -6.84/-0.01	0.02	

[1] Estimated binding energy (on top) and two of its components: van der Waals, hydrogen bonding and desolvation energy/electrostatic interaction (on bottom).

Looking at the first five docking results for the affinity system GFP/LA-A4C7, we can state that this is probably a good ligand for this system since it presents good (very negative) estimated energies of binding, and the top ranking clusters are highly populated. The van der Waals, hydrogen bonds and desolvation energy are the major contributors to the low estimated free energies of binding. However, this information is not enough to decide which docking solution to chose.

Regarding GFP_s1M, the first ranked model (Solution 12) binds to the surface of the β-barrel structure, near β7 and β8. The solid support binding group is solvent-exposed and the

calculated Ka ($1.43x10^5M^{-1}$) is in the same order of magnitude as the experimentally determined Ka ($2.38x10^5M^{-1}$). It was the solution with the lowest estimated free binding energy (-7.03kcal/mol). Another remark is that cluster 1 is highly populated (10 solutions grouped in the same cluster, within a threshold of 2Å), representing one of the preferential docking site at the target protein surface.

As the location, free binding energies and Ka for the other top-rated solutions are very similar, no other solutions were further analyzed in dynamics for this affinity system.

The affinity system GFP_R/LA-A4C7 presents good (very negative) free binding energies. Contrary to the results for the affinity system GFP_M/LA-A4C7, the clusters are not as populated.

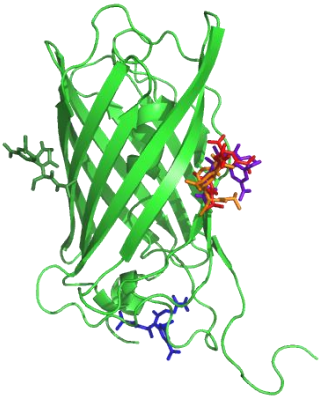
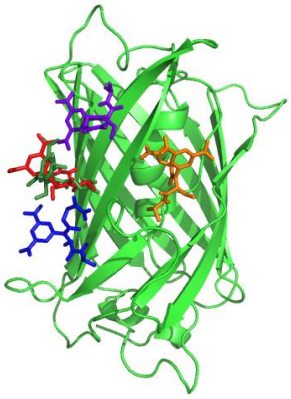
Solution 71, the top ranked solution for this system binds to the β-barrel structure close to the top of the structure, near β10. As seen by the orientation of the C-terminal residues (to the left instead of the right) the docking site is very different from the docking sites found for the GFP_M/LA-A4C7 affinity system. The calculated Ka ($1.06x10^5M^{-1}$) has the same order of magnitude as the experimentally determined Ka ($2.38x10^5M^{-1}$). This model was chosen to proceed with the modeling task.

Solution 14, the second best ranked docking solution is located under the protein structure, near the loop between β9 and β10, binding to the bottom loops. However the calculated Ka ($0.72x10^5M$) is already an order of magnitude lower than the experimentally determined Ka and was chosen to proceed with the modeling task.

Docking results of the negative control candidate (NA-A7C1)

The results for the affinity system GFP/NA-A7C1 are summarized in Table 4.5.

Table 4.5 - Summary of the docking results for the affinity system GFP/NA-A7C1.

Target protein	Cluster Rank	Number of solutions per cluster	Model			
			Solution	Binding Energy (kcal/mol) [1]	Ka ($x10^5 M^{-1}$)	Ligand binding site
GFP_s1M	1	2	49	-6.76 -8.19/-1.5	0.9	
	2	2	248	-6.47 -7.99/0.87	0.55	
	3	6	55	-6.39 -6.96/-2.18	0.48	
	4	7	215	-6.01 -7.41/-1.14	0.25	
	5	1	111	-5.81 -7.41/-1	0.18	
GFP_R	1	4	142	-5.82 -5.62/-2.27	0.18	
	2	18	110	-5.76 -5.75/-2.11	0.16	
	3	3	236	-5.73 -6.44/-2.04	0.16	
	4	1	133	-5.6 -7.67/-0.52	0.13	
	5	6	83	-5.46 -5/-2.34	0.10	

[1] Estimated binding energy (on top) and two of its components: van der Waals, hydrogen bonding and desolvation energy/electrostatic interaction (on bottom).

Overall, we can see less populated clusters and more positive binding energies than for the affinity system GFP/LA-A4C7 (Table 4.4), which would be expected for a negative control.

The first ranked solution for the receptor GFP_s1M (Solution 49) binds on the bottom of the barrel, to the loop between the central helix structure and β4, and the chemical group necessary to bind to the solid support is unavailable.

The second top-ranked solution (Solution 248) has a different binding site, located near β7 and β8. The calculated Ka is an order of magnitude lower ($0.55 \times 10^5 M^{-1}$) than the experimental

value for the GFP/LA-A4C7 system and the resin-binding group is solvent accessible. Therefore, this model was chosen to proceed with the modeling studies as a negative control.

Regarding the analysis for the affinity pair GFP_R/NR-A7C1, there are some solutions that could not bind to the resin due to the inaccessibility of the resin-binding group to solvent.

Solution 110 belongs to a very populated cluster and its docking site is located between β8 and β9. The solid phase-binding group is solvent-exposed and it has a lower $Ka=0.16\times10^{5}M^{-1}$, one order of magnitude smaller than the calculated Ka values for the lead ligand. This solution was chosen as a negative control for this affinity system.

Solution 236 belongs to the third cluster and its docking parameters are very similar to the previous solution, so it was not chosen to proceed with the modeling task.

Some notes have to be in mind when performing a docking protocol. Docking calculations can be hampered by a number of reasons: (1) the ligand binds to deep specific pockets of the protein structure [16]; (2) the docking protocol does not consider the presence of solvent, but a dielectric continuum instead, which can be crucial for hydrogen bonding [71]; (3) attachment of the ligand to a solid surface via a spacer arm [16]; (4) ligands with high flexibility; (5) weak interactions between the ligand and the protein; (6) large-scale motions of the peptide backbone [99]. However, new optimizations and extensions are being developed into existing programs to overcome these drawbacks.

4.2.4. Extensive molecular dynamic simulations

The molecular systems selected to proceed for MD simulations are summarized in Table 4.6. Before using the GROMOS protocol, the ligands have to be parameterized according to the GROMOS96 53a6 force field, in order to assure that the used atom types are compatible with the GROMOS force field used in this work. The online server PRODRG was used for this purpose.

The next step was to perform a MD simulation with the complexes summarized in Table 4.6. Afterwards possible interactions between the molecular complexes were assessed, according to the thresholds described in the Methods section. The results regarding possible hydrogen bonds and hydrophobic interactions between moieties or atoms of the studied ligands and GFP are summarized in Table 4.7.

Table 4.6 - Chosen docking poses to proceed for extensive MD for the different affinity pairs analyzed with the GFP system.

Affinity Ligand	Homology Model	Chosen Solution
LA-A4C7	GFP_s1M	12
	GFP_R	71
NA-A7C1	GFP_s1M	248
	GFP_R	110

Table 4.7 - Extensive MD simulation results regarding the interactions between the ligands and GFP.

Ligand	Receptor	Solution	Main interaction	Type interaction [1]		Contribution (%)
LA-A4C7	GFP_s1M	12	TYR145:$C^{\varepsilon 1}$---MOL:CAD	Hyd		91,46
			HIS169:$C^{\varepsilon 1}$---MOL:CAD	Hyd		90,17
			HIS148:C^{γ}---MOL:CAD	Hyd		63,02
			ARG168:N^{ε}---MOL:OBI		Hb	28,86
			LYS166:N---MOL:O		Hb	17,43
			TYR151:C^{ζ}---MOL:CAA	Hyd		10,1
	GFP_R	71	PHE223:C^{γ}---MOL:CAA	Hyd		67,29
			TYR143:$C^{\varepsilon 1}$---MOL:CAD	Hyd		29,08
			PHE223:$C^{\delta 2}$---MOL:CAH	Hyd		23,66
			GLN204:$N^{\varepsilon 1}$---MOL:OBI		Hb	5,77
			PHE223:$C^{\delta 1}$---MOL:CAD	Hyd		4,92
			GLN204:$N^{\varepsilon 2}$---MOL:O		Hb	2,94
NA-A7C1	GFP_s1M	248	HIS148:C^{δ}---MOL:CAC	Hyd		26,64
			TYR200:O^{η}---MOL:O		Hb	20,23
			TYR200:C^{ζ}---MOL:CAC	Hyd		17,17
			ASN149:$N^{\gamma 2}$---MOL:OAZ		Hb	8,73
			TYR200:O^{η}---MOL:OBA		Hb	5,79
			ASN149:N---MOL:O		Hb	5,52
	GFP_R	110	TYR182:O^{η}---MOL:OAZ		Hb	3,84
			SER99:O^{γ}---MOL:OAT		Hb	2,65
			ASN164:$O^{\gamma 1}$---MOL:NAW		Hb	0,96

Note: GFP atoms are identified according to the recommended atom identifiers following the 1969 IUPAC-IUB guidelines (consult [100]).
[1] The interactions are classified as hydrophobic (Hyd) or hydrogen bond (Hb).

Table 4.7 shows many relevant interactions (above 10% of the simulation time) between GFP and LA-A4C7. Due to its hydrophobic character the hydrophobic interactions are predominant. Some of the interactions between LA-A4C7 Solution 12 and GFP_s1M are shown in Figure 4.8. The most relevant hydrophobic interactions were found between a heavy atom of the pyrene ring (CAD) and the benzene ring of the TYR145 side chain (Figure 4.8a). By its stable distance profile, we can observe that the ligand has found a stable docking pose. Other hydrophobic interactions (e.g. between the ligand and HIS169, HIS148 or TYR151) contribute for the binding of the ligand to this docking pose. On the other hand hydrogen bonding also contributes for the ligand binding (e.g. between ARG168 side chain and the ligand (OBI) at the end of the simulation - Figure 4.8b).

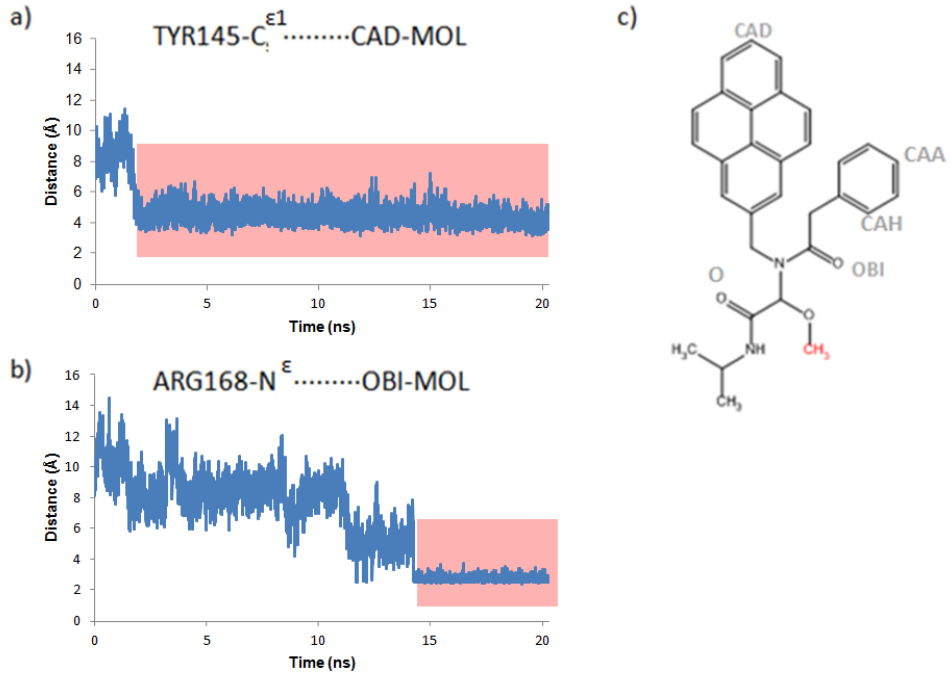


Figure 4.8 - Monitoring the distance between GFP/LA-A4C7 Solution 12 during the MD simulation trajectory (a) between TYR145:$C^{\varepsilon 1}$ and LA-A4C7:CAD, and (b) between ARG168:N^{ε} and LA-A4C7:OBI. (c) Ligand LA-A4C7 and atom types involved in the main interactions in Table 4.7.

Although some relevant interactions were found between Solution 71 and GFP_R, a larger number of relevant interactions were found between Solution 12 and GFP_s1M. This is in agreement with the docking results described before for both models: the estimated binding energy of Solution 71 (-6.87kcal/mol) is more positive than the estimated binding energy of Solution 12 (-7.03kcal/mol). With this information we can conclude that the docking pose of Solution 12 (between β7 and β8) is a more probable binding site location than the docking pose of Solution 71. On the other hand, and according to the results found for Solution 71 in

terms of estimated binding energy, Ka and relevant interactions, this solution should not be discarded as an alternative binding site.

Regarding ligand NA-A7C1 Solution 110, no relevant interactions were found with GFP. The interactions shown in Table 4.7 occur at the beginning of the simulation, after which the ligand distances itself from the protein surface.

On the other hand for the NA-A7C1/GFP_s1M Solution 248 system some relevant interactions were found, mostly hydrophobic. The distance profiles for these interactions are shown in Figure 4.9. However, contrary of what was observed for LA-A4C7, the interactions happen only at the beginning of the simulation (Figure 4.9a), occur in short periods (Figure 4.9b) or occur only once in the middle of the simulation (Figure 4.9c).

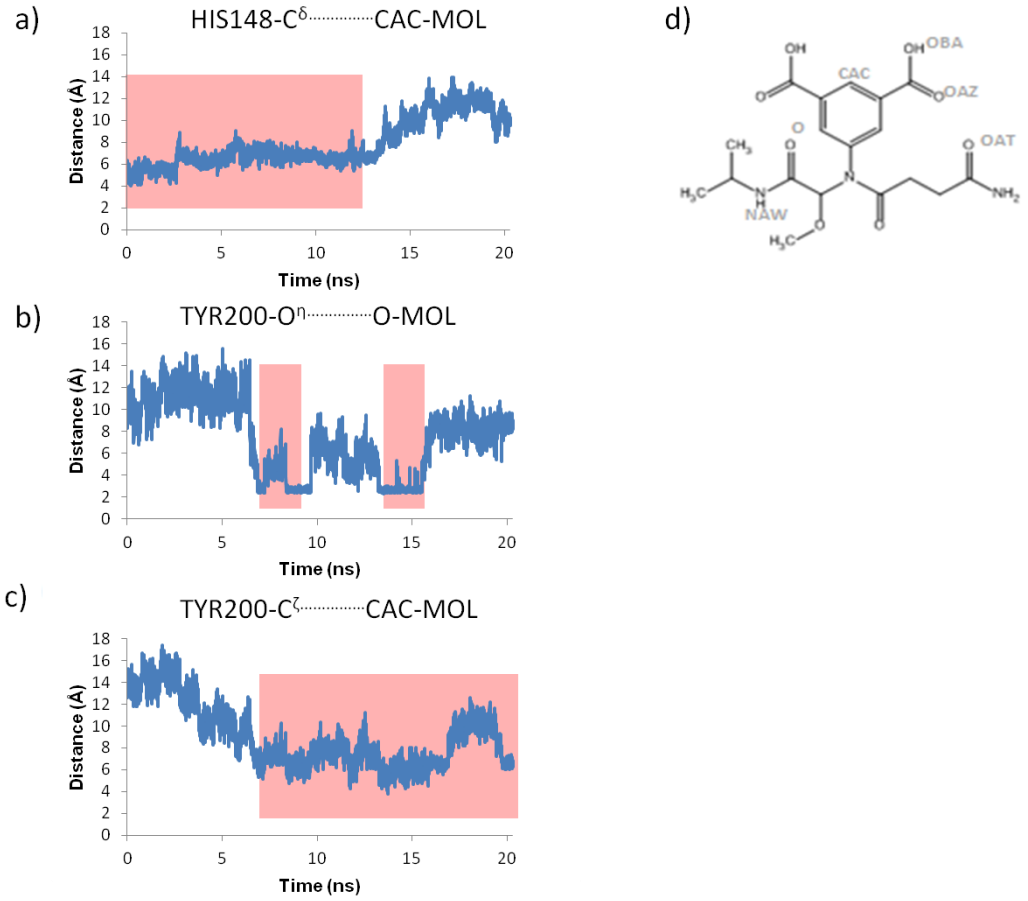


Figure 4.9 – Monitoring the distance between GFP/NA-A7C1 Model 248 during the MD simulation, (a) between HIS148:C^{δ} and NA-A7C1:CAC, (b) between TYR200:O^{η} and NA-A7C1:O, and (c) between TYR200:C^{ζ} and NA-A7C1:CAC. (d) Ligand NA-A7C1 and atom types involved in the main interactions in Table 4.7.

According to the modeling results and comparing them with the results found for the LA-A4C7/GFP affinity system, NA-A7C1 acts according to a poor ligand for this system and has a decreased probability of displaying affinity towards GFP.

In reality NA-A7C1 was used as the lead ligand for RK-GFP. This will be considered during the analysis of the results for this affinity pair system in Section 4.4.

4.3. *In silico* 2nd generation library of ligands for binding of GFP

The goal of this section was to create a second generation of ligands with affinity towards GFP which presented better theoretical estimated binding energy and affinity constant than ligand LA-A4C7. For this purpose two strategies were followed, which will be described below.

4.3.1. Docking-based second generation library

The first strategy was based on a characterization of the docking pose of solution 71 of the GFP_R/LA-A4C7 pair. This docking solution was characterized by an estimated binding energy of -6.87kcal/mol and an affinity constant of $1.06\times10^{5}M^{-1}$. Its location is shown in Figure 4.10a.

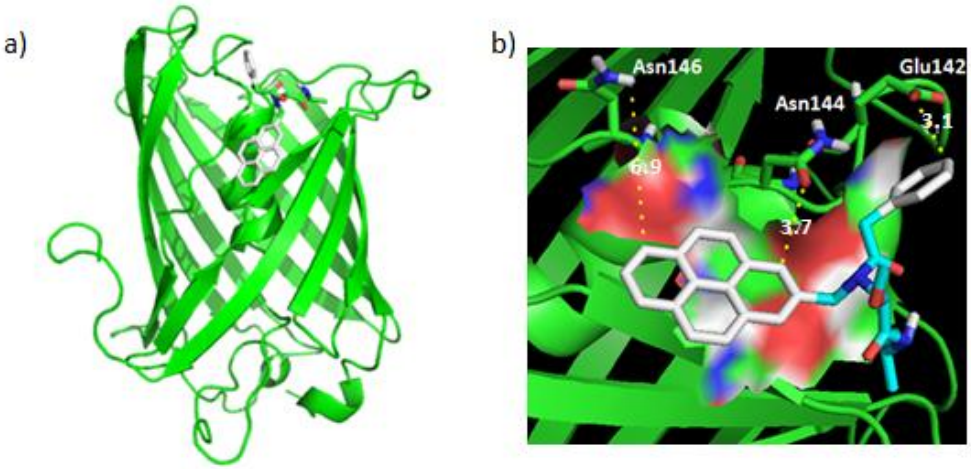


Figure 4.10 - Solution 71: lowest estimated free binding energy docking hypothesis for the affinity system GFP_R/LA-A4C7. a) Location of the ligand (colored by atom type) on the surface of the GFP_R structure (in green). b) Close-up to the docking location in a). An hydrophobic patch is shown by surface display and key amino acids are shown in sticks.

As shown in Figure 4.10b the pyrene ring of LA-A4C7 establishes hydrophobic interactions with a hydrophobic patch located at the surface of GFP formed by Leu141, Tyr143, Tyr145 and His148. The benzene ring of LA-A4C7 establishes hydrophobic interactions with these residues. However, there is a lack of chemical groups that could interact via hydrogen bonds with surrounding residues. There are some residues in a close distance (namely Asn146, Asn144 and Glu142, shown in Figure 4.10b) that could interact via hydrogen bonds with nearby hydroxyl, carboxyl or other hydrogen-bonding capable groups from the ligand.

Based on the need of hydrophobic and polar groups, a library of amine and carboxylic compounds was constructed (Table 4.8).

4.3.2. Nanobodies-based second generation library

The second strategy was based on an extensive PDB search for existing structures of molecules binding to GFP. Until date, the only available structures were published by Kirchhofer and colleagues [101]. Their work was regarding two small proteins (named nanobodies) that could bind to GFP and affect its fluorescence, increasing it (Enhancer) or decreasing it (Minimizer). Three binding sites where chosen (shown in Figure 4.11), based on their influence in both binding partners (Minimizer and Enhancer) and the proximity between residues (e.g. consecutive triads).

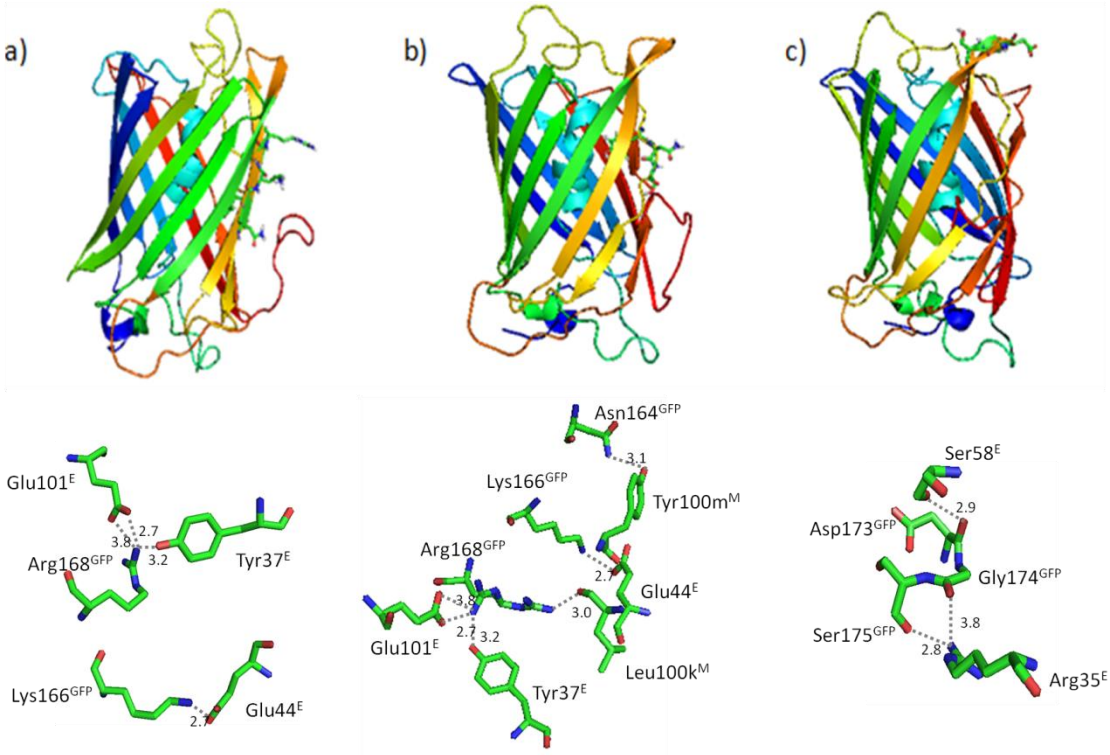
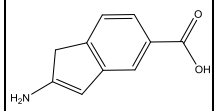
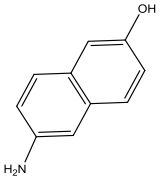
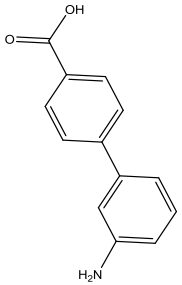
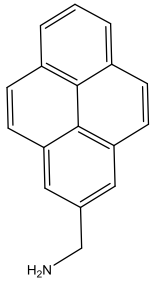
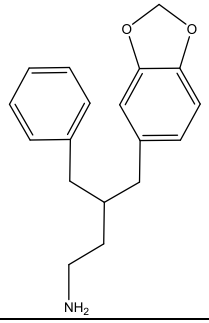
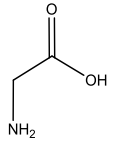
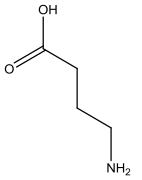
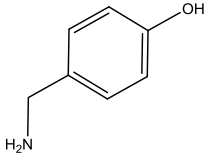
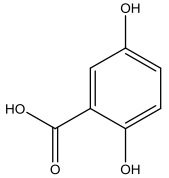
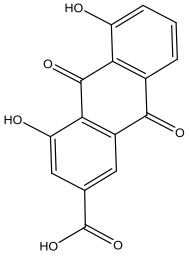
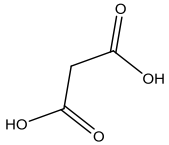
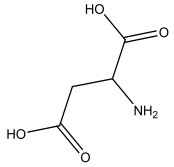
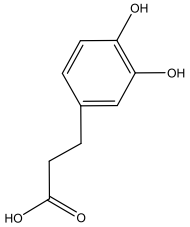
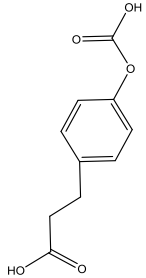
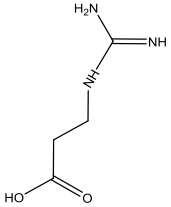


Figure 4.11 – Three amino acid triads (a, b and c) that are part of binding sites described in [101] for the Enhancer (E) and the Minimizer (M) to GFP. Below each binding site, the key residues are represented in stick display, as well as the distances between the chemical groups involved in hydrogen bonding (in Å).

The key residues involved are glutamic acids, tyrosines and arginines for both examples (Minimizer and Enhancer). The second generation Ugi library (Table 4.8) was based on the side chains of these amino acids, resulting in a selection of more hydrophilic compounds than the previous library.

Table 4.8 - Library of 2nd generation of Ugi-based ligands resulting from the docking-based (A1, A2, A3, A4, A5, C1 and C2) and the nanobodies-based (A6, A7, A8, C3, C4, C5, C6, C7) strategies.

Amines	**A1** 2-amino-1H-benzimidazole-5-carboxylic acid 	**A2** 6-amino-naphtol 	**A3** 4-biphenyl-3'-amino-acetic acid 	**A4** 1-pyrene methylamine hydrochloride 	**A5** 3-(1,3-benzodioxol-5-yl)-4-phenyl-1-butanamine 	**A6** Aminoacetic acid 	**A7** 4-aminobutyric acid 	**A8** 4-hydroxybenzylamine 
Carboxylic Acids	**C1** 2,3-dioxoindoline-7-carboxylic acid 	**C2** 4,5-dihydroxy-9,10-dioxo-9,10-dihydro-2-anthracenecarboxylic acid 	**C3** Malonic acid 	**C4** β-glutamic acid 	**C5** 3,4-dihydroxyhydrocinnamic acid 	**C6** 4-Acetoxy hydrocinnamic acid 	**C7** 3-Guanidino propionic acid 	

The next step was to perform a blind docking to GFP and chose the ligands that presented better estimated binding energy and affinity constant when compared to LA-A4C7.

To combine the characteristics of each library, both libraries were combined. Therefore, a total of 56 Ugi-based ligands (8amines×7carboxylic acids) were tested by a blind docking search onto the GFP receptor.

4.3.3. Affinity pair docking

The results of the blind docking with the 56 ligands from the 2nd generation library are summarized below.

45 ligands out of 56 (80.35%) had better estimated binding energy and Ka than LA-A4C7 (solution 71, -6.87kcal/mol and 0.106µM^{-1}, respectively). The first (A4C2) and the last (A7C7) classified of the 56 ligands are shown in Table 4.9. The best scored ligand displays a very negative binding energy and an affinity constant four orders of magnitude higher than LA-A4C7, being a very promising ligand.

Table 4.9 - First (A4C8) and last (A8C14) classified ligands of the 2nd generation library.

Ligand	Cluster 1			
	Docking solution	**Binding energy (kcal/mol)**	**Population/cluster**	**Ka (µM^{-1})**
A4C2	6	-13.47	2	7463.24
A7C7	107	-7.15	1	0.18

Figure 4.12 summarizes the frequency of each amine and carboxylic compounds that are part of the 56 best-rated ligands. The amines that were more common were A2, A3, A4 and A5, all belonging to the first library. The carboxylic acids also corroborated this assumption, as the most common compounds belonged to the first hydrophobic library (C2) and the second was chosen to mimic the tyrosine side-chain in the second library (C5). Both corroborated the need of a hydrophobic moiety to contribute for the binding to GFP, as it was observed by the hydrophobic ligand LA-A4C7.

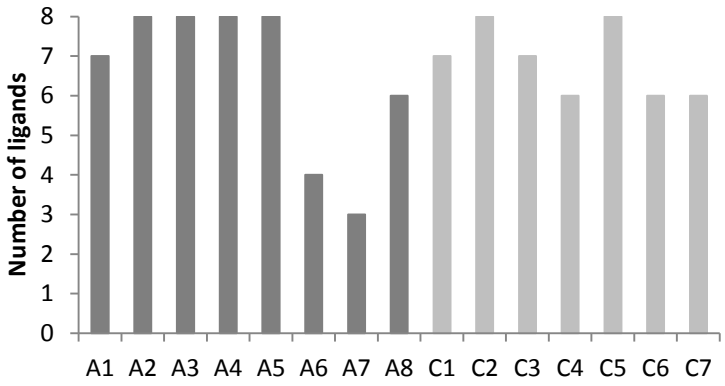


Figure 4.12 – Number of ligands which contain each amine or carboxylic acid summarized in Table 4.8

Figure 4.13 can give an idea of the range of binding energy values of the 56 analyzed ligands. Ligand A4C2, the best-rated ligand (see Table 4.9), is represented in green (very negative binding energy value) and ligand A7C7 is represented in red. These results are a mirror of the results shown in Figure 4.12, where more abundant compounds belong to ligands with better binding energy (e.g. A4 and C2).

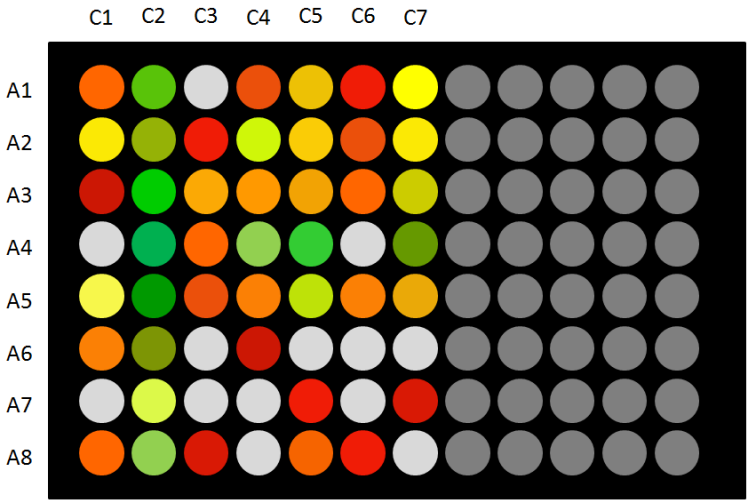


Figure 4.13 - Schematic representation of a 96-well microplate colored by the value of each ligand's estimated energy of binding to GFP (green>yellow>orange>red). In light grey: ligands with binding energy values more positive than LA-A4C7 M71.

All 56 initial ligand can be synthesized on a solid support and the experimental results compared to the blind docking results. One should notice that the results can vary due to the molecular docking limitations: (1) the docking is done between one ligand and one receptor, and (2) the ligand is solvent-free and not bound to a solid support.

4.4. *In silico* study of affinity ligands for the purification of RK-GFP

4.4.1. Homology modeling

The nucleotide and amino acid sequences corresponding to the RK-GFP used in this work are shown in Figure 3.1b, in the Methods section. The difference between GFP and RK-GFP protein sequence is a set of eighteen amino acids fused to the N-terminal: "RKRKRK-PPPDDDDKG-TGS".

One of the main goals of the RK-GFP modeling task is to understand the structural implications of the presence of an extra N-terminal peptide, and if it can have a deleterious effect on the GFP structure due to intramolecular interactions. It has been reported that insertions into the GFP sequence can alter GFP fluorescence [102] or stability [103]. However many crystallographic structures have been deposited in the PDB regarding GFP-fusion proteins and the characteristic β-sheet barrel of GFP is conserved (e.g. GFP-fusion myosin, PDB ID code 4ANJ; calmudolin, 3EVR; cytochrome b562, 3U8P).

The first step according to the homology modeling protocol is to find suitable templates. Homologous structures with the query sequence "RKRKRK-PPPDDDDKG-TGS" were not found in a preliminary NCBI BLASTp search.

Using the RK-GFP protein sequence as query, the same templates as for GFP (see Table 4.1) are retrieved and the same set of homology modeling software were used. The resulting models were evaluated based on the same scores as before and the homology modeling evaluation results are shown in Table 4.10.

Selected homology models for RK-GFP

Using the protein sequence of RK-GFP as a query in SWISS-MODEL did not return a model that included the N-terminal tag residues. The rigid body assembly method used by SWISS-MODEL [49] is not adequate to model sequences that do not have an associated template or when the alignment contains large gaps.

According to the evaluation results present in Table 4.10, two of the obtained homology models for RK-GFP were chosen to proceed with the modeling task: homology model obtained by Modeller (RK-GFP_s1M) and by Rosetta (RK-GFP_R) (Figure 4.14).

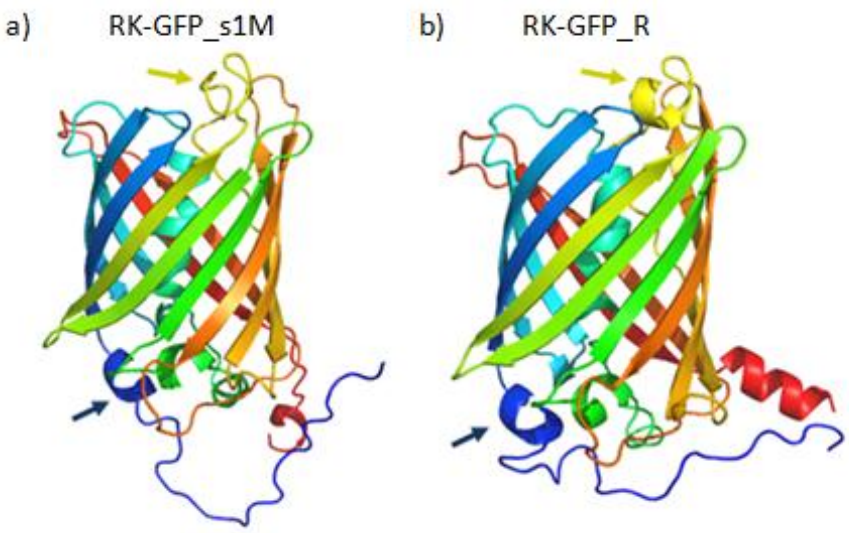


Figure 4.14 - Best homology models obtained for RK-GFP using (a) Modeller, and (b) Rosetta. The backbone structure is colored in gradient from blue (at N-terminal) to red (at C-terminal) of the structure. Two helix structures are highlighted with yellow and blue arrows. Displayed using cartoon representation in PyMol.

As shown in Figure 4.14, the two models have similar structure and packing and none have an associated secondary structure for the tag residues (colored in dark blue). The main structural differences are: (1) the orientation of the C-terminal residues and corresponding secondary structure (in red); (2) the orientation of the tag residues in N-terminal (in dark blue); (3) the length of the helix structure highlighted by the dark blue arrow; (4) the subtle difference in orientation and length of some of the loops between the beta-sheets; and (5) the presence of a helix structure (yellow-colored loop) in the RK-GFP_R model but not in the GFP_s1M.

As described before, the interest in proceeding with the two models in parallel is to do a systematic search for structural inconsistencies and to avoid any computational bias.

Table 4.10 - Summary of the comparative homology modeling evaluation for the homology models obtained for the RK-GFP amin oacid sequence. The best result for each score is highlighted in green.

	Template	Identity [2]	ANOLEA [3]	GROMOS [4]	RMSD (Å)	SOLVX	VERIFY3D [5]	QMEAN (z-score)						
								Overall Score	Cβ interaction	All-atom interaction	Solvatation	Torsion	Secondary structure agreement	Solvent accessibility agreement
SWISS-MODEL [1]	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Modeller	1QYO	98.31%	89.53%	53.88%	1.024	-139.3	0.78% 5.81%	0.4	-1.07	-1.46	0.45	-1.51	0.8	1.13
Rosetta	2WUR **[6]**	97.90%	94.96%	69.38%	0.902	-139.4	3.10% 4.56%	-0.64	-1.24	-1.51	-0.28	-0.68	0.66	-0.42
I-TASSER	1B9C, 3NED, 1KYP, 3U8P, 2VZX, 3CGL **[7]**	**[8]**	45.35%	27.48%	6.2 **[9]**	-143.3	1.94% 0.78%	-1.43	-1.17	-2.14	0.63	-4.01	-0.89	0.57

[1] SWISS-MODEL could not model the RK-GFP sequence due to the tag residues having no alignment correspondence.
[2] The value corresponds to the identity when aligned to the GFP sequence.
[3] Percentage of residues that were scored positively (>0) by the ANOLEA score.
[4] Percentage of residues that were scored positively (>0) by the GROMOS score.
[5] Percentage of residues that were scored as bad (0-1.1; first line) and very bad (<0; second line) by the VERIFY3D score.
[6] Protocol does not allow choosing the template, which is chosen based on the multi pair-wise alignment preformed by HHpred.
[7] Most of the selected templates were also chosen for the GFP homology modeling_ .
[9] RMSD value calculated by the I-TASSER webserver.

4.4.2. Structural relaxation of homology models

MD simulations of the models found for RK-GFP (RK-GFP_s1M and RKGFP_R) were carried out. Analyses of the obtained results are summarized in Figure 4.15 for the RK-GFP_s1M and Figure 4.16 for the RK-GFP_R.

By Figure 4.15a, one can see that the predicted C-terminal associated α-helix structure of RK-GFP_s1M vanished after 10ns of simulation. This unfolded structure was also shown by the GFP_s1M after structural dynamical relaxation (see Figure 4.5). The conserved β-barrel structure of the GFP protein [75] undergoes some point flexibility compared to the tag's flexibility, as represented by the variation of the RMSF values in Figure 4.15b, in blue. The profile of RMSD has higher values than the ones for the GFP_s1M model (Figure 4.15b, in red) and this is a direct result of the high flexibility of the N-terminal tag residues (Figure 4.15c, in blue, and Figure 4.15d, in red), which deviate from the reported 0.5-2.0Å for proteins in equilibrium [78]. Although the unfolding of the C-terminal residues resulted in high RMSF values, higher values of RMSF are obtained for the flexible N-terminal tag residues.

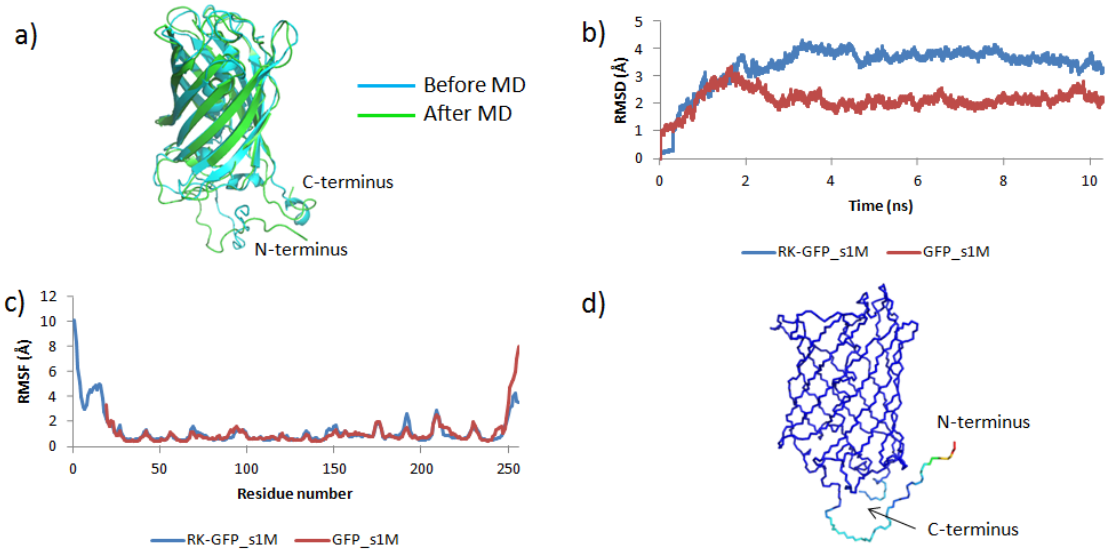


Figure 4.15 - Results obtained after the first MD simulation for the RK-GFP_s1M. a) Superimposition of the model obtained by Modeller before MD simulation (in blue) and after MD simulation (in green). b) Variation of the RMSD values along the MD simulation trajectory. c) RMSF values per residue. d) B-factor representation of structure flexibility colored in gradient. Blue represents residues with less flexibility and red represents residues with more flexibility average over the entire trajectory.

The RK-GFP_R had a very similar behavior to the one predicted by Modeller (Figure 4.16a). The N-terminal residues remain with no associated secondary structure and solvent-exposed. The helix-structure highlighted by the yellow arrow in Figure 4.14 is lost and the long helix-

structure in the C-terminal residues is reduced, as shown in Figure 4.16a2. The RMSD values tend to converge among the two models (RK-GFP_R and RK-GFP_s1M) despite the RMSF values for the N-terminal residues are lower for RK-GFP_R (Figure 4.16c). This can result from the close proximity between the N- and C-terminal residues as shown in Figure 4.16a1.

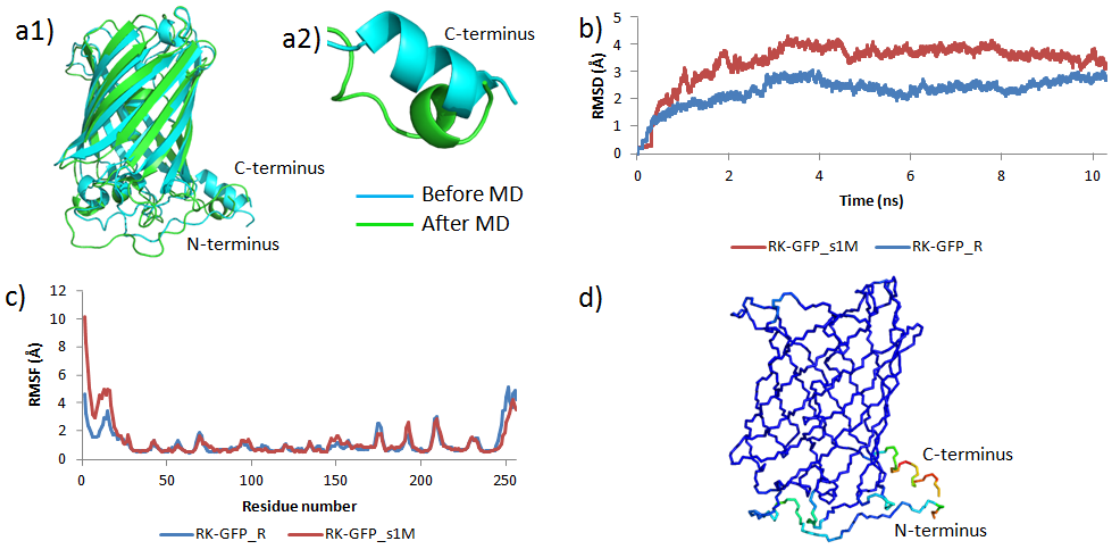


Figure 4.16 - Results obtained after the first MD simulation for the RK-GFP_R. a) Superimposition of the model obtained by Rosetta before MD simulation (in blue) and after MD simulation (in green). b) Variation of the RMSD values along the MD simulation trajectory. c) RMSF values per residue. d) B-factor representation of the structure flexibility colored in gradient. Blue represents residues with less flexibility and red represents residues with more flexibility average over the entire trajectory.

4.4.3. Automated molecular docking

The homology models obtained for RK-GFP were used as receptors to dock the affinity ligands LR-A7C1 (lead ligand) and NR-A5C2 (worst ligand). The coordinates of the LR-A7C2 (Figure 4.17a) and NR-A5C2(Figure 4.17b) were set up based on previous experimental solid phase affinity results of the complex [47]. The experimentally determined affinity constant for the affinity pair RK-GFP/LR-A7C2 was Ka= $2.45\times10^{5}M^{-1}$.

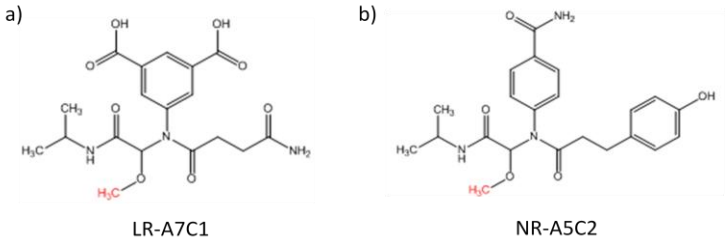


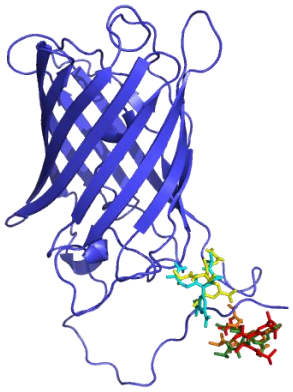
Figure 4.17 - Molecular structures of the solid phase screened ligands for RK-GFP system. a) Lead ligand with the best affinity for the RK-GFP and b) Negative control ligand with the worst affinity for RK-GFP system. The chemical group used to immobilize the ligand into the solid support is highlighted in red.

After assigning the correct atomic parameterization and molecular topology to the ligands, they were docked onto the receptor using the Lamarckian genetic algorithm as implemented in the Autodock software package.

Docking results of the lead candidate (LR-A7C1)

The results for the affinity system RK-GFP/LR-A7C1 are summarized in Table 4.11.

Table 4.11 - Summary of the docking results for the affinity system RK-GFP/LR-A4C1.

Target protein	Cluster Rank	Number of solutions per cluster	Model			
			Solution	Binding Energy (kcal/mol) [1]	Ka ($x10^5 M^{-1}$)	Ligand binding site
RK-GFP_s1M	1	19	9	-7.69 -7.64/-2.71	4.33	
	2	4	148	-7.13 -5.97/-3.34	1.06	
	3	7	183	-6.92 -5.76/-2.75	1.00	
	4	6	66	-6.81 -6.7/-2.29	0.98	
	5	2	211	-6.40 -6.27/-2.14	0.49	
RK-GFP_R	1	2	170	-7.18 -5.71/3.7	1.55	
	2	6	163	-7.07 -6.00/-3.48	1.50	
	3	7	160	-6.91 -7.20/-2.28	1.17	
	4	4	72	-6.60 -6.38/-2.63	0.69	
	5	6	230	-6.56 -5.77/-3.30	0.65	

[1] Estimated binding energy (on top) and two of its components: van der Waals, hydrogen bonding and desolvation energy/electrostatic interaction (on bottom).

Overall, top ranking hits achieved very good energies of binding and correspondent docking clusters are highly populated, especially for the RK-GFP_s1M receptor. Interestingly, all five top ranking poses are located near the N-terminal tag residues.

The first ranked model for the RK-GFP_s1M receptor (Solution 9) might interact via hydrogen-bonds with tag residues (R3, K4 and K6) as with other residues (G246 and N216). The anchoring chemical group used for immobilization is solvent-exposed which fulfills one of the

criteria stated before. Moreover, the predicted Ka=4.33x$10^5$$M^{-1}$ is in the same order of magnitude as the experimentally determined Ka=2.45x$10^5$$M^{-1}$. Solution 9 was the solution with the lowest estimated free binding energy (-7.69kcal/mol). On the other hand, cluster 1 is highly populated (19 solutions).

Solution 148 is the second top ranking solution for this system. This solution has a very similar binding site to the previous model, with a different orientation. In this solution, the ligand is located at a distance consistent with the formation of hydrogen bond interactions with R1, K2 and R5. Solution 148 is an interesting model to investigate further since observed interactions occur to different N-terminal tag residues than Solution 9. On the other hand, all other criteria are fulfilled.

No more solutions listed for the RK-GFP_s1M receptor were further analyzed as the corresponding clusters were not as populated and the binding location is very similar to the previous models.

The affinity system RK-GFP_R/LR-A7C1 presents good binding energies, although the clusters are not as populated as for the previous homology model. This receptor presents two main binding sites locations. Solution 160, from Cluster 3 in the ranking (in green) was discarded due to higher value of estimated energy of binding.

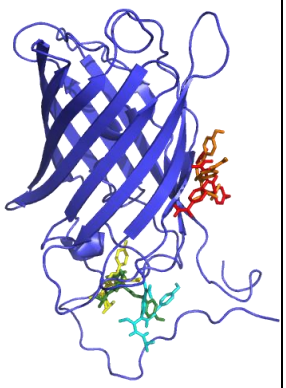
The binding site of Solution 170 in Cluster 1 allows hydrogen-bond interactions with N-terminal tag residues, namely R1, K4 and R5. The predicted Ka=1.06x$10^5$$M^{-1}$ has the same order of magnitude as the experimentally determined Ka and the anchoring chemical group necessary to bind to the resin is solvent-exposed. This model was chosen to proceed with molecular dynamics simulations.

The next top ranked model, Solution 163, also satisfies the given assumptions about the support-binding group, Ka and preferential binding to N-terminal residues. This is the docking solution which includes a higher number of possible interactions with different tag residues (R1, R3, K4 and R5).

Docking results of the negative control candidate (NR-A5C2)

The results for the affinity pair RK-GFP_s1M/NR-A5C2 are summarized in Table 4.12.

Table 4.12 - Summary of the docking results for the affinity system RK-GFP/NR-A5C2.

Target protein	Cluster Rank	Number of solutions per cluster	Model			
			Solution	Binding Energy (kcal/mol) [1]	Ka ($x10^5 M^{-1}$)	Ligand binding site
RK-GFP_s1M	1	2	238	-7.96 -9.61/0.50	6.80	
	2	14	164	-6.73	0.86	
	3	1	160	-6.68 -9.5/-0.11	0.78	
	4	4	27	-6.67 -8.38/-0.24	0.77	
	8	1	89	-6.38 -8.97/0.05	0.47	
RK-GFP_R	1	1	173	-8.18 -10.28/-0.46	9.90	
	2	1	240	-7.13 -10.28/-0.46	1.67	
	3	1	227	-6.89 -9.10/-0.51	1.13	
	5	1	184	-6.47 -7.12/-0.42	0.55	
	8	4	243	-6.29 -7.89/-0.35	0.41	

[1] Estimated binding energy (on top) and two of its components: van der Waals, hydrogen bonding and desolvation energy/electrostatic interaction (on bottom).

Contrary to what was expected for a negative control ligand, docking solution revealed very negative estimated free energies of binding. However the docking sites are much more disperse and most docking solutions are not located near the tag residues.

The first ranked solution for the receptor RK-GFP_s1M (Solution 238) binds underneath the structure, to the loop between the central helix structure and β4, leaving the chemical group necessary to bind to resin unavailable. The second top-ranked model (Solution 164) has a similar binding site to the previous hit. The calculated Ka is an order of magnitude lower ($0.8\times10^5 M^{-1}$) than the experimental one, but in this case the resin-binding group is solvent

accessible. These models were not chosen to proceed with the modeling task due to the non-relevant binding location.

Solution 160 of cluster 3 was chosen to proceed as there is a possible hydrogen bond interaction between the ligand and K6 and the resin-binding group is solvent-accessible. It has a predicted Ka= $0.8\times10^{5}M^{-1}$, one order of magnitude lower than the experimentally determined Ka for LR-A4C7.

The next four models do not have a binding site near the tag residues, so they were not further analyzed.

Solution 89, the eighth model for the RK-GFP_s1M receptor represents a promising docking solution. The resin-binding group is solvent-accessible and there is a possible hydrogen bond interaction with K6 and H95. It has a low Ka=$0.4\times10^{5}M^{-1}$ compared to the experimentally determined Ka for LR-A4C7. This model was chosen to proceed for MD simulations, as a negative control.

Regarding the analysis for the affinity pair RK-GFP_R/NR-A5C2 the clusters are not as populated as for the pair RK-GFP_s1M/NR-A5C2, but there are docking results that display good estimated free energy of binding. Most docking solutions do not bind near the N-terminal tag residues, while others could not binding to the solid support. The models from cluster 5 (in orange) and cluster 8 (in purple) will be further analyzed and were chosen to precede with the MD simulations.

Solution 184 is the only solution of cluster 5 and its docking site is located near the N-terminal tag residues establishing a hydrogen bond interaction with R5. The resin-binding group is solvent-exposed and it has a lower Ka=$0.5\times10^{5}M^{-1}$, one order of magnitude smaller than the calculated Ka values for the lead ligand.

Solution 243 belongs to eighth ranked cluster, but it was analyzed because of its binding location, at interacting distance through hydrogen bonds with R3 and D252 (located near the C-terminal). The estimated free energy of binding and the calculated Ka are similar to the last analyzed solution.

4.4.4. Extensive Molecular Dynamic Simulations

The docking solutions chosen to proceed with the MD simulations are summarized in Table 4.13.

Table 4.13 - Chosen docking poses to proceed for extensive MD for the different affinity pairs analyzed with the RK-GFP system.

Affinity Ligand	Homology Model	Chosen Solution
LR-A7C1	RK-GFP_s1M	9 148
	RK-GFP_R	170 163
NR-A5C2	RK-GFP_s1M	160 89
	RK-GFP_R	184 243

At this stage we were looking for significant interactions (>10% of the simulation time) between the best ligand LR-A7C1 and RK-GFP and no significant interactions between the negative control ligand NR-A5C2 and RK-GFP. Mostly hydrogen bond interactions were studied as the tag residues have a hydrophilic character in Nature. The results are summarized in Table 4.14 and the atoms of the ligands involved in the interactions analyzed are shown in Figure 4.18.

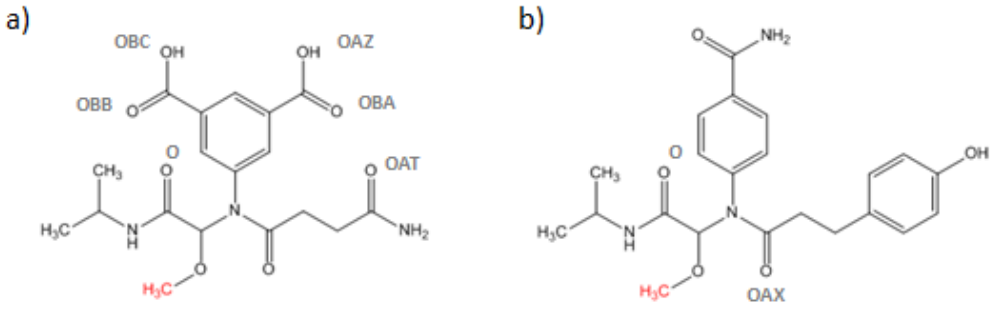


Figure 4.18 - Ligands (a) LR-A7C1 and (b) NR-A5C2 and the atoms involved in the studied interactions in Table 4.14.

Table 4.14 - Extensive MD simulation results regarding the interactions between the ligands and RK-GFP.

Ligand	Receptor	Model	Main interaction	Type interaction [1]		Contribution (%)
LR-A7C1	RK-GFP_s1M	9	LYS2:N---MOL:OAZ		Hb	6.87
			LYS2:N^{ζ}---MOL:OBA		Hb	3.3
			LYS6:N^{ζ}---MOL:OBA		Hb	0.12
		148	LYS2:N---MOL:OAZ		Hb	22.14
			LYS2:N---MOL:OBA		Hb	21.6
			LYS2:N^{ζ}---MOL:O		Hb	8.83
			LYS2:N---MOL:OBC		Hb	5.11
			LYS2:N---MOL:OBC		Hb	5.11
			ARG1:N^{ε}---MOL:OBB		Hb	1.57
			ARG1:N^{ε}---MOL:OBC		Hb	1.48
			ARG1:$N^{\eta 2}$---MOL:OBC		Hb	0.89
	RK-GFP_R	170	LYS4:N^{ζ}---MOL:O		Hb	1.81
			ARG1:N---MOL:OAT		Hb	1.4
		163	LYS4:N---MOL:OAT		Hb	6.54
			ARG5:N---MOL:OBB		Hb	0.97
			ARG1:N^{ε}---MOL:OAT		Hb	0.92
NR-A5C2	RK-GFP_s1M	160	No detected interaction			
		89	LYS6:N^{ζ}---MOL:OAX		Hb	3.85
			LYS6:N^{ζ}---MOL:O		Hb	2.65
	RK-GFP_R	184	No detected interaction			
		243	ARG3:N^{ε}---MOL:OAX		Hb	5.86

Note: GFP atoms are identified according to the recommended atom identifiers following the 1969 IUPAC-IUB guidelines (adapted from [100]).
[1] The interactions are classified as hydrophobic (Hyd) or hydrogen bond (Hb).

As expected the docking solutions corresponding to NR-A5C2 had fewer interactions with the protein and did not give any hits on significant hydrogen bond interactions. NR-A5C2 has lower experimental Ka and that was supported by the MD results. The longest interactions found represent less than 6% of the simulation time and occur mostly at the beginning of the simulation (near the initial docking pose).

A coherent behavior was observed for the LR-A7C1/RK-GFP system. Most of the models studied did not present significant interactions between RK-GFP and the ligand, breaking up

this complex after a few nanoseconds. In addition to presenting the higher number of interactions with RK-GFP Solution 148 showed an interesting feature (Figure 4.19a).

This interaction is a hydrogen bond between the second lysine residue of the tag and two adjacent oxygen atoms of ligand LR-A7C1 (in Figure 4.19b). Both oxygen atoms contributed for a hydrogen interaction that lasts 42.89% of the total simulation time. This is an on/off interaction: when LYS2 is not interacting with oxygen atom, it is able to interact with the adjacent oxygen atom. Although the ligand in some part of the simulation is detached from the protein, the interaction is strong enough to occur again at the end of the simulation time between both oxygen atoms.

The same functional group is found on the other side of the molecule and it would be predictable that there would be an interaction between these atoms and the protein. However this was not found during the MD conformational search.

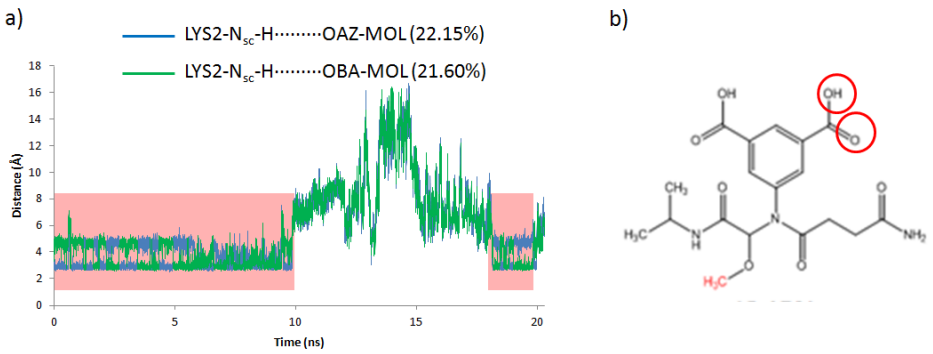


Figure 4.19 - (a) Variation of the distance between LYS2:N and LR-A7C1:OAZ and between LYS2:N and LR-A7C1:OBA during the second MD trajectory. Results regarding RK-GFP_s1M and Solution 148. Notes: BB- backbone; SC – side-chain. (b) Ligand LR-A7C1. The atoms OAZ and OBA are circled in red.

4.5. Conclusions and Future Directions

The results for the systems GFP and RK-GFP and their correspondent lead and negative control ligands were quite promising to explain the experimental results previously obtained in [47].

Many relevant hydrophobic interactions were found between GFP and LA-A4C7, the best ligand, over 90% of the MD simulation that can explain the experimental results. The interactions between GFP and NA-A7C1 were less predominant during the MD simulation, which is coherent with the decreased affinity towards GFP [47].

On the other hand, for the RK-GFP receptor only one of the docking solutions was in agreement with the experimental results [47], with an interaction over 40% of the simulation time. All other models had interactions that were present less than 10% of the simulation, including docking of the best ligand, which creates doubts about some key aspects.

First, is the sampling efficient enough? Choosing more docking poses to study by extensive MD would give a higher probability of finding more than one accordant solution. However it is computationally demanding.

Second, would a larger sampling influence the negative results that were obtained for the negative control ligand? However doing a significant sampling is very time consuming and data extensive.

Third, one of the criteria to choose the docking poses was that the ligand would bind near the tag residues in RK-GFP. Although most of the best-scored clusters were located near the RK-tag, this location can provide few stabilizing interactions. The results found for the GFP/NA-A7C1 system in section 4.2.4 confirm this hypothesis – there can be other alternative binding sites for ligand LR/NA-A7C1. When a different docking pose was chosen (e.g. Solution 248) some relevant interactions were found to occur between the ligand and the GFP surface. However when comparing the docking results for two examples (e.g. GFP/NA-A7C1 M248 and RK-GFP/LA-A7C1 M148) the first model has a more positive binding energy and a lower Ka value, which supports the selectivity of the ligand NA/LA-A7C1 to the tag residues.

Other important remarks have to be done about this theoretical study for both for GFP and RK-GFP systems.

We studied a simplification of the system, in terms of protein, ligand and salt concentrations. MD methodological issues are related with solution conditions, the accuracy of the force field and sufficient sampling [79].

On the other hand the ligand is in its free form (soluble in solution), while experimentally it is tethered to a chromatographic surface. Although some measures were taken to take this into account, its immobilization affects the ligands' mobility and conformational space available for protein interaction, as described in Freed *et al* [79].

Therefore, this study can only be used as a first approach to explain the experimental results previously obtained [47].

Other interesting studies could be done. One would include using the same systems while simulating the elution conditions used by Pina [47] or with different elution conditions (pH, ionic strength).

Regarding the design of a 2nd generation library towards GFP, all 56 initial ligands can be synthesized on agarose beads or other solid supports and the experimental results can be compared to the theoretical results. One should notice that the results can vary due to the docking limitations. Furthermore, the binding and elution conditions can be optimized for the new affinity system.

5. MAGNETIC PARTICLES FOR PURIFICATION OF FUSION PROTEINS

5.1. Purification methodologies for GFP

The GFP gene can be manipulated using standard molecular biology tools and it is widely used to create fusion constructs and labelled-proteins [104], to monitor protein expression [47], study protein stability [105], folding/unfolding phenomena [106], overexpression [107] and many other applications. However, few purification techniques for GFP-fusion proteins have been described.

GFP recovery is mostly based on IMAC [108,109] or Q-Sepharose columns [110,111]. An organic and chromatographic extraction has been described [112], as well as Ni-NTA-functionalized magnetic particles for the purification of His-tagged GFP [113]. The His-tagged version of nanobodies described in [101] was used to bind GFP to a Ni-NTA column. Fukuda *et al* have described a four step purification protocol that relies on differential precipitation, two gel filtrations and ion-exchange chromatography [114]. Commercially, anti-GFP antibodies-based affinity purifications are available. Milteniyi offers superparamagnetic microbeads (μMACS® technology) and Chromotek has monovalent matrices (agarose beads, magnetic particles or 96-multiwell plates) functionalized with GFP-TRAP® for the purification of GFP and GFP-fusion proteins.

Therefore there is a need for a rapid and efficient purification method for GFP. In [47] a novel affinity ligand for GFP has been tethered to agarose beads and allowed the recovery of a very pure GFP fraction.

5.2. Purification methodologies for tagged-proteins

Fusing an affinity tag to the target protein allows its purification, with high yields and purity. Nowadays many options are available according to the final application. A summary of some available tags and their correspondent affinity ligand is available in Table 6.1

Table 6.1 - Examples of currently used tags and specific affinity ligands in protein purification strategies. Adapted from [17,20].

Tag	Size (aa)	Ligand	Other properties	Examples
His-tag	5-15	Ni^{2+}-nitrilotriacetic acid, Co^{2+}-carboxymethylaspartate	-	[107]
FLAG	8	Anti-FLAG monoclonal antibody	-	[115]
Strep-tag	8	Strep-tactin (modified streptavidin)	-	[116]
HA-tag	9	Co^{2+}- carboxymethylaspartate	-	[117]
myc-tag	10	Monoclonal antibody	-	[117]
S-tag	15	S-fragment of RNaseA	-	[118]
Elastin-like peptides	18-320	Non-chromatographic	Response to ionic strength and temperature	[119]
Chitin-binding domain	52	Chitin	-	[120]
Gluthathione-S-tranferase	201	Glutathione	Increase protein solubility	[121]
Maltose binding protein	396	Cross-linked amylose	Increase protein solubility	[121]

There is a constant search for novel affinity tags that allow high purity and recovery of the target protein. In [47] the properties of novel proteic tags and their correspondent affinity ligand have been studied and allowed the recovery of a pure tagged-GFP fraction. In this work we will study the application of "RKRKRK" as affinity tag.

5.3. Objective

The small size and large surface area to volume ratio of magnetic nanoparticles in addition to their superparamagnetic properties make them highly attractive as solid supports for adsorptive separations [2]. Many proteins have been purified using magnetic techniques [1], where the affinity is conferred by the coating [7,122], metal chelants at the surface [113], affinity ligands (examples are shown in Table 1.1), or other molecules – see [11].

In this chapter we described the use of a new solid support for the purification of GFP and tagged-GFP. For this purpose, we accomplished two goals: (1) we have produced GFP and RK-GFP in *E. coli*, and (2) we have functionalized magnetic nanoparticles with the most promising affinity ligands described in [47] and shown in Table 2.1 for the purification of GFP and RK-GFP.

5.4. Production of GFP and RK-GFP in *E. coli*

5.4.1. Fragment DNA design and DNA purification

The DNA fragments that were cloned into the pET-21c plasmid by Geneart™ are shown in the Methods section, Figure 3.1, resulting in the plasmids pAP001 (expressing GFP) and pAP006 (RK-GFP). Both plasmids were amplified in NZY5α competent cells. Afterwards the plasmid DNA (pDNA) was isolated using the isolation NZYMiniprep kit and the final purity and concentration were determined by spectroscopic analysis (Table 5.1). An agarose gel analysis was performed to confirm that there had been no hydrolysis of the pDNA (Figure 5.1).

Table 5.1 - pDNA concentration and purity of the plasmids used for the large scale production of GFP and RK-GFP.

Plasmid (protein encoded)	Elution	DNA concentration (ng/µl)	DNA purity	
			A_{260}/A_{280}	A_{260}/A_{230}
pAP001 (GFP)	1st	76.4	1.76	1.76
	2nd	10.2	1.34	0.64
pAP006 (RK-GFP)	1st	152.4	1.80	1.81
	2nd	80.4	1.64	1.66

The isolation of both plasmids resulted in the successful recovery of pDNA with high concentration and purity for the first elution step. The second elution of pAP006 was too diluted and it is not visible in the agarose gel (Figure 5.1, lane 3). The purification and isolation of pAP006 resulted in a single distinct band, observed in lane 9 for the 1st elution and lane 11 for the 2nd elution (Figure 5.1).

The resulting DNA samples of the first elution were used to transform BL21(DE3) cells to obtain final crude extracts containing the protein of interest.

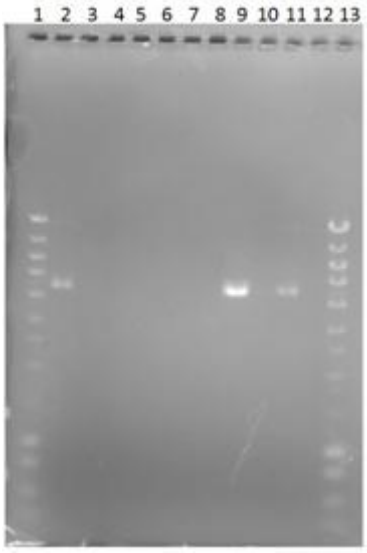


Figure 5.1 - Agarose gel for pDNA analysis. Lanes: 1- NZYDNA Ladder III, 2- pAP001 (GFP) 1st elution, 3- pAP001 (GFP) 2nd elution 9 -pAP006 (RK-GFP) 1st elution, 11- pAP006 (RK-GFP) 2nd elution, 13- NZYDNA Ladder III. Lanes 4-8, 10 and 12 are empty. The gels present 0.8% agarose and were afterwards stained with Green Safe.

5.4.2. Large scale production of GFP and RK-GFP

The optimal conditions for the large scale production of GFP and RK-GFP in BL21(DE3) cells were studied previously [47]. Protein production was monitored by fluorescence intensity and OD_{600nm} measurements (Figure 5.2) as well as SDS-PAGE analysis to visualize the relative amount of protein in the various fractions (Figure 5.3).

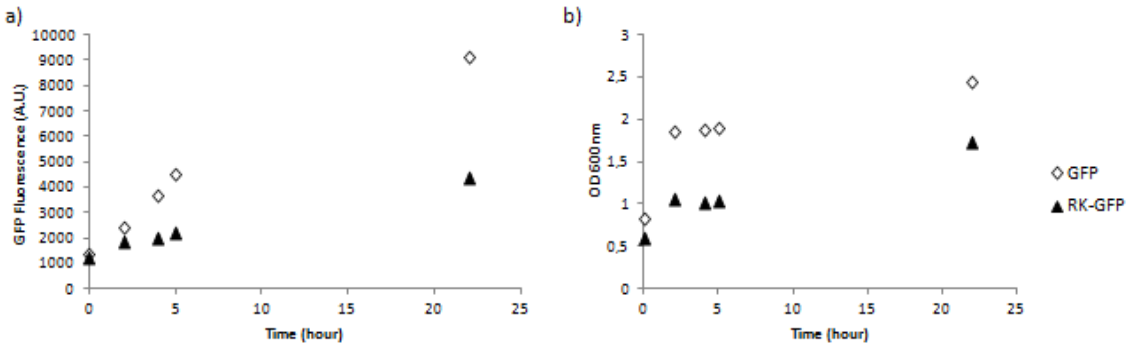


Figure 5.2 - (**a**) GFP fluorescence and (b) optical density measurements after the induction (t=0) during the large scale expression of GFP and RK-GFP.

We can see in Figure 5.2 that there is a correlation between the two measurements with the increasing induction time. The amount of protein produced by the cells increases with the time after induction and the amount of protein was significantly higher for GFP than for RK-GFP. This is confirmed by the SDS-PAGE analysis, as shown in Figure 5.3. It also shows that we only have protein production after IPTG-induction. GFP has a molecular weight of approximately 29kDa (confirmed by the literature [104]) and RK-GFP of approximately 31kDa.

The results from the analysis of the SDS-PAGE gels confirm the previous results: there is a higher production of GFP (GFP represents 25.7% of the total protein at 22h after induction) than of RK-GFP (8.8% of the total protein).

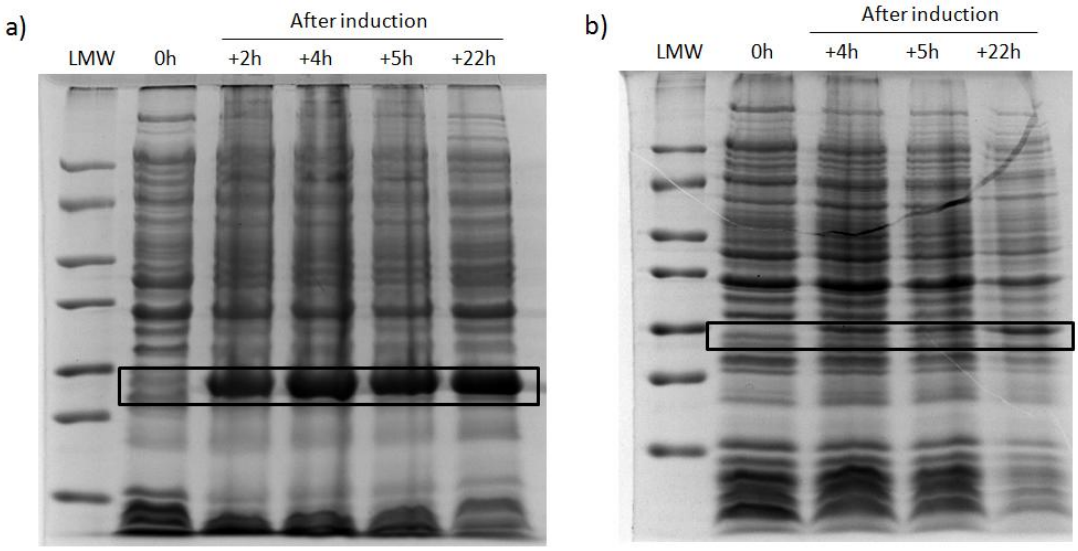


Figure 5.3 - SDS-PAGE time course of large scale production of (a) GFP and (b) RK-GFP in *E. coli* BL21(DE3). Lanes: LMW – Low molecular weight maker; 0h – sample collected before induction; + xh – sample collected x hours after the induction. GFP is highlighted inside the box. The gels present 12.5% acrylamide and were stained with Coomassie Blue.

Afterwards, the cells were harvested and lysed to obtain the final crude extract used for the screening assays in the next section.

For the growing conditions used in this work GFP is presented as a soluble protein [47] and was produced mostly in its soluble form, as shown in Figure 5.4c. Lane PC in Figure 5.4a should contain the pellet sample obtained after the low speed centrifugation and as shown in Figure 5.4c there is a very low amount of GFP present in this sample. However a large band is visible as a result from a contamination from the adjacent lane. The supernatant sample obtained after the ultracentrifugation step (SU) contains a high amount of GFP and was used for the screening assays.

RK-GFP is produced mostly as soluble protein (Figure 5.4b). After the ultracentrifugation, most of the protein is located in the supernatant in high concentration. This sample was used for the screening assays. As shown before and as is corroborated in Figure 5.4c, RK-GFP was produced in less concentration.

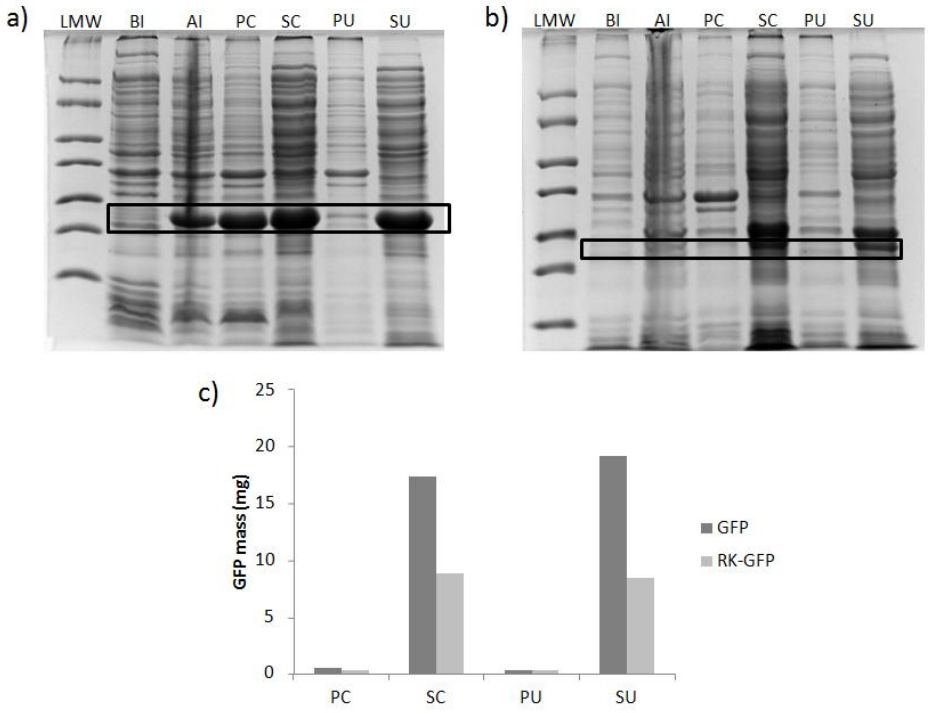


Figure 5.4 – SDS-PAGE cellular fractionation gel of (a) GFP and (b) RK-GFP in *E. coli* BL21(DE3). Lanes: LMW – Low molecular weight maker; BI - sample collected before induction; AI – sample collected 2h after induction; PC – sample of ressuspended pellet after low-speed centrifugation; SC – sample of supernatant after low-speed centrifugation; PU – sample of ressuspended pellet after ultra centrifugation; SU – sample of supernatant after ultra centrifugation. The protein of interest is highlighted inside the box. The gels present 12.5% acrylamide and stained with Coomassie Blue. (c) Amount of GFP (in mg) present in each sample of the previous gels.

5.5. Magnetic particle selection and characterization

Magnetic particles were produced by the co-precipitation method and their surface was chemically modified with the affinity ligands LA-A4C7, NA/LR-A7C1 and NR-A5C2. The particles were then characterized according to their hydrodynamic diameter and zeta potential.

DLS is a widely used technique for the determination of average particle size in a colloidal solution [123]. This technique is based on the interaction between a source of light and the particles present in a solution. The scattering effect depends on the Brownian movements of the particles in solution, which correlates with the translational diffusion coefficient. The Stokes-Einstein equation is used to estimate the mean size distribution of the particles within the sample [124]. This measurement will depend on the particle size, surface structure, concentration and medium.

The dispersed particles in a colloidal sample will have a charged surface due to ionization or adsorption of charged species. Therefore, there will be an electric double layer surrounding

the particle: the Stern layer where ions are more strongly attached, and an outer, more diffused layer. The zeta potential relates with the potential at the slipping plane, a notional boundary where the particles and the ions of the outer layer form a stable entity.

The magnitude of the zeta potential can be related with the potential stability of the colloidal system. If a large zeta potential is measured (higher than +30mV or lower than -30mV), the repulsion forces will prevail and the particles will remain dispersed in solution. For zeta potential values in between, the particles tend to flocculate. The zeta potential is influenced by temperature, ionic strength and particularly by pH [125].

In this section we analyze the hydrodynamic diameter and zeta potential of the MNP_DEX, MNP_DEX_NH_2 and the aminated-supports functionalized with the affinity ligand for GFP (MNP_DEX_NH_2_LA-A4C7) and for RK-GFP (MNP_DEX_NH_2_LR-A7C1), as well as negative controls (MNP_DEX_NH_2_NA-A7C1) for GFP, and MNP_DEX_NH_2_LR-A5C2 for RK-GFP) [47].

All samples were measured using the concentration 0.05mg/ml in deionized water (pH 5.80). The results are shown in Figure 5.5.

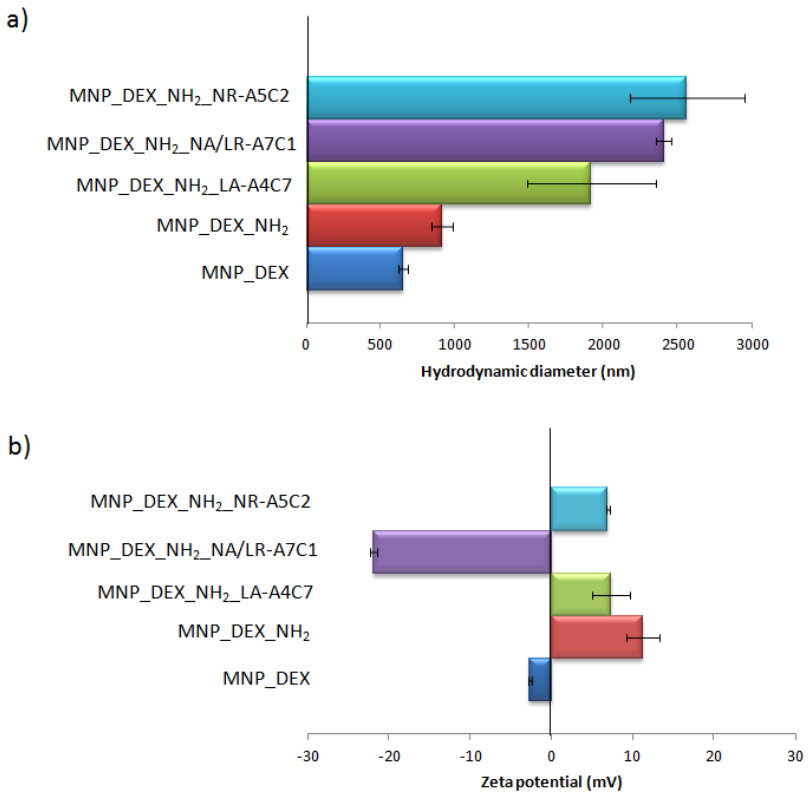


Figure 5.5 – (a) Hydrodynamic diameter (nm) and (b) zeta potential (mV) results for the magnetic supports used for the screening assay of GFP and RK-GFP (n=3).

MNPs synthesized by the co-precipitation methods usually have individual magnetic cores of 10-12nm in diameter [22]. However, their small size and large surface area tend to induce

aggregation [7]. Furthermore, additional functionalizations contribute for the increase of this property, as seen in Figure 5.5a. The coating material contributes for the increase of the particle diameter, due to the formation of bridges between coating molecules [22]. APTES can crosslink and form extended networks, contributing for the increase of the MNPs' diameter. Additional Ugi functionalization uses glutaraldeyde, which can react with two amine groups present on the surface of two adjacent MNPs and contribute for the formation of larger aggregates. If we attend to the error bars for the results of the ligand-functionalized MNPs (Figure 5.5a) we can see that their diameter is quite similar. However, DLS is a reliable technique for the measurement of homogeneous colloidal solutions with particle size lower than 1µm, which is not the case. Larger particles limit the DLS technique to diluted samples due to the contribution from multiple scattering [123]. Therefore, the results obtained by this technique are not very reliable and different equipments should be used for the determination of the size.

The zeta potential can be used to confirm the functionalization of the particle surface (Figure 5.5b). Bare magnetite nanoparticles have a isoelectric point of 6.8 [3], which would result in a positive zeta potential value. When coated with the neutral polysaccharide dextran, this value approaches zero [2]. Aminated-MNPs present low positive values of zeta potential, due to the amine groups at the surface of the MNPs [22]. The zeta potential of the functionalized-MNPs will be result of the pKa of the compounds that form the ligands. For LA-A4C7- and NR-A5C2-functionalized MNPs, the zeta potential is very similar to the non-functionalized MNPs. This could mean the lack of functionalization – which is not the case as shown in the next section – or a low yield of ligand coupling. On the other hand, the low pKa of the carboxyl groups in amine A7 contributes to lower the zeta potential of MNP_DEX_NH_2_NR/LR-A7C1.

Furthermore, all supports present zeta potential values between -30mV and +30mV, which confirms the low stability of the supports and aids the formation of large aggregates (see Figure 5.5a). However, for the purpose of protein purification large aggregates facilitate their recovery by a relatively weak permanent magnet.

The supports characterized above have been used for the purification of GFP (section 5.6) and RK-GFP (section 5.7) from crude extracts.

5.6. MNP as solid supports for the purification of GFP

For the purification of GFP, MNP_DEX_NH_2 were functionalized with LA-A4C7, the Ugi ligand that has been identified to display affinity towards GFP [47]. As negative controls, MNP_DEX_NH_2 were used during the screening, as well as MNP functionalized with NA-A7C1, due to its low affinity for GFP [47].

5.6.1. Solid support functionalization and ligand leaching

The amine A4 is 1-pyrenemethylamine which has a fluorescence signal ($\lambda_{excitation}$=485nm and $\lambda_{emission}$=535nm) – see Figure 5.6. It shows a clear fluorescence sign for MNP_DEX_NH_2_LA-A4C7, which is absent for the non-functionalized MNPs. This qualitative analysis cannot confirm 100% that the Ugi ligand is present on the MNP_DEX_NH_2, only that pyrene molecules are present on the surface of the MNPs.

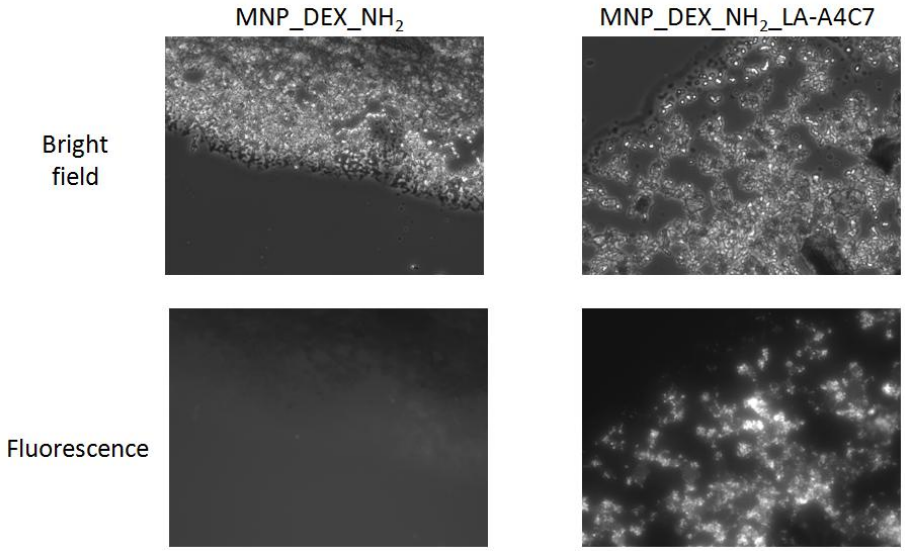


Figure 5.6 - Fluorescence microscopy results. Bright field images of MNP_DEX_NH_2 (a) and MNP_DEX_NH_2_LA-A4C7 (b). Fluorescence microscopy images (exposition time=1s) of MNP_DEX_NH_2 (c) and MNP_DEX_NH_2_LA-A4C7 (d).

Due to the pyrene's fluorescence in the same range as GFP, ligand leaching could influence the results. To determine this possibility a screening assay was performed using PBS instead of crude extract in the first step. All washes were quantified and the pyrene present was quantified using a pyrene calibration curve. No pyrene leaching was identified during the assay.

5.6.2. GFP binding to the solid supports

For the screening assay, the three supports described previously (MNP_DEX_NH_2, MNP_DEX_NH_2_LA-A4C7 and MNP_DEX_NH_2_NA-A7C1) were tested simultaneously for the purification of GFP. Different conditions have been tested: (1) MNP concentration, (2) the crude extract loading concentration and (3) the amination of the supports. The results will be discussed regarding two aspects: binding (Figure 5.7) and elution (Figure 5.8).

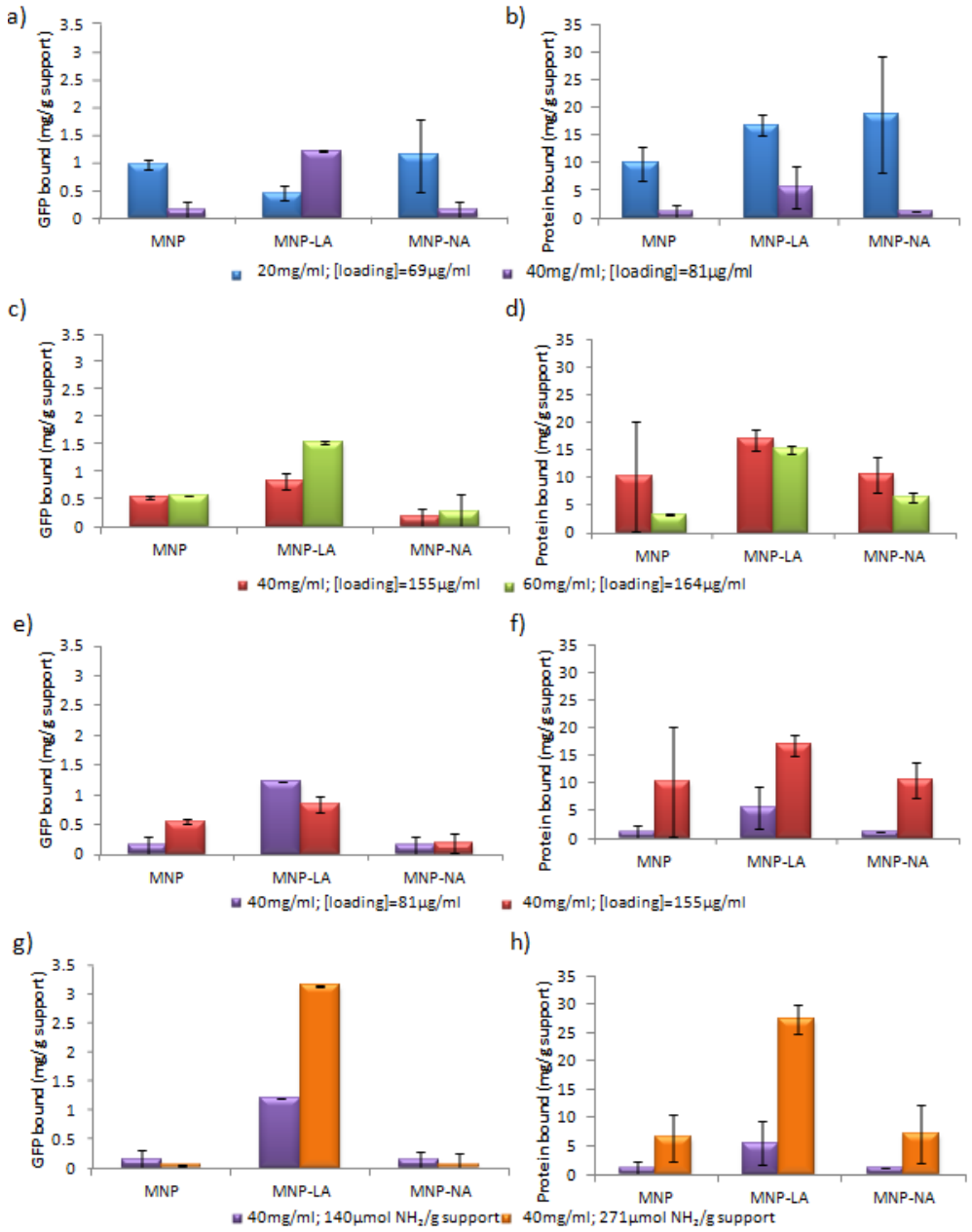


Figure 5.7 - Screening results for GFP binding regarding GFP (a, c, e and g) and total protein (b, d, f and h) with the following supports: MNP_DEX_NH_2 (MNP), MNP_DEX_NH_2_LA-A4C7 (MNP-LA) and MNP_DEX_NH_2_NA-A7C1 (MNP-NA). The first concentration in the legend is the concentration of MNP used, the loading concentration refers to the GFP concentration in the loading solution.

Influence of the MNP concentration

For this purpose two loading conditions have been tested: 75±8µg GFP/ml (Figure 5.7a and Figure 5.7b) and 159±6µg GFP/ml (Figure 5.7c and Figure 5.7d). It is important to refer that the MNPs used have the same amination content (140umol NH_2/g support). So assuming that we have a batch with a homogenous functionalization, we can test the effect of changing the initial MNP concentration.

For the first loading condition, two MNP concentrations were tested: 20mg/ml and 40mg/ml. A higher concentration of MNPs translated into a higher concentration of ligand LA-A4C7, resulting in a higher percentage of GFP binding. A higher ligand density translates into a higher protein adsorption [19]. Although the coating should contribute to lower the non-specific adsorption (as observed by a higher binding of IgG to bare MNPs *vs* gum arabic-coated MNPs [22]), the coating can contribute for non-specific binding and decrease the selectivity of the purification process [113]. This can explain the high amount of protein that binds to all the solid supports.

On the other hand, screening the crude extracts with 20mg/ml resulted in less GFP bound to the solid support MNP_DEX_NH_2_LA-A4C7 than to the controls. The low amount of ligand does not allow a specific binding of GFP.

The second loading condition tested (159±6µg GFP/ml) confirmed the previous results. A larger concentration of MNP translates into a higher amount of ligand and consequently a higher amount of GFP bound to the support.

Ditsch *et al* [7] have described a successful protein purification system using only 1mg magnetic adsorbent/ml. Although handling higher MNP concentration does not present any immediate disadvantage [122], working with a very high concentration of particles is hard to handle. Therefore, 40mg/ml is the best tested concentration for protein purification purposes.

Influence of the concentration of the loading sample

Herein we compared for the same concentration of MNP (40mg/ml), different loading concentrations: 81µg GFP/ml and 155µg GFP/ml (Figure 5.7e and Figure 5.7f).

A higher concentration of GFP in the loading solution does not result in a higher amount of GFP bound to the supports. This can be explained due to the higher concentration of other proteins presented in the more concentrated loading solution (confirmed by the BCA results,

Figure 5.7f). A higher percentage of non-specific interactions with other protein present in the crude extract do not allow GFP to bind to LA-A4C7.

Therefore, for a similar assay with the same incubation conditions, a more diluted loading solution allowed a higher percentage of GFP bound to the MNP_DEX_NH_2_LA-A4C7 support.

Influence of the level of MNP amination

Two different MNP samples have been tested, which had different amination contents (Figure 5.7g and Figure 5.7h) for similar loading concentrations (86±8mg GFP/ml). The amine density on solid supports is strongly influenced by reaction conditions [22], which makes it complicated to obtain evenly aminated supports. Performing the Kaiser test, we determined that the first batch of MNP had 140±4µmol NH_2/g support (from which all previous results are from) and the second batch had 271±52µmol NH_2/g before ligand attachment. If we assume that we have 100% amine functionalization, it confirms the results: a higher amination (which corresponds to a higher concentration of ligand at the surface of the MNPs) results in a higher amount of GFP bound to the solid support.

The determination of ligand concentration at the surface of the MNPs is a complicated task.

Authors have determined ligand density through the extent of epoxy-activation of the agarose beads used as solid support [15,19]. Similarly, we tried to use the amine content to determine ligand density. The results are summarized in Table 5.2.

Table 5.2 - Degree of amination before and after ligand coupling for both batches of MNP_DEX_NH_2_LA-A4C7 and other relevant information.

	MNP_DEX_NH_2_LA-A4C7	
	Batch 1	**Batch 2**
Amination		
Before functionalization (µmol NH_2/g support)	140±4	271±52
After functionalization (µmol NH_2/g support)	90±1	89±1
Reacted amines (µmol NH_2/g support)	49	182
Percentage of functionalization (%)	35	67

If we assume that the amines that react during the Ugi reaction are in a 1:1 proportion to the glutaraldeyde coupling reaction, the concentration of reacted amines will correspond to ligand concentration. This is still according to the results for GFP binding: a higher concentration of ligand translates into a higher concentration of GFP bound to the particles. On the other hand, a higher amount of amines can favor the reaction between the compounds in solid-phase and the compounds in liquid-phase, which will favor a higher functionalization percentage.

Unfortunately the Kaiser test can only be used as a semi-quantitative method and has several limitations, making it difficult to predict ligand concentration. It is performed under quite harsh reaction conditions (sample incubation at 100°C), which can reveal hidden amines or allow the release of other interfering components. If these conditions are harsh enough to promote ligand release, we can be measuring more free amines and increase the percentage of functionalization.

Based on the previous results, we can conclude that the best conditions for GFP binding would be to use a higher concentration of MNP (40mg/ml), with a higher amination content (271µmol NH_2/g support) and a more diluted loading solution (81µg GFP/ml). This is confirmed by the elution results, which will be presented below.

5.6.3. GFP elution from the solid supports

The elution conditions are dependent of each system, protein to purify, affinity ligand and solid-support used. For GFP elution, the magnetic microbeads µMACS® (Milteniyi) uses a denaturing elution buffer, containing low concentrations of dithiothreitol and SDS, at pH6.8, while the GFP-TRAP® uses an elution buffer with low pH (pH=2).

For all experimental conditions tested and presented in Figure 5.7, two elution conditions were screened in this work: 0.1mM glycine-NaOH pH9 (Figure 5.8a) and 0.1mM glycine-NaOH pH9 50% (v/v) ethylene glycol (Figure 5.8b), as previously studied in [47].

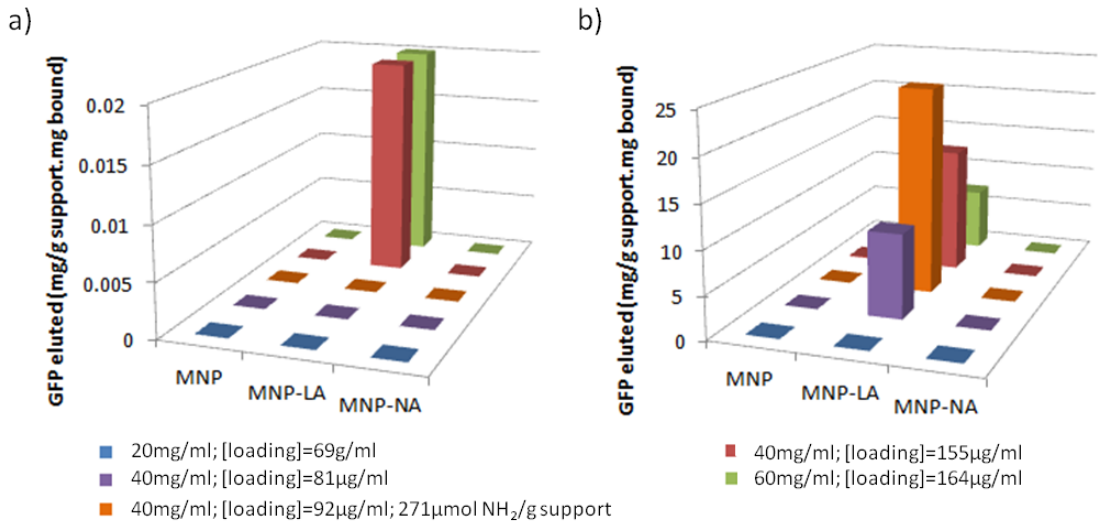


Figure 5.8 - Screening results for GFP elution with the following supports: MNP_DEX_NH_2 (MNP), MNP_DEX_NH_2_LA-A4C7 (MNP-LA) and MNP_DEX_NH_2_NA-A7C1 (MNP-NA), for (a) elution condition A and (b) elution condition B. The first concentration in the legend is the concentration of MNP used, the loading concentration refers to the GFP concentration in the loading solution.

The best elution condition to elute GFP was 0.1mM glycine-NaOH pH9 50% (v/v) ethylene glycol (Figure 5.8b). Ethylene glycol is responsible for disrupting hydrophobic interactions [126]. This result is in accordance with the results from section 4 – the interaction between GFP and LA-A4C7 is mainly hydrophobic. All tested conditions (except for the first (Figure 5.8b, in blue, due to low GFP binding) resulted in roughly 25% elution of the GFP bound to the solid support. The experiment where we had the most GFP eluted (Figure 5.8b in orange) corresponds to the experiment were we had the most GFP bound to the MNPs.

An important remark has to be done about these elution results, as ethylene glycol influences the GFP fluorescence, increasing the measured fluorescence [47]. Therefore, a lower amount of GFP has been eluted compared to the results in Figure 5.8b.

Another important parameter for protein purification is the purity of the protein during elution. His-GFP can be purified by a Ni-NTA column with a final purity over 95% [127].

BCA does not have enough resolution to quantify the low amount of protein that was eluted (quantification limit of 0.2mg/ml), while GFP allows a quantification limit of 0.0001mg/ml (according to the calibration curve used in this work). SDS-PAGE was used to evaluate the purity of the eluted GFP (Figure 5.9).

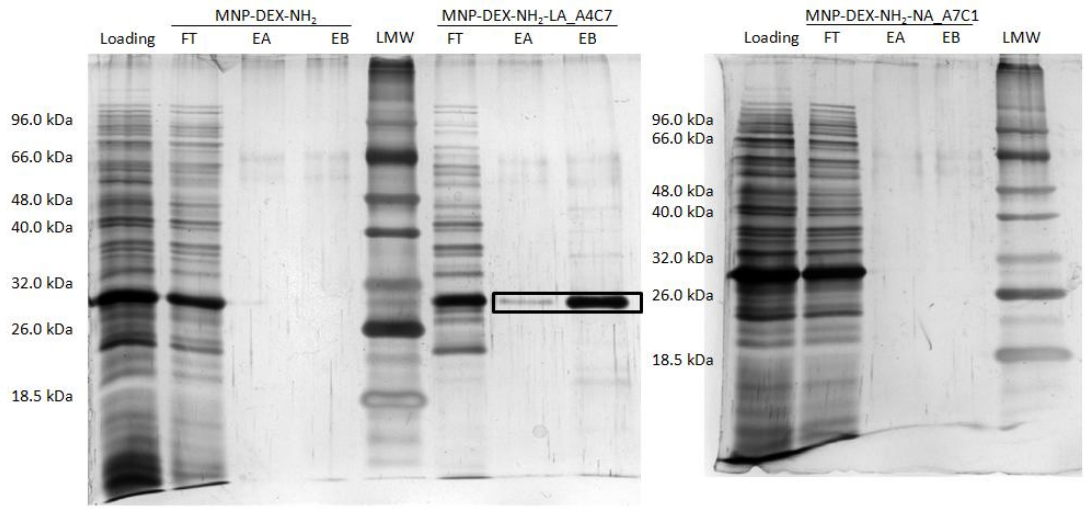


Figure 5.9 - SDS-PAGE analysis of the molecular weight marker (LMW) and flow-through (FT), first elution sample with elution condition A (EA) and condition B (EB) for all the supports tested. Results regarding 40mg MNP/ml, loading=91.6µg GFP/ml, 271µmol NH_2/g support (orange, in Figure 5.7 and Figure 5.8). The GFP corresponding bands are shown in a box. Gels stained by silver staining.

Figure 5.9 corresponds to the experiment where the best conditions were tested: 40mg MNP/ml of the most aminated batch (217µmol NH_2/g support) and less concentrated loading solution (92µg GFP/ml) - regarding Figure 5.7g and Figure 5.7h. No GFP was eluted when using the negative control particles. Purity and yield results regarding this system are summarized in Table 5.3.

Table 5.3 – Percentage of recovery, purity and yield of GFP calculated from the results presented in Figure 5.9.

Sample	GFP Recovery (%)	Purity (%)	Yield (%)
MNP_DEX_NH_2_LR-A7C1 EA	0 **[1]**	65	0 **[1]**
MNP_DEX_NH_2_LR-A7C1 EB	48	86	44

[1] Although there is a visible band corresponding to GFP in the MNP_DEX_NH_2_LR-A7C1 EA lane, the concentration was not quantifiable or measurable by fluorescence intensity.

The SDS-PAGE gel allows a qualitatively analysis that we have more protein binding to MNP_DEX_NH_2_LA-A4C7 than to the negative supports. The selectivity of this system is low (12%), as LA-A4C7 also binds to other proteins present in the crude extract. We have the most amount of GFP eluting from the LA-A4C7 functionalized-MNPs when treated with elution condition B. Although some other proteins are eluted when using elution condition B and we only achieved a yield of 44%, we were able to retrieve 46% of the total GFP bound to the support, with a purity of 86%. Although this value of purity is still low for an affinity-based

purification system, it is quite high for a novel system using a novel solid-support, where the used elution conditions have not been optimized.

5.7. MNP as solid supports for the purification of RK-GFP

The RKRKRK-tag has shown promising results for the purification of tagged-GFP using agarose beads [47]. For the purification of RK-GFP, aminated dextran-coated magnetic nanoparticles (MNP_DEX_NH_2) were functionalized with LR-A7C1, the Ugi ligand that has been identified to display affinity towards RK-GFP [47]. As negative controls, MNP_DEX_NH_2 were used during the screening, as well as MNP functionalized with NR-A5C2, due to its low affinity for RK-GFP [47].

Based on the GFP screening results, an initial concentration of 40mg/ml of MNPs has been used for this screening. Only one loading condition was tested (12µg RK-GFP/ml), and the amount of RK-GFP and total protein bound and eluted from the magnetic supports is shown in Figure 5.10 and Figure 5.11, respectively.

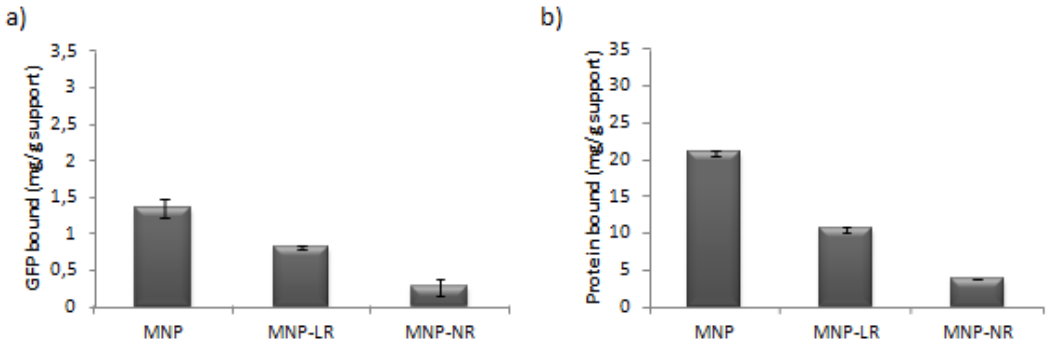


Figure 5.10 - Screening results for RK-GFP binding regarding (a) RK-GFP and (b) total protein with the following supports: MNP_DEX_NH_2 (MNP), MNP_DEX_NH_2_LR-A7C1 (MNP-LR) and MNP_DEX_NH_2_NR-A5C2 (MNP-NR). Experimental conditions: 40mg MNP/ml, loading 12µg RK-GFP/ml, 271µmol NH_2/g support.

Contrary to what we observed for the purification of GFP, we observed a higher amount of GFP binding to the non-functionalized MNPs than to MNP_DEX_NH_2_LR-A7C1. (Figure 5.10a), as well as non-specific binding to total protein (Figure 5.10b). This system is even less selective than the affinity system LR-A7C1/GFP, with only 8% selectivity. We could only bind 0.81±0.03mg RK-GFP/g support, while we had a binding maximum of 3.13mg GFP/g support. This can be for two reasons. On one hand, the ligand LR-A7C1 has a lower affinity for RK-GFP than LA-A4C7 for GFP. On the other hand, the lower concentration of RK-GFP in the crude extract facilitates the unspecific binding of other proteins, limiting the contact between RK-

GFP and the ligand. This can be supported by the results obtained for the study of the loading concentration influence for the GFP binding assay: a higher concentration of loading solution did not result in a higher amount of GFP bound (see Figure 5.7e).

However these conditions should be optimized in order to obtain a higher amount of RK-GFP bound, as only 10% of the RK-GFP present in the loading solution was bound to MNP_DEX_NH_2_LR-A7C1.

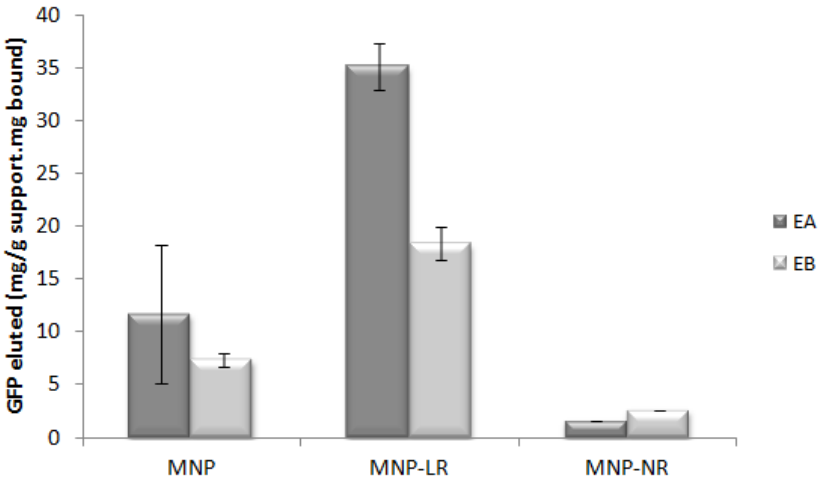


Figure 5.11 - Screening results for RK-GFP elution with the following supports: MNP_DEX_NH_2 (MNP), MNP functionalized with ligand LR-A7C1 (MNP-LR) and MNP functionalized with ligand NR-A5C2 (MNP-NR), for elution condition A (EA) and elution condition B (EB).

Figure 5.11 summarizes the elution results for RK-GFP. The elution conditions tested were PBS pH 7.4, 500mM arginine (EA) and 0.1M glycine-NaOH pH 11, 150mM NaCl (EB). Contrary to the GFP results (Figure 5.8), we had elution in all cases, including for the negative controls. However, we can see that we had a higher amount of RK-GFP eluted from MNP_DEX_NH_2_LR-A7C1 in comparison with the negative controls.

A higher concentration of RK-GFP was eluted when using EA, which is based on competition. These results are not according to the ones obtained in the modeling studies (Section 4) - the interactions were mainly between LR-A7C1 and the lysine residues in the tag. However, contrary to the results regarding GFP/LA-A4C7, we did not obtain very high percentages of interaction and it may not be able to explain theoretically the main interactions between LR-A7C1 and RK-GFP. However we could retrieve 70% of the RK-GFP bound to MNP_DEX_NH_2_LR-A7C1.

SDS-PAGE analysis was performed to access RK-GFP purity in the elution factions (Figure 5.12).

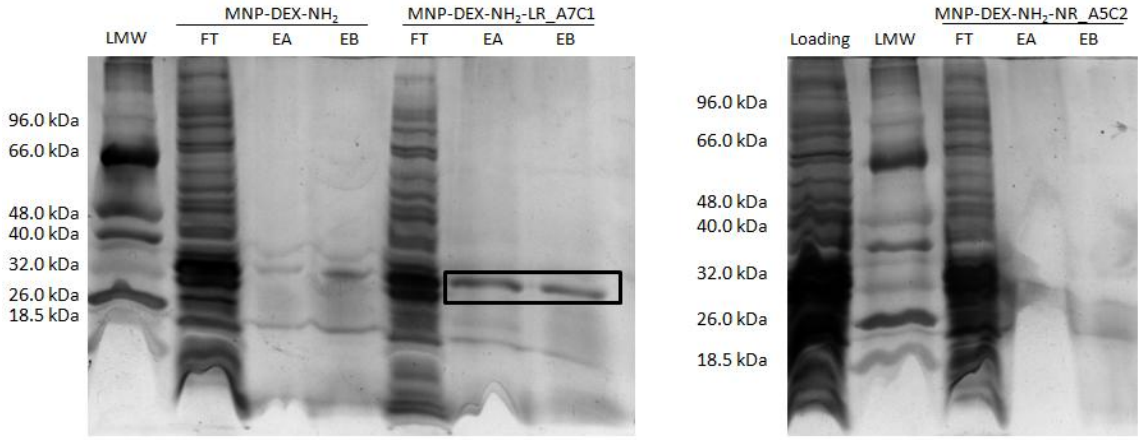


Figure 5.12 - SDS-PAGE analysis of the molecular weight marker (LMW) and flow-through (FT), first elution sample with elution condition A (EA) and condition B (EB) for all the supports tested. Gels stained by silver staining.

We can see that we have RK-GFP elution in the lanes corresponding to the negative controls (non-functionalized MNPs). The calculated purity and yield for RK-GFP in the different lanes is shown in Table 5.4.

Table 5.4 - Percentage of recovery, purity and yield of RK-GFP calculated from the results presented in Figure 5.12..

Sample	RK-GFP Recovery (%)	Purity (%)	Yield (%)
MNP_DEX_NH_2 EA	25	21	0.9
MNP_DEX_NH_2 EB	11	36	0.3
MNP_DEX_NH_2_LR-A7C1 EA	58	62	6
MNP_DEX_NH_2_LR-A7C1 EB	38	78	3
MNP_DEX_NH_2_NR-A5C2 EA	0	0	0
MNP_DEX_NH_2_NR-A5C2 EB	0	0	0

A less pure amount of RK-GFP was recovered after elution of non-functionalized support. The elution of the LR-A7C1 functionalized support resulted in a higher purity of RK-GFP, although lower than for GFP elution (see Figure 5.9). On the other hand, although elution condition A resulted in a higher yield of RK-GFP eluted, it displays a lower purity.

The low values obtained are result of using a novel solid-support, for which the conditions have not been optimized.

5.8. Static partition equilibrium studies for the determination of binding constants

The binding constants that describe the reversible interaction between an affinity ligand and a target protein in an affinity purification method are the affinity binding constant (Ka) and the maximum binding capacity of the support (Q_{max}), usually obtained from the application of a Langmuir model to data from partition equilibrium studies (Figure 5.13).

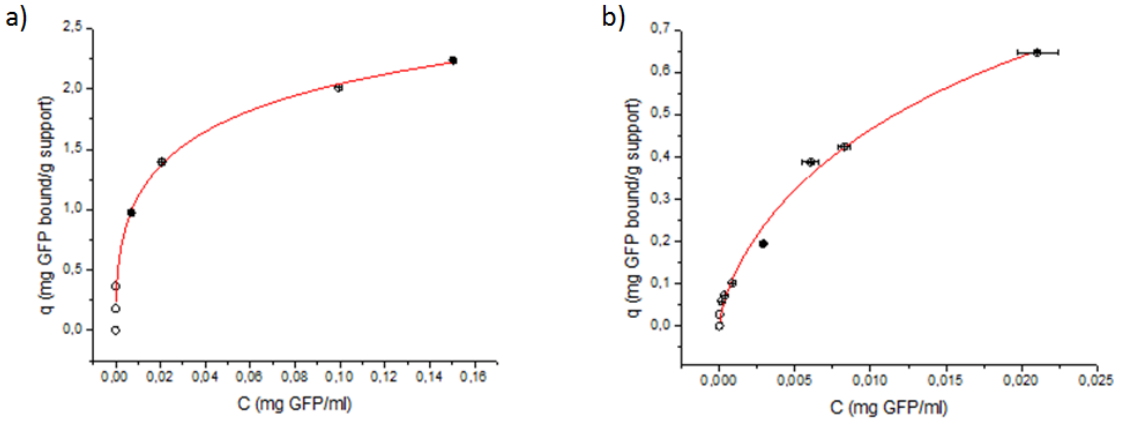


Figure 5.13 - Binding isotherms for the affinity pairs (a) GFP and LA-A4C7 and (b) RK-GFP and LR-A7C1. The experimental results have been fitted to the Langmuir model (in red).

The binding isotherm for GFP binding to LA-A4C7 (Figure 5.13a) is linear for lower concentrations of GFP and it starts to reach the maximum binding capacity. The isotherm for RK-GFP (Figure 5.13a) displays a more linear profile due to the lower concentration of RK-GFP present in the crude extracts. The Langmuir isotherm is the most popular model employed in affinity chromatography and is described by Equation 5.1 [128]:

$$q = \frac{Q_{max} \times Ka \times C}{1 + Ka \times C} \quad \text{(Equation 5.1)}$$

where q is the bound protein per mass of support (mg/g support) and C corresponds to the concentration of unbound protein in equilibrium (mg/ml). The Langmuir isotherm assumes that all binding sites are identical and that each one retains one target molecule and have uniform and independent energies of adsorption [128]. The resulting constants are summarized in Table 5.5.

Table 5.5 - Binding constants for the studied affinity systems – GFP/LA-A4C7 and RK-GFP/LR-A4C7 – with different solid supports and by theoretical study after fitting with Langmuir model.

Solid support and Affinity pair	**Ka ($\times 10^5$ M^{-1})**	**Q_{max} (mg/g)**	**R^2**
Magnetic nanoparticles			
GFP and LA-A4C7	0.83	3.99±1.61	0.99
RK-GFP and LR-A7C1	3.21	1.58±0.27	0.98
Agarose beads ([47]**)**			
GFP and LA-A4C7	2.38	1.02±6.85	0.89
RK-GFP and LR-A7C1	2.45	0.47±0.57	0.99
None (modeling study)			
GFP and LA-A4C7	1.43	-	-
RK-GFP and LR-A7C1	4.43		

The Ka values obtained were within the range for suitable purification methods: high specificity affinity purification methods display Ka within 10^3-$10^9 M^{-1}$ [94]. Ka values are within the same order of magnitude as the values obtained for agarose beads as the solid support and for the ligand in its free form (modeling study). Natural affinity systems usually display higher Ka values (e.g. glutathione/glutathione-S-transferase [129], staphylococcal protein A/IgG [130] with 10^5 and 10^8 M^{-1}, respectively). Affinity systems with biomimetic ligands usually display lower Ka (e.g. triazine-based ligands 22/8 [18] and 8/7 [19] and Ugi-based ligand [15] that bind human IgG with a Ka of 10^5, 10^4 and 10^6 M^{-1}, respectively. Compared to these literature examples, we can state that both affinity pair described has a very promising Ka value for a protein purification system, particularly RK-GFP/LR-A7C1.

Using MNP as a solid support has resulted in higher values of Q_{max} than when using agarose beads. However it is complicated to use the maximum capacity of the support as we are limited to the amount of GFP presented in the crude extracts. However, using a more diluted loading solution has shown to increase the binding of GFP to the solid support (see Figure 5.7e).

The obtained values of Q_{max} are low when compared with other affinity systems described in the literature. The magnetolipossomes of Bucak *et al* [122] or the polymer-coated MNPs of Ditsch *et al* [7] present a value of 800mg protein/g particles and 640mg/g, respectively. In general, biomimetic affinity systems also display lower Q_{max} values. The biomimetic Ugi-based ligands for the purification of IgG presented a value of 16mg/ml agarose [15]. However, we

have to have in mind that the magnetic supports tested had an unspecific behavior towards GFP binding.

5.9. Conclusions and Future Directions

Herein we have described the application of novel affinity ligands for the purification of GFP and tagged-GFP using magnetic beads.

First, we have amplified and isolated successfully pure pDNA, which was confirmed by spectroscopic and gel electrophoresis analysis. Furthermore, we were able to obtain a crude extract containing GFP or RK-GFP using *E. coli* BL21 as host. These crude extracts were successfully used for the purification assay with magnetic particles.

Magnetic particles were functionalized with the same lead and negative control affinity ligands employed in Chapter 3, namely LA-A4C7 for GFP studies and LR-A7C1 for RK-GFP studies.

When employing particles MNP_DEX_NH_2_LA-A4C7, we were able to bind and elute GFP at different experimental conditions. Based on the results, we concluded that the best conditions for GFP binding were a higher concentration of MNPs (40mg/ml) with a higher amination content (271µmol NH_2/g support), and a more diluted loading solution of protein (81µg GFP/ml). However we have observed that the magnetic support presents an unspecific binding behavior towards GFP binding, contributing for the binding of a large amount of other proteins present in the crude extract.

Still, we were able to selectively elute GFP from the LA-A4C7 functionalized-MNPs. The best elution condition for GFP elution was 0.1mM glycine-NaOH pH9 50% (v/v) ethylene glycol, as ethylene glycol contributes for the disruption of the hydrophobic interactions between the lead ligand and GFP. However the yield obtained was low (44%) despite the fact that the purity of the eluted fractions was quite high (81%).

It is complicated to perform a comparison between our magnetic-based system for GFP purification and other commercially available kits (µMACS®, from Milteniyi; GFP-TRP®, from Chromotek), due to the lack of information provided by the suppliers. However, both suppliers claim low non-specific binding of their systems, which is a clear advantage over the system

described in this work (selectivity of 12%). Both systems rely on the affinity between an anti-GFP antibody and GFP, with quite different elution conditions (µMACS® (Milteniyi) uses a denaturing elution buffer at pH6.8 and GFP-TRAP® uses an elution buffer with low pH (pH=2).

The best conditions described above (40mg MNP/ml) with a higher amination content, and a more diluted loading solution of protein) have been tested for the purification of RK-GFP. The LR-A7C1 functionalized-MNPs also showed low specificity towards RK-GFP, as lower purity and yield of RK-GFP eluted than for the GFP/LR-A4C7 system. The best elution condition for RK-GFP elution was PBS pH 7.4, 500mM arginine, in which the arginine competes for the same binding locations as the tag-residues of the protein.

Most conditions that have been chosen due to its good results when using agarose beads [47]. However different solid supports require different optimized conditions. A next possible route would be to optimize binding and elution conditions in order to increase yield and final protein purity while using MNPs as a solid support. These optimizations include different incubation and elution buffers, with different ionic strengths, pH values or others chemicals that can disrupt hydrophobic interactions or induce competition.

Partition equilibrium studies have been performed and the experimental results were fitted according to a Langmuir isotherm. For GFP/LA-A4C7 and RK-GFP/LR-A7C1 affinity systems we achieved quite promising Ka values (0.83×10^5 and $3.21\times10^5 M^{-1}$, respectively), comparable to the studies done for the same affinity pairs using agarose beads [47] and in the theoretical study (Section 4), as well as with other affinity systems described in the literature. On the other hand, Q_{max} values were quite low. These values can be correlated with the results achieved in this chapter, specifically for the low amount of RK-GFP bound to the magnetic supports. Although the Langmuir isotherm was chosen as the best fitting model it might not be the most appropriated. Other isotherm models could have resulted in better Ka and Q_{max} results for this affinity system.

6. MAGNETIC NANOPARTICLES AS ANTIMICROBIAL DEVICES

6.1. Objective

We aimed to develop novel antimicrobial devices and to assess their antimicrobial potential based on a new HTS method. For this purpose magnetic nanoparticles were functionalized with an antimicrobial peptide by two separate methods and tested against Gram-positive and Gram-negative bacteria.

The chosen antimicrobial peptide for this work was RWRWRW. The antimicrobial activity of $(RW)_x$ has been described by many authors as well as its low toxicity against eukaryotic cells [32,36,131].

6.2. Peptide immobilization onto MNPs and characterization

Two different methods for peptide immobilization onto MNPs were evaluated and compared: via sulfo-coupling and via EDC-coupling, as schematically represented in Figure 3.4 and Figure 3.5, respectively. Different biopolymers were used for coating, as dextran, PEG and a double coating of silica (Si) and dextran.

The sulfo-coupling relies on the use of sulfo-SMCC, a water-soluble amine-to-sulfhydryl crosslinker. It provides a 8.3Å spacer arm between the two reactive groups – the amine groups at the surface of the MNPs and the thiol group at the N-terminal of the Cys-PP-RWRWRW peptide (Cys-$(RW)_3$). The EDC-coupling relies on the use of two compounds: EDC and NHS. EDC is a water-soluble carbodiimide crosslinker. It allows a neutral linkage between two reactive groups – the amine groups at the surface of the MNPs and the carboxyl group of the RWRWRW peptide ($(RW)_3$). The NHS allows the formation of a more stable intermediate of the coupling reaction.

The DLS and zeta potential characterization of the resulting solid supports is shown in Figure 6.1.

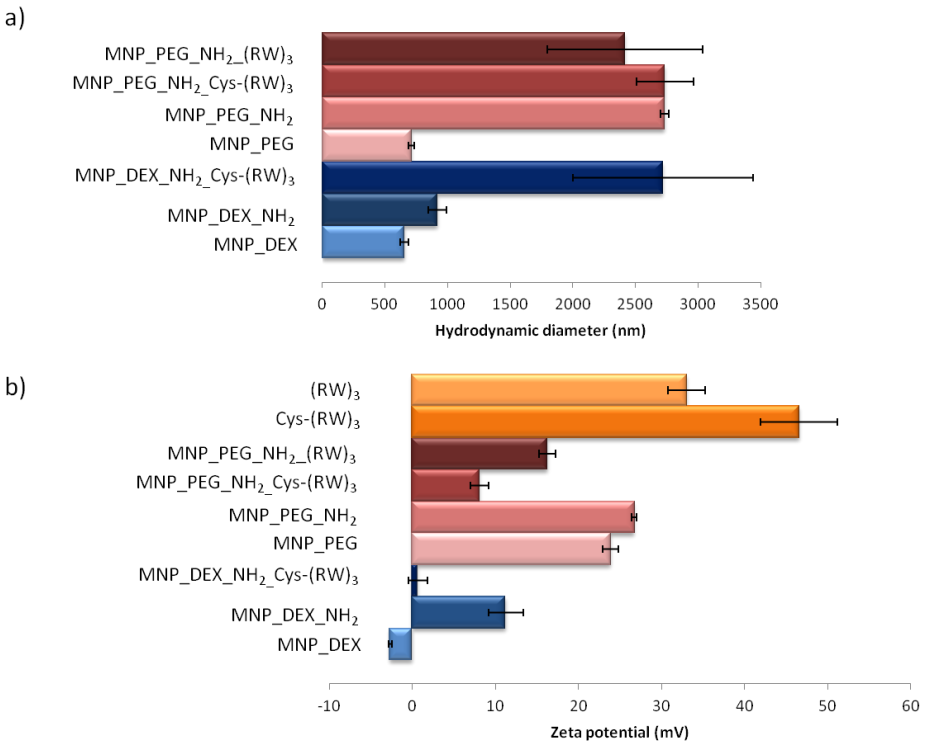


Figure 6.1 - a) Hydrodynamic diameter (nm) and (b) zeta potential (mV) results for the magnetic supports and peptides used for the antimicrobial assay (n=3).

As described in the previous section, additional functionalization contribute for the increase of the hydrodynamic diameter, as seen in Figure 6.1a. A similar diameter has been reported for the dextran- and PEG-coated MNPs after peptide tethering.

The zeta potential for the peptides $(RW)_3$ and Cys-$(RW)_3$ is positive (Figure 6.1b, in orange), which would be expected due to the cationic behavior of the peptide. The positive charges are vital for the peptide's antimicrobial mechanism [36].

On the other hand, all functionalized-solid supports present zeta potential values between -30mV and +30mV, which is a confirmation of the low stability of the supports. The main disadvantage of larger particles is the decreased specific surface area available. However the larger aggregates allow the facilitated magnetic removal of the MNPs from solution. Further observations related with the zeta potential of the functionalized-solid supports while be discussed below.

For an initial comparison of the two methods of functionalization we have compared the percentage of immobilization of the peptide onto MNP with different coatings. The percentage of immobilization is summarized in Table 6.2.

Table 6.2 - Percentage of functionalization for the different coated-MNP used as antimicrobial devices.

Biopolymer coating	Functionalization method	Peptide	% Immobilization	% Immobilization (negative control)
Dextran	Sulfo-coupling	Cys-$(RW)_3$	92±4 (n=3)	46 (n=1)
Si-dextran	Sulfo-coupling	Cys-$(RW)_3$	36 (n=1)	-
Si-PEG	Sulfo-coupling	Cys-$(RW)_3$	92 (n=1)	0 (n=1)
	EDC-coupling	$(RW)_3$	18 (n=1)	5 (n=1)
PEG	Sulfo-coupling	Cys-$(RW)_3$	99 (n=1)	50 (n=1)
	EDC-coupling	$(RW)_3$	42±28 (n=3)	4 (n=1)

Note: n represents the number of times that the strategy was tested. The results with (-) were not performed in this work.

The first strategy used was to functionalize MNP_DEX_NH_2 with Cys-$(RW)_3$, which yield a high percentage of immobilization (92%). However when these particles were tested against *E. coli* we did not observe the expected antimicrobial effect (see section 6.3). In fact, a low positive zeta potential value has been measured for the MNP_DEX_NH_2_Cys-$(RW)_3$ (Figure 6.1b) and a more positive charge should be determinant for the interaction with the negatively charged bacterial membrane [132]. On the other hand, although the percentage of immobilization was high, it appeared that the peptide was not immobilized in such a way that could facilitate its interactions with the microbial cell walls. In addition, the amount of "correctly" immobilized peptide is not known. It is possible that the cysteine residue from Cys-$(RW)_3$ is interacting directly with the iron oxide surface, which would hidden the antimicrobial group. Cysteine has been described to absorb to magnetite in a wide range of pH and ionic strengths [133]. A negative control has been performed, where the MNP_DEX_NH_2 were incubated with the peptide in the absence of the cross-linker, and the results confirm that the peptide was being attached to the magnetic core *via* the cysteine residue.

MNP_PEG_NH_2 were also functionalized with the antimicrobial peptide *via* sulfo-coupling with a high percentage of functionalization (over 90%) and did not show any antimicrobial effect against *E. coli* (data not shown). The negative control performed with these particles demonstrated that the peptide was interacting with the magnetic surface. The AMP was not available to interact with the microbe membrane and would not contribute for its disruption, which is supported by the low positive zeta potential (Figure 6.1b).

A new double coating was tested: silica/dextran- and silica/PEG-coated MNPs were functionalized with Cys-$(RW)_3$ via sulfo-coupling. We expected that the double coating would decrease the probability of the cysteine interacting with the MNP surface. The negative control

performed with the silica/PEG-coated MNPs supports this hypothesis: no peptide interacts *via* cysteine, as the double coating provides a better coverage of the magnetic core.

To overcome the described disadvantages a different functionalization strategy was used: via EDC-coupling. Poly(ethylene glycol) (PEG) is described by its anti-adhesive [44] properties and it is used as coating for anti-fouling surfaces [4,44]. PEG contributes for MNP stability when used as coating - confirmed by higher zeta potential values (Figure 6.1b). Therefore only PEG- and silica/PEG-coated MNP were tested. The negative controls performed confirm that there is a low non-specific coupling of the peptide to these particles. However, as described in Table 6.2 EDC-coupling seems a non-reproducible method for $(RW)_3$ coupling due to the characteristics of the peptide. The EDC-coupling strategy is based on the bond formation between a carboxylic group and an amine group. The amines in solution are not only the amines from the aminated support but also the amines from the arginines' side chains of the peptide, which leads to a cross-link reaction within the peptides present in solution, contributing to lower the available free peptide in solution. On the other hand, if the mesh of peptide is immobilized it may not have the correct conformation to disrupt the microbial membrane or the concentration we are expecting may not be correctly available. The EDC-coupling strategy must be further optimized to be used for the immobilization of this peptide.

However the functionalization of the PEG-coated MNPs has yield particles which have demonstrated antimicrobial activity, as shown in Figure 6.3 and Figure 6.4. It also has a more positive zeta potential value than the MNP_PEG_NH_2_Cys-$(RW)_3$ (see Figure 6.1b), which supports the antimicrobial activity results.

6.3. Testing magnetic antimicrobial devices against *E. coli* K12 and *B. subtilis* 168

Two bacteria strains have been tested against antimicrobial devices in this work: *E. coli* K12 and *B. subtilis* 168. The negatively-charged bacterial cell membrane and the outer lipopolysaccharide layer of Gram-negative are the perfect target for the electrostatic interaction with the peptide [134].

The peptide $(RW)_x$ alone has shown to display antimicrobial activity against *E. coli* [32,131] and *B. subtilis* [134]. The chain length of $(RW)_x$ peptides correlates with the antimicrobial

activity and the hexapeptide has been described to exhibit a strong antimicrobial activity [36]. To confirm these results both strains used in this work have been tested against this peptide (Figure 6.2).

A very high concentration of the peptide has been used in this work compared to the concentrations reported in the literature – a maximum of 0.08mM [131], 0.15mM [134] and 0.20mM [32].

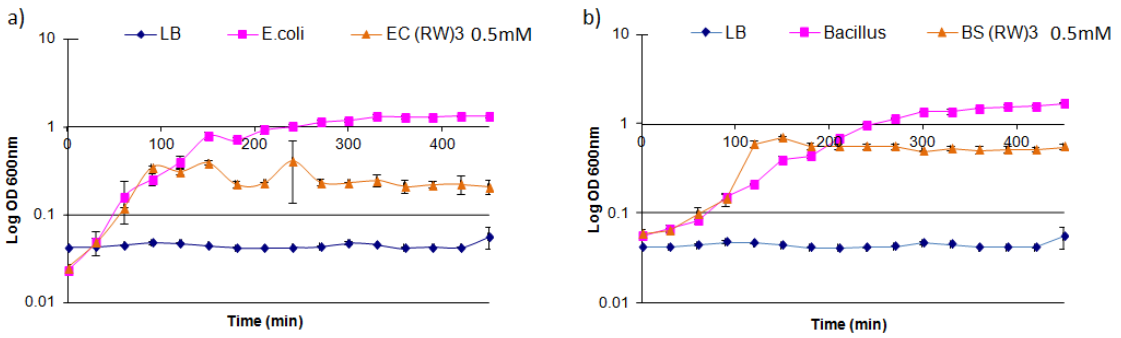


Figure 6.2 - (a) *E. coli* (EC) and (b) *B. subtilis* (BC) growth behavior in shaking flasks in the presence of $(RW)_3$ antimicrobial peptide. The antimicrobial agent was added at (a) t=60min and (b) t=90min.

Figure 6.2 clearly shows that $(RW)_3$ affects bacteria growth. The minimum inhibitory concentration (MIC) has not been determined in this work, but in a similar study [47] it was determined that 0.25mM of $(RW)_3$ was enough to inhibit the growth of *E. coli* K12. Although the action of this AMP has not been reported against *B. subtilis*, a lower concentration would be expected due to the lack of the outer layer – a lower MIC has been reported against Gram-positive *S. aureus* compared to *E. coli* and other Gram-negative bacteria [36,134].

We observe a bacteriostatic effect of the peptide for both strains – the peptide inhibits the growth of the bacteria present in solution. Although this peptide is described as having a bactericidal effect [32], the protocol to derive this observation was different, which was based on the inoculation of agar plates after 3h incubation of *E. coli* with the $(RW)_3$ peptide. This protocol may not reflect adaptation mechanisms that may occur.

Herein we tested some of the functionalized magnetic supports described above for their antimicrobial potential against Gram+ and Gram- bacteria. Two platforms have been compared for the bacterial growth: shaking flasks and 96-well titer plates.

Nowadays available methods for antimicrobial studies are based on non-practical methodologies. These include plating bacterial suspensions that have been pre-incubated with antimicrobial solutions [33,44], by disk diffusion tests [45] or cell viability tests with

fluorescent markers [33]. However these do not allow high-throughput screening (HTS) of antimicrobial peptides or minimum inhibitory concentration (MIC) determination.

To overcome these drawbacks, HTS antimicrobial assays on multi-well format have been developed. The determination of bacterial viability relies on the measurement of a fluorescence signal [41,43] or the optical density [32,44].

The comparison of two of the functionalized magnetic supports is summarized in Figure 6.3.

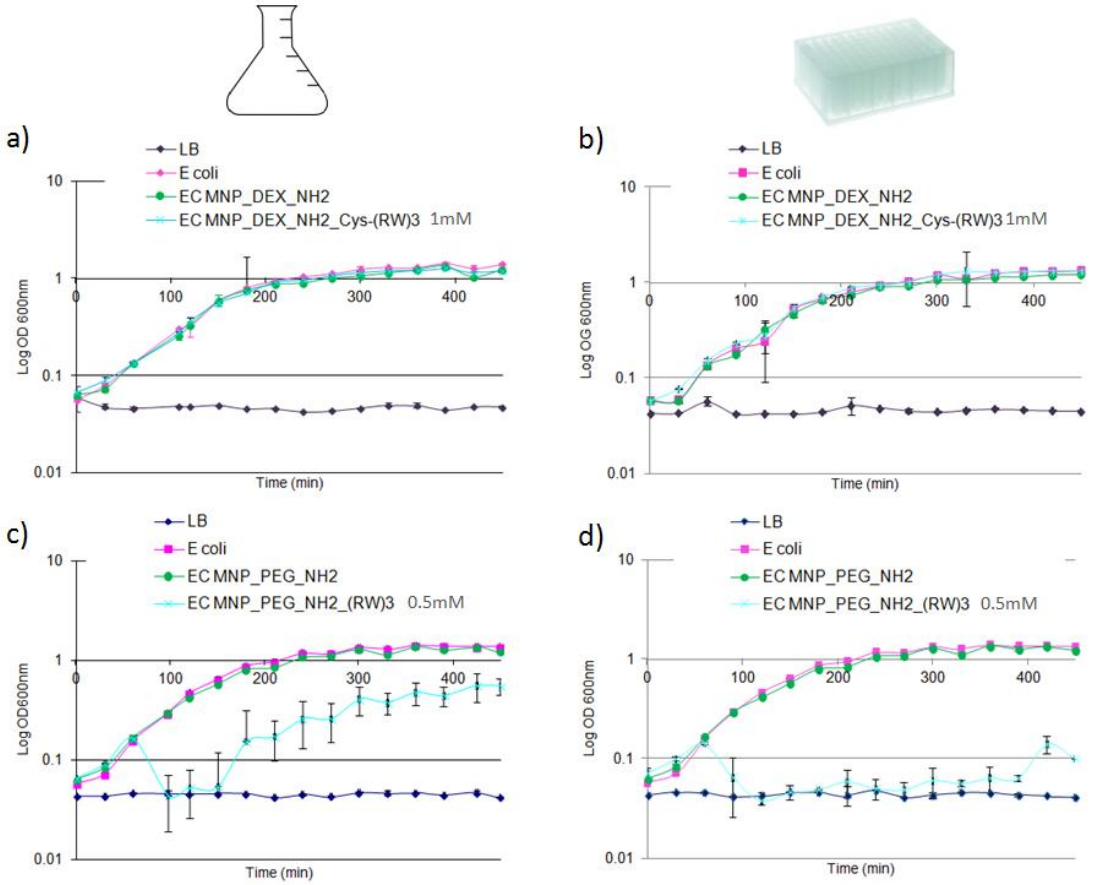


Figure 6.3 - *E. coli* (EC) growth behavior in the presence of $(RW)_3$-functionalized dextran-coated MNPs (a and b) and in the presence of $(RW)_3$-functionalized PEG-coated MNPs (c and d). Two growth supports have been compared: shaking flasks (a and c) and 96-well titer plate (b and d). The antimicrobial agent was added at t=60min.

If we compare the two methodologies (shaking flasks - Figure 6.3a and Figure 6.3c – and 96-well titer plates - Figure 6.3b and Figure 6.3d) we can see that we have obtained similar growth behaviors for most cases. This shows that we can successfully use the 96-well titer plates to obtain the same results as with shaking flasks, which allows a more efficient screening, including HTS.

Two polymers have been used to coat the MNPs displayed in Figure 6.3. Dextran and PEG are non-toxic polymers that have previously been used for MNP coating [4,135]. As expected, polymer-coated MNPs do not have any antimicrobial effect in the bacteria growth (Figure 6.3, in green).

The first support tested was dextran-coated MNPs functionalized with the Cys-$(RW)_3$ peptide via sulfo-coupling. No effect was observed on the growth of *E. coli* (Figure 6.3a and Figure 6.3b, in blue). The same behavior was observed for PEG-coated MNP functionalized by sulfo coupling (data not shown). A very high concentration has been tested (1mM), which should be enough to display antimicrobial activity [36]. As discussed in section 6.2, this was probably a consequence of the functionalization method and the peptide was not free to interact with the microbe membrane. The membrane disruption mechanism is dependent on the concentration of adsorbed peptide [136] and although we thought we were testing *E. coli* against an estimated concentration of 1mM we probably have a much lower peptide concentration at the MNP surface. On the other hand, the graft density of the peptide at the surface of the MNPs has an important role in its antimicrobial activity [41]. However this parameter is hard to control. Moreover it has been described that a lower grafting of the peptide onto the solid support may contribute for the partial adsorption of the bacteria to the particles and limit the peptide's availability [33].

Other functionalization method was tested to conjugate $(RW)_3$ to the PEG-coated MNPs with $(RW)_3$. The peptide concentration tested was 0.5mM, the same concentration that showed to display antimicrobial activity against both strains (Figure 6.2). When we added the functionalized MNPs with 0.50mM of peptide at t=60min we can see that it disturbed the microbial growth, causing a decrease of the number of cells in solution. Moreover the same behavior is observed in both shaking flasks and 96-well titer plates - Figure 6.3c and Figure 6.3d, respectively. Comparing the death rate (slope between t=60min (when the magnetic device was added) and the time when the microbial growth reaches its lowest value) we see they are in the same order of magnitude: $0.0030h^{-1}$ in the shaking flasks (t=60-90min) and $0.0017h^{-1}$ in the 96-well titer plates (t=60-120min).

In the shaking flasks, *E. coli*'s growth is inhibited for 1h30min, after which we can see an adaptation behavior and it slowly starts growing although with a 39% lower growth rate than *E. coli* alone ($0.1247h^{-1}$ *vs* $0.3199h^{-1}$, respectively). This percentage is very similar to the 36% decrease in the growth rate reported by Hou *et al* while testing 0.20mM $(RW)_3$ against *E. coli* [32]. On the other hand, a higher growth rate was reported for the $(RW)_3$-incubated *E. coli* [32]

than in this study, which can be a first indication of a higher antimicrobial potential of the $(RW)_3$-functionalized MNPs.

On the other hand, when using the 96-well titer plates as support for the microbial growth a different profile was observed. *E. coli* only started to show some growth behavior 5h after incubation. This can be an indirect result of the lower solution homogenization due to the 96-well titer plate wells' geometry – each well is narrower than a shaking flask.

The results in Figure 6.3c and Figure 6.3d were obtained for the incubation of the microbial culture with a final concentration of 0.50mM of peptide. An effective antimicrobial agent should have a high antimicrobial activity at a concentration that is non-toxic for mammalian cells [44]. $(RW)_3$ has only displayed hemolytic activity against human red blood cells at 0.745mM [134], which means that these devices would have low probability of showing toxicity against mammalian cells.

Due to the MNP_PEG_NH_2_$(RW)_3$'s antimicrobial properties, these particles have been also tested against *B. subtilis* (Figure 6.4).

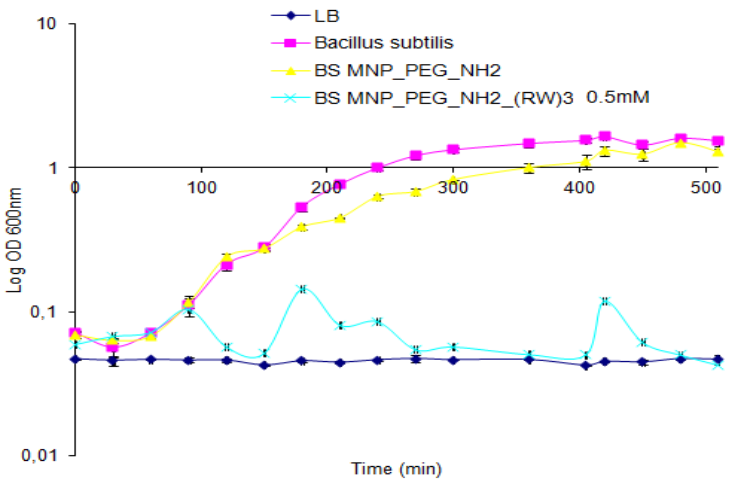


Figure 6.4 - *B. subtilis* (BS) growth behavior in 96-well titer plate in the presence of $(RW)_3$-functionalized MNPs in shaking flask. The antimicrobial agent was added at t=90min.

Once again we observed a bactericidal effect of the $(RW)_3$-functionalized MNPs. The growth behavior of *B. subtilis* is very similar to the results in Figure 6.3d for *E.coli*. *B. subtilis* does not seem to be able to display a constant growth rate and lacks the adaptation capacity against the peptide. A longer experiment with *E. coli* may have shown the same behavior.

The calculated death rate (t=90-150min) is higher than for *E. coli* in the same conditions (Figure 6.3d) - $0.0525h^{-1}$ *vs* $0.0017h^{-1}$, which would be expected due to the lack of outer layer of the Gram-positive bacteria.

Comparison with other AMP-functionalized nanodevices described in the literature (see Table 6.2) is a rather challenging task, as no author has used HTS analysis and most results are displayed as inhibition or cellular decrease percentages. On the other hand, the antimicrobial effects of $(RW)_3$ have never been studied when tethered to a surface or other solid support.

Zhang *et al* [38] functionalized magnetic iron oxide particles with bacitracin that displayed a higher antimicrobial activity than bacitracin itself. However this is not a typical behavior, as the solid support usually decreases the peptides' antimicrobial activity [33,34].

Using MNP_PEG_NH_2_$(RW)_3$ we achieved close to 100% decrease of the microbial population for both strains. Blin *et al* [33] only achieved a 75% decrease of the microbial population while using magainin I-functionalized silica magnetic particles and conclude that better results could be achieved by tuning the concentration, size and treatment time.

According to the literature, PEG seems to be a better choice as coating material. Moreover, Forbes and co-workers [44] have functionalized discs composed of various materials with antimicrobial apoE derivates and conclude that the two tested strains displayed the lowest percentage of survival when incubated with the PEG discs.

6.4. Conclusions and Future Directions

Two different methods to couple the peptide $(RW)_3$ onto MNPs have been tested - *via* sulfo-coupling and EDC-coupling – and compared according to the immobilization efficiency and the antimicrobial potential of the final device.

Although the sulfo-coupling methodology yielded higher immobilization percentages, the resulting particles did not display antimicrobial activity. It is probable that the terminal cysteine of the peptide used (Cys-$(RW)_3$) was interacting with the magnetic core of the particles, which would result in lower amount of peptide available for interaction. This hypothesis was supported by the negative controls, which were performed in the absence of the sulfo-SMCC cross-linker: a large amount of peptide would interact non-specifically with the particles. On the other hand, the AMPs used in this work present a very positive zeta potential, which is a requisite to occur the interaction between the cationic peptide and the negatively charged bacteria membrane. The low amount of peptide at the MNP surface contributed for a lower zeta potential value, which may not be positive enough to interact membrane.

Better coating options should be explored to be able to use the sulfo-coupling strategy. The coating of iron oxide particles with a silica layer previously to the polymer layer showed promising results and decreased non-specific interaction of the peptide with the particles.

To overcome the described disadvantages a different functionalization strategy was used: via EDC-coupling. Although lower immobilization percentages were achieved, the EDC-coupling methodology has resulted into a potential novel antimicrobial nanodevice with antimicrobial activity against *E.coli* and *B. subtilis*. A negative control was always performed to confirm that the antimicrobial activity was a result of the peptide and not of the non-functionalized particles. Furthermore, two growing platforms have been tested (shaking flasks *vs* 96-well titer plates). Comparable results were achieved in terms of death rate when MNP_PEG_NH_2_$(RW)_3$ was tested against *E. coli* ($0.0030h^{-1}$ and $0.0017h^{-1}$, respectively). On the other hand, a higher death rate was recorded for *B. subtilis*, which would be expected due to the lack of outer layer of the Gram-positive bacteria.

However the work described above is only a first step and there are still many hypotheses to be explored. EDC-coupling is not the most efficient functionalization method for the $(RW)_3$ peptide, due to the large amount of amines of the peptide itself. The amount of peptide that we can tether to the MNPs is variable and the method should be optimized. Other functionalization techniques can be explored, as physical (adsorption, layer-by-layer assembly) or chemical (covalent bonding) methods. Onaizi and Leong have written a good review on this topic [137].

Comparison with other AMP-functionalized nanodevices described in the literature is complicated due to the different methodologies that have been used to analyze the antimicrobial potential. However, preliminary results are quite promising for a novel antimicrobial nanodevice.

Due to their magnetic properties it should be interesting to re-use the particles above and analyze their antimicrobial potential after multiple rounds.

In this work we were not able to test the double coating particles for their antimicrobial activity. However they have shown promising results when functionalized by the sulfo-coupling methodology, as well as a higher zeta potential (+27mV) than the devices described above.

When testing novel antimicrobial molecules or devices two parameters are typically measured – the MIC and the concentration where 50% of the red blood cells are lysed when exposed to the antimicrobial molecule [37]. To determine the MIC the same protocol can be

used, as the HTS method described above would allow the screening of a high number of concentrations simultaneously.

7. CONCLUDING REMARKS

This work explored many different areas and techniques to achieve the proposed goals.

In the first chapter several modeling techniques have been used successfully to explain the molecular interaction between two affinity pairs. The methodology involved building a homology model based on the target protein sequence, relaxing the model to avoid structural crashes through a first MD, finding the most promising docking poses and finally using a more extended MD simulation to analyze relevant interactions between the protein surface and the affinity ligand.

The interaction between GFP/LA-A4C7 is mainly hydrophobic during a very significant percentage of the simulation time (over 90%). On the other hand, the interaction between RK-GFP/LR-A7C1 is due to hydrogen bonding between tag residues and oxygen atoms of the ligand. However fewer docking results supported this hypothesis.

Several methodological limitations do not allow us to conclude definitely about these interactions, as inefficient sampling, system simplification and accuracy of the force field. On the other hand, the criteria used to choose the docking poses may not be the most accurate. Therefore, this study can only be used as a first approach to explain the experimental results in [47].

Furthermore, a theoretical second generation for an Ugi affinity ligand for the purification of GFP has been developed with some potential better affinity ligands.

In the second chapter we amplified and isolated pDNA containing the gene that expressed the protein of interest and used *E. coli* BL21 to produce crude extracts containing GFP or RK-GFP. Afterwards, the affinity pairs described in [47] and studied in the 1st chapter were tethered onto MNPs and their affinity potential was accessed.

We were able to bind and elute GFP using MNP_DEX_NH_2_LA-A4C7, while testing different experimental conditions. The best conditions for GFP binding include higher MNP concentration with higher amination content and lower loading concentrations. However the solid support presented unspecific binding towards GFP. The best elution condition for GFP contains ethylene glycol, which contributes for the disruption of the hydrophobic interaction between GFP and LA-A4C7. However we could only obtain a yield of 44% and a final purity of 86% in the elution fraction.

On the other hand, the best elution condition for RK-GFP is based on a competition system with arginines. The LR-A7C1 functionalized-MNPs showed low specificity towards RK-GFP, as well as lower purity and yield of RK-GFP eluted than for the GFP/LR-A4C7 system

For GFP/LA-A4C7 and RK-GFP/LR-A7C1 affinity systems we achieved quite promising Ka values for an affinity-based protein purification system. However several experimental conditions have to be previously optimized for this new solid support.

In the third chapter, two different peptide coupling methods onto MNPs have been tested - *via* sulfo-coupling and EDC-coupling

Although the sulfo-coupling methodology yielded higher immobilization percentages, the resulting particles did not display antimicrobial activity due to the interaction between the terminal cysteine of Cys-$(RW)_3$ and the magnetic core and low positive zeta potential. Better coating options (e.g. silica/PEG) should be explored to be able to use this immobilization strategy.

To overcome the described disadvantages a different functionalization strategy was used: via EDC-coupling, which resulted in a novel antimicrobial nanodevice with antimicrobial activity against Gram-negative and Gram-positive bacteria. A higher death rate was recorded for *B. subtilis* than for *E. coli*, which would be expected due to the lack of outer layer of the Gram-positive bacteria. However the EDC-coupling methodology is not the most efficient functionalization method for the amine-rich $(RW)_3$ peptide

We also proved the reproducibility of the results when using classical growing supports (shaking flasks) and a HTS method (96-well titer plates), with comparable results. The HTS method allows a more efficient and rapid screening method.

There are still many hypotheses to be explored, including a more efficient functionalization method, reusability of the particles, MIC determination of the novel nanodevice and different coating options.

8. BIBLIOGRAPHY

1 Safarik, I. and Safarikova, M. (2009) Magnetic nano- and microparticles in biotechnology. *Chem. Pap.* 63, 497–505

2 Dias, A.M.G.C. *et al.* (2011) A biotechnological perspective on the application of iron oxide magnetic colloids modified with polysaccharides. *Biotechnol. Adv.* 29, 142–155

3 Laurent, S. *et al.* (2008) Magnetic iron oxide nanoparticles: synthesis, stabilization, vectorization, physicochemical characterizations, and biological applications. *Chem. Rev.* 108, 2064–2110

4 Boyer, C. *et al.* (2010) The design and utility of polymer-stabilized iron-oxide nanoparticles for nanomedicine applications. *NPG Asia Mater.* 2, 23–30

5 Wu, W. *et al.* (2008) Magnetic iron oxide nanoparticles: synthesis and surface functionalization strategies. *Nanoscale Res. Lett.* 3, 397–415

6 Lu, A.-H. *et al.* (2007) Magnetic nanoparticles: synthesis, protection, functionalization, and application. *Angew. Chem. Int. Ed. Engl.* 46, 1222–1244

7 Ditsch, A. *et al.* (2006) Ion-exchange purification of proteins using magnetic nanoclusters. *Biotechnol. Prog.* 22, 1153–1162

8 Horák, D. *et al.* (2007) Preparation and properties of magnetic nano- and microsized particles for biological and environmental separations. *J. Sep. Sci.* 30, 1751–1772

9 Santana, S.D. *et al.* (2012) Immobilization of enterokinase on magnetic supports for the cleavage of fusion proteins. *J. Biotechnol.* 161, 378–382

10 Santana, S. *et al.* (2012) Dextran-coated magnetic supports modeified with a biomimetic ligand for IgG purification. *ACS Appl. Mater. Interfaces* 4, 5907–5914

11 Safarik, I. and Safarikova, M. (2004) Magnetic techniques for the isolation and purification of proteins and peptides. *Biomagn. Res. Technol.* 2, 7

12 Martínez, J.L. *et al.* (2012) Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation. *Curr. Opin. Biotechnol.* 23, 965–971

13 Fong, B.A. *et al.* (2010) The potential role of self-cleaving purification tags in commercial-scale processes. *Trends Biotechnol.* 28, 272–279

14 Roque, A.C.A. and Lowe, C.R. (2006) Advances and applications of de novo designed affinity ligands in proteomics. *Biotechnol. Adv.* 24, 17–26

15 Haigh, J.M. *et al.* (2009) Affinity ligands for immunoglobulins based on the multicomponent Ugi reaction. *J. Chromatogr. B* 877, 1440–1452

16 Branco, R.J. *et al.* (2012) Understanding the molecular recognition between antibody fragments and protein A biomimetic ligand. *J. Chromatogr. A* 1244, 106–15

17 Arnau, J. *et al.* (2011) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr. Purif.* 48, 1–13

18 Teng, S.F. *et al.* (2000) Affinity chromatography on immobilized "' biomimetic '" ligands Synthesis , immobilization and chromatographic assessment of an immunoglobulin G-binding ligand. *J. Chromatogr. B* 740, 1–15

19 Roque, A.C.A. *et al.* (2005) An artificial protein L for the purification of immunoglobulins and Fab fragments by affinity chromatography. *J. Chromatogr. A* 1064, 157–167

20 Lichty, J.J. *et al.* (2005) Comparison of affinity tags for protein purification. *Protein Expr. Purif.* 41, 98–105

21 Li, Y. (2011) Self-cleaving fusion tags for recombinant protein production. *Biotechnol. Lett.* 33, 869–881

22 Batalha, I.L. *et al.* (2010) Gum Arabic coated magnetic nanoparticles with affinity ligands specific for antibodies. *J. Mol. Recognit.* 23, 462–471

23 Betz, N. (2004) Purifying His-tagged proteins from insect and mammalian cells. *Cell notes* 9, 6–9

24 Safarík, I. and Safaríková, M. (1999) Use of magnetic techniques for the isolation of cells. *J. Chromatogr. B, Biomed. Sci. Appl.* 722, 33–53

25 Lupo, L.D. *et al.* (2012) Application of µMACS ™ Streptavidin MicroBeads for the analysis of HIV-1 directly from patient plasma. *MACS&more* 8, 16–19

26 Odabaşı, M. and Denizli, A. (2004) Cibacron Blue F3GA incorporated magnetic poly(2-hydroxyethyl methacrylate) beads for lysozyme adsorption. *J. Appl. Polym. Sci.* 93, 719–725

27 Odabaşı, M. and Denizli, A. (2004) Cibacron Blue F3GA-attached magnetic poly(2-hydroxyethyl methacrylate) beads for human serum albumin adsorption. *Polym. Int.* 53, 332–338

28 O'Brien, S.M. *et al.* (1996) Non-porous magnetic chelator supports for protein recovery by immobilised metal affinity adsorption. *J. Biotechnol.* 50, 13–25

29 Girault, S. *et al.* (1996) Coupling of MALDI-TOF mass analysis to the separation of biotinylated peptides by magnetic streptavidin beads. *Anal. Chem.* 68, 2122–6

30 Prioult, G. *et al.* (2000) Rapid purification of nisin Z using specific monoclonal antibody-coated magnetic beads. *Int. Dairy J.* 10, 627–633

31 Khng, H. *et al.* (1998) The synthesis of sub-micron magnetic particles and their use for preparative purification of proteins. *Biotechnol. Bioeng.* 60, 419–424

32 Hou, S. *et al.* (2010) Effects of Trp- and Arg-containing antimicrobial-peptide structure on inhibition of Escherichia coli planktonic growth and biofilm formation. *Appl. Environ. Microbiol.* 76, 1967–1974

33 Blin, T. *et al.* (2011) Bactericidal microparticles decorated by an antimicrobial peptide for the easy disinfection of sensitive aqueous solutions. *Biomacromolecules* 12, 1259–1264

34 Chen, G. *et al.* (2009) Nanolayer biofilm coated on magnetic nanoparticles by using a dielectric barrier discharge glow plasma fluidized bed for immobilizing an antimicrobial peptide. *Nanotechnology* 20, 1–6

35 Li, P. *et al.* (2012) Antimicrobial macromolecules: synthesis methods and future applications. *RSC Adv.* 2, 4031–4044

36 Liu, Z. *et al.* (2007) Length effects in antimicrobial peptides of the (RW) n series. *Antimicrob. agents Chemother.* 51, 597–603

37 Engler, A.C. *et al.* (2012) Emerging trends in macromolecular antimicrobials to fight multi-drug-resistant infections. *Nano Today* 7, 201–222

38 Zhang, W. *et al.* (2012) Bacitracin-conjugated superparamagnetic iron oxide nanoparticles: synthesis, characterization and antibacterial activity. *Chemphyschem* 13, 3388–3396

39 Liu, L. *et al.* (2013) The potent antimicrobial properties of cell penetrating peptide-conjugated silver nanoparticles with excellent selectivity for Gram-positive bacteria over erythrocytes. *Nanoscale* 5, 3834–3840

40 Qi, X. *et al.* (2011) Covalent immobilization of nisin on multi-wall carbon nanotubes: superior antimicrobial and anti-film properties. *Nanoscale* 3, 1874–1880

41 Gao, G. *et al.* (2011) Antibacterial surfaces based on polymer brushes: investigation on the influence of brush properties on antimicrobial peptide immobilization and antimicrobial activity. *Biomacromolecules* 12, 3715–3727

42 Héquet, A. *et al.* (2011) Optimized grafting of antimicrobial peptides on stainless steel surface and biofilm resistance tests. *Colloids surfaces B Biointerfaces* 84, 301–930

43 Gao, G. *et al.* (2011) The biocompatibility and biofilm resistance of implant coatings based on hydrophilic polymer brushes conjugated with antimicrobial peptides. *Biomaterials* 32, 3899–3909

44 Forbes, S. *et al.* (2013) Comparative surface antimicrobial properties of synthetic biocides and novel human apolipoprotein E derived antimicrobial peptides. *Biomaterials* 34, 5453–5464

45 Kazemzadeh-Narbat, M. and Lai, B. (2013) Multilayered coating on titanium for controlled release of antimicrobial peptides for the prevention of implant-associated infections. *Biomaterials* 34, 5969–5977

46 Chudasama, B. *et al.* (2010) Enhanced antibacterial activity of bifunctional Fe3O4-Ag core-shell nanostructures. *Nano Res.* 2, 955–965

47 Pina, A.S. (2013) , Novel affinity pairs “tag-receptor” for the purification of recombinant proteins.

48 Goujon, M. *et al.* (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* 38, W695–W699

49 Bordoli, L. *et al.* (2008) Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* 4, 1–13

50 Šali, A. *et al.* (2008) *MODELLER. A Program for Protein Structure Modeling*,

51 Rohl, C.A. *et al.* (2004) Protein structure prediction using Rosetta. *Methods Enzymol.* 383, 66–93

52 McGuffin, L.J. *et al.* (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16, 404–405

53 Söding, J. *et al.* (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–248

54 Fan, F. *et al.* (2010) Expression and purification of two different antimicrobial peptides, PR-39 and Protegrin-1 in Escherichia coli. *Protein* 73, 147–151

55 (2011) , SWISS MODEL. [Online]. Available: http://swissmodel.expasy.org/

56 Benkert, P. *et al.* (2008) QMEAN: A comprehensive scoring function for model quality assessment. *Proteins Struct. Funct. Bioinforma.* 71, 261–77

57 Benkert, P. *et al.* (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27, 343–350

58 Holm, L. and Ander, C. (1992) Evaluation of protein models by atomic solvation preference. *J. Mol. Biol.* 225, 93–105

59 Biotecnology, I. of (2012) , SolvX Server. [Online]. Available: http://ekhidna.biocenter.helsinki.fi/solvx/start. [Accessed: 08-Jan-2013]

60 Rauscher, S. and Pomès, R. (2010) Molecular simulations of protein disorder. *Biochem. cell Biol.* 290, 269–290

61 Eisenberg, D. *et al.* (1997) VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol.* 277, 396–404

62 Hess, B. *et al.* (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. theory compuatation* 4, 435–447

63 Berendsen, H.J.C. *et al.* (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690

64 Bussi, G. *et al.* (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101–7

65 Hess, B. *et al.* (1997) LINCS: a linear constraint solver for molecular simulations. *J. compuatational Chem.* 18, 1463–1472

66 Schüttelkopf, A.W. and van Aalten, D.M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Biol. Crystallogr.* 60, 1355–1363

67 Rozas, I. (2007) On the nature of hydrogen bonds: an overview on computational studies and a word about patterns. *Phys. Chem. Chem. Phys.* 9, 2782–2790

68 Bahar, I. and Jernigan, R.L. (1997) Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.* 266, 195–214

69 Morris, G.M. *et al.* (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662

70 Hetényi, C. and van der Spoel, D. (2002) Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci.* 11, 1729–1737

71 Ritchie, D.W. (2008) Recent progress and future directions in protein-protein docking. *Curr. protein Pept. Sci.* 9, 1–15

72 Roda, A. (2010) Discovery and development of the green fluorescent protein, GFP: the 2008 Nobel Prize. *Anal. Bioanal. Chem.* 396, 1619–1622

73 Iafolla, M. a J. *et al.* (2008) Dark proteins: effect of inclusion body formation on quantification of protein expression. *Proteins* 72, 1233–42

74 Craggs, T.D. (2009) Green fluorescent protein: structure, folding and chromophore maturation. *Chem. Soc. Rev.* 38, 2865–2875

75 Remington, S.J. (2011) Green fluorescent protein : A perspective. *Protein Sci.* 20, 1509–19

76 Shaner, N.C. *et al.* (2007) Advances in fluorescent protein technology. *J. Cell Sci.* 120, 4247–4260

77 Ong, W.J. *et al.* (2011) Function and structure of GFP-like proteins in the protein data bank. *Mol. Biosyst.* 7, 984–992

78 Stavrakoudis, A. *et al.* (2011) Homology modeling and molecular dynamics simulations of MUC1-9/H-2K(b) complex suggest novel binding interactions. *J. Mol. Model.* 17, 1817–29

79 Freed, A.S. *et al.* (2011) Molecular simulations of multimodal ligand-protein binding: elucidation of binding sites and correlation with experiments. *J. Phys. Chem. B* 115, 13320–13327

80 Misura, K.M. *et al.* (2006) Physically realistic homology models built with ROSETTA can be more accurate than their templates. *Proc. Natl. Acad. Sci. united states Am.* 103, 5361–5666

81 Fiser, A. and Šali, A. (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* 374, 461–491

82 Wallner, B. and Elofsson, A. (2009) All are not equal: a benchmark of different homology modeling programs. *Protein Sci.* 14, 1315–1327

83 Wallner, B. and Elofsson, A. (2005) All are not equal: a benchmark of different homology modeling programs. *Protein Sci.* 14, 1315–1327

84 RosettaCommons Rosetta 3.4 User Guide. [Online]. Available: http://www.rosettacommons.org/manuals/archive/rosetta3.4_user_guide/. [Accessed: 26-Sep-2012]

85 Roy, A. *et al.* (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5, 725–738

86 Allen, M.P. (2004) Introduction to Molecular Dynamics Simulation. In *Computational Soft Matter: From Synthetic Polymers to Proteins* 23pp. 1–28

87 Jensen, F. (2001) *Introduction to Computational Chemistry*,

88 Durrant, J.D. and McCammon, J.A. (2011) Molecular dynamics simulations and drug discovery. *BMC Biol.* 9, 71

89 Brooks, B. *et al.* (2010) CHARMM: the biomolecular simulation package. *J. Comput. cgemistry* 30, 1545–1614

90 Case, D.A. *et al.* (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688

91 Oostenbrink, C. *et al.* (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* 25, 1656–1676

92 Daura, X. *et al.* (1998) Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J. Comput. Chem.* 19, 535–547

93 Jochens, H. *et al.* (2011) Protein engineering of α/β-hydrolase fold enzymes. *A Eur. J. Chem. Biol.* 12, 1508–1517

94 Gray, J.J. (2006) High-resolution protein-protein docking. *Curr. Opin. Struct. Biol.* 16, 183–193

95 Huey, R. *et al.* (2007) A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* 28, 1145–1152

96 Bamborough, P. and Cohen, F.E. (1996) Modeling protein-ligand complexes. *Curr. Opin. Struct. Biol.* 6, 236–241

97 Hartmann, C. *et al.* (2009) Docking and scoring with alternative side-chain conformations. *Proteins Struct. Funct. Bioinforma.* 74, 712–726

98 Morris, G. and Huey, R. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. cgemistry* 30, 2785–2791

99 Gutteridge, A. and Thornton, J. (2005) Conformational changes observed in enzyme crystal structures upon substrate binding. *J. Mol. Biol.* 346, 21–28

100 Markley, J.L. *et al.* (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J. Mol. Biol.* 280, 933–52

101 Kirchhofer, A. *et al.* (2010) Modulation of protein properties in living cells using nanobodies. *Nat. Struct. Mol. Biol.* 17, 133–138

102 Cetinkaya, M. *et al.* (2006) How do insertions affect green fluorescent protein? *Chem. Phys. Lett.* 419, 48–54

103 Baird, G.S. *et al.* (1999) Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci. U. S. A.* 96, 11241–11246

104 Dedecker, P. *et al.* (2013) Fluorescent proteins: shine on, you crazy diamond. *J. Am. Chem. Soc.* 135, 2387–402

105 Moreau, M.J.J. *et al.* (2010) Quantitative determination of protein stability and ligand binding using a green fluorescent protein reporter system. *Mol. Biosyst.* 6, 1285–92

106 Abedi, M.R. *et al.* (1998) Green fluorescent protein as a scaffold for intracellular presentation of peptides. *Nucleic Acids Res.* 26, 623–630

107 Drew, D. *et al.* (2011) A scalable , GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci.* 14, 2011–2017

108 Sokalingam, S. *et al.* (2012) A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS One* 7, 1–12

109 Stepanenko, O. V *et al.* (2004) Comparative Studies on the Structure and Stability of Fluorescent Proteins EGFP, zFP506, mRFP1, “dimer2”, and DsRed1. *Biochemistry* 43, 14913–14923

110 Huang, J. *et al.* (2007) Stable intermediate states and high energy barriers in the unfolding of GFP. *J. Mol. Biol.* 370, 356–71

111 Khan, F. *et al.* (2006) 19F NMR studies of the native and denatured states of green fluorescent protein. *J. Am. Chem. Soc.* 128, 10729–10737

112 Samarkina, O.N. *et al.* (2009) Universal and rapid method for purification of GFP-like proteins by the ethanol extraction. *Protein Expr. Purif.* 65, 108–113

113 Gu, H. *et al.* (2006) Biofunctional magnetic nanoparticles for protein separation and pathogen detection. *Chem. Commun. (Camb).*

114 Fukuda, H. *et al.* (2000) Folding of green fluorescent protein and the cycle3 mutant. *Biochemistry* 39, 12025–12032

115 Einhauer, A. and Jungbauer, A. (2001) The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods* 49, 455–65

116 Cass, B. *et al.* (2005) Purification of recombinant proteins from mammalian cell culture using a generic double-affinity chromatography scheme. *Protein Expr. Purif.* 40, 77–85

117 Sumandea, C.A. *et al.* (2011) Cardiac troponin T, a sarcomeric AKAP, tethers protein kinase A at the myofilaments. *J. Biol. Chem.* 286, 530–541

118 Raines, R. *et al.* (2000) The S-Tag fusion system for protein purification. *Methods Enzymol.* 326, 362–376

119 Ge, X. *et al.* (2005) Self-cleavable stimulus responsive tags for protein purification without chromatography. *J. Am. Chem. Soc.* 127, 11228–9

120 Hashimoto, M. *et al.* (2000) Expression and characterization of the chitin-binding domain of chitinase A1 from Bacillus circulans WL-12. *J. Bacteriol.* 182, 3045–3054

121 Smyth, D.R. *et al.* (2003) Crystal structures of fusion proteins with large-affinity tags. *Protein Sci.* 12, 1313–1322

122 Bucak, S. *et al.* (2003) Protein separations using colloidal magnetic nanoparticles. *Biotechnol. Prog.* 19, 477–484

123 Urban, C. and Schurtenberger, P. (1998) Characterization of turbid colloidal suspensions using light scattering techniques combined with cross-correlation methods. *J. Colloid Interface Sci.* 207, 150–158

124 Sharma, M. and Yashonath, S. (2007) Size dependence of solute diffusivity and Stokes-Einstein relationship : effect of van der Waals interaction. *Diffus. Fundam.* 7, 1–15

125 Wang, N. *et al.* (2013) Influence of metal oxide nanoparticles concentration on their zeta potential. *J. Colloid Interface Sci.* 407, 22–28

126 Ochoa, J. (1978) Hydrophobic (interaction) chromatography. *Biochimie* 60, 1–15

127 Peternel, S. *et al.* (2009) Bacterial growth conditions affect quality of GFP Expressed inside inclusion bodies. *Acta Chim. Slov.* 56, 860–867

128 Limousin, G. *et al.* (2007) Sorption isotherms : A review on physical bases , modeling and measurement. *Appl. geochemistry* 22, 249–275

129 Schramms, V.L. *et al.* (1984) Kinetic Studies and active site-binding properties of glutathione S- transferase using spin-labeled glutathione , a product analogue. *J. Biol. Chem.* 259, 714–722

130 Nilsson, B. *et al.* (1985) Immobilization and purification of enzymes with staphylococcal protein A gene fusion vectors. *EMBO J.* 4, 1075–1080

131 Chen, X. *et al.* (2011) Control of bacterial persister cells by Trp/Arg-containing antimicrobial peptides. *Appl. Environ. Microbiol.* 77, 4878–4885

132 Costa, F. *et al.* (2011) Covalent immobilization of antimicrobial peptides (AMPs) onto biomaterial surfaces. *Acta Biomater.* 7, 1431–1440

133 Vieira, A.P. *et al.* (2011) Adsorption of cysteine on hematite, magnetite and ferrihydrite: FT-IR, Mössbauer, EPR spectroscopy and X-ray diffractometry studies. *Amino Acids* 40, 205–214

134 Strøm, M.B. *et al.* (2002) Antimicrobial activity of short arginine- and tryptophan-rich peptides. *J. Pept. Sci.* 8, 431–437

135 Hradil, J. *et al.* (2007) Dextran-modified iron oxide nanoparticles. *China Particuology* 5, 162–168

136 Strömstedt, A. a *et al.* (2009) Oligotryptophan-tagged antimicrobial peptides and the role of the cationic sequence. *Biochim. Biophys. Acta* 1788, 1916–1923

137 Onaizi, S.A. and Leong, S.S. (2011) Tethering antimicrobial peptides: current status and potential challenges. *Biotechnol. Adv.* 29, 67–74