



Ana Sofia Fidalgo Pombo Mendes Pina

Mestre em Biotecnologia

***Novel affinity pairs "tag-receptor" for
the purification of fusion proteins***

Dissertação para obtenção do Grau de Doutor em
Biotecnologia, ramo Biotecnologia

Orientador: Prof. Doutora Ana Cecília Afonso Roque
Co-orientador: Prof. Doutor Christopher Robin Lowe



Março 2013

NOVEL AFFINITY PAIRS "TAG-RECEPTOR" FOR THE PURIFICATION OF FUSION PROTEINS

"Copyright"

Ana Sofia Fidalgo Pombo Mendes Pina
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

ACKNOWLEDGEMENTS

The PhD has always been a challenge during these four years allowing me the opportunity to work in different areas and therefore to have a broader view of science. However, the performance of this PhD was not possible without the support and encouragement of several people.

First of all, I would like to sincerely acknowledge my supervisors Prof. Cecília Roque and Prof. Chris Lowe, that gave me the opportunity to work on this project and together we were able to conduct a successful work. Dear Prof. Cecilia, all these achievements were not possible without your support, confidence, exigency, constructive criticism and friendship, and mostly helped me to be a more mature person at scientific level. Dear Prof. Chris Lowe, I am very thankful to your pertinent suggestions that gave me the motivation to improve the work developed during my PhD, and also for your support and trust. For all these reasons, I do not have enough words to express my gratitude.

I also would like to thank to Biomolecular Engineering Group: Íris Batalha, Telma Barroso, Margarida Dias, Susana Palma, Vijaykumar Dhadge, Dr Abid Hussain, Dr Ricardo Branco and most recently Henrique Carvalho, Cláudia Fernandes and Bianca Gonçalves! All members contributed to the goals achievement during my PhD in a friendly and fun environment. Íris Batalha, Telma Barroso and Dr. Abid Hussain were with me since the very beginning, many thanks for all your support and friendship! Dr. Abid Hussain, I am very grateful for all the help regarding the affinity chromatography technique and your patience to introduce me all the methods for the synthesis of affinity ligands. Íris Batalha, my great friend with an “eagle eye”, thank you for your critical personality that is unbelievable, and obviously for your companionship! Telma Barroso, this endurance test, that is the PhD, became easier with you! More than a colleague, you are a really good friend, always there for me! A special thanks to dear Margarida Dias who helped to improve this work, for the helpful discussions, friendship and unique personality. Thanks also to Sara Santana for the unforgettable brainstorming. Susana Palma and Vijay, you were a pleasant surprise! To all members from the past and present, Many Thanks!

To all my lab colleagues at the Institute of Biotechnology in Cambridge, my sincere acknowledgments specially to Dr Graziella El Khoury, Dr Isik Ustok, Dr Rhian Grainger Jacky Kourieh, Nan Li, Chen Chen, Srishti Gupta and Anil Bagha. My dear “Doctors Graziella”, you were tireless with your help on studies with the automated liquid chromatography ÄKTA avant 25 equipment. Thank you for your motivation and trust! Dear Isik, thank you for your interest and availability in helping me with production of GFP fusion proteins for different applications of my PhD.

I would like to express my gratitude to Prof Alice Pereira and Dr Márcia Guilherme to introduce me to molecular biology and protein expression methods as well as their availability to reply and clarify all my doubts. A special thanks to Molecular Biophysics lab in particular to Dr Filipe Folgosa, Sara Matias and Daniela Penas for their helpful discussions. I also like to thank Dr Américo Duarte and Ana Teresa Lopes for their availability and assistance.

I also like to thank to Dr Pedro Vidinha and Diana Garcia for providing *E.coli* K12 and helping me on the antimicrobial assays. Dr Paulo Lemos, a sincere acknowledgement for helping on the Fluorescence Microscopy and Image J.

A special thanks to “Gang do Tupperware” for provide me relaxed and amusing moments every day during lunch time for this last four years. All these moments helped me on my motivation to carry out my PhD thesis. Leonor Morgado and Tânia Carvalho, I am really thankful for your support and friendship during my PhD, especially in the last moments of panic! Apart from REQUIMTE colleagues, you are more than good friends for me. A special thanks to Inês Grilo for the helpfull discussions on the molecular biology experiments and obviously for the friendship,

I would like to acknowledge D. Maria José Carapinha, Isabel Rodrigues and Cara Bootman for their assistance on solving bureaucratic work and to D. Maria da Palma Afonso for providing clean tools for our research.

To all my friends and “Gang do Regabofo”, many thanks for always being there for me (even if we are far away on the map, we are close by heart), for your support and friendship! My life is much richer with you around me!

To all my family, especially to my parents, I am very grateful to have you supporting me. My dear grandparents, I am very thankful for your interest in understanding the science more specifically what I have been doing this last four years. Mum and Dad, you are the true inspiration in my life! I am thankful to you for the person that I am today with all the values and principles that you taught me during my entire life and I will never have enough words to thank you and I love you very much!

My dear Zé, you know what you mean for me! Thank you for love, patience, encouraging words and for always believes in me! My life is much better with you by my side!

Lastly, I would like to thank to Fundação para a Ciência e Tecnologia (FCT, Portugal) for the financial support through the PhD fellowship SFRH/BD/48804/2008 and the project PTDC/BIO/65383/2006 assigned to Prof. Cecília Roque and also to Associate Laboratory REQUIMTE (Pest-C/EQB/LA0006/2011).

ABSTRACT

This work aimed at the development of novel affinity pairs “tag-receptor” as generic platforms for the purification of recombinant proteins. Five different hexapeptides with the sequences RWRWRW, WFWFWF, NWNWNW, RKRKRK and NNNNNN were selected as affinity tags to be fused to Green Fluorescent Protein (GFP). The respective DNA fragments were designed, cloned in pET21c expression vector and expressed in *Escherichia coli* host as soluble proteins except for the WFWFWF tagged GFP. A solid phase combinatorial library based on the Ugi reaction was synthesized with 64 affinity ligands displaying complementarity functionalities towards the designed tags. The library was screened by microscale affinity chromatography in a 96-well format for binding to GFP and GFP-tagged proteins, leading to the selection of four putative lead ligands A4C8, A3C4, A7C1, and A4C7 for WFWFWF, NWNWNW and RKRKRK fusion proteins and non-tagged GFP, respectively. The affinity pair WFWFWF tagged GFP – A4C8 was explored for matrix assisted refolding. The affinity pair NWNWNW tagged GFP-A3C4 revealed extremely low protein recovery. The affinity pair RKRKRK tagged GFP-A7C1 emerged as a promising pair, as the ligand yielded 30% of protein recovery with a purity of 50% by using arginine as competitor agent. Additionally, the pair displays a K_a of $2.45 \times 10^5 \text{ M}^{-1}$ and a Q_{\max} of 0.5 mg GFP-tagged per gram of moist gel. A novel strategy for the purification of non-tagged GFP or GFP fusion proteins was also accomplished with ligand A4C7 that allowed a selective recovery of GFP with 92% yield and 94% purity. Moreover, the pair displayed a suitable K_a of $2.3 \times 10^5 \text{ M}^{-1}$ and a Q_{\max} of 1 mg of GFP per gram of moist gel.

KEYWORDS: Affinity tag, Fusion protein, Synthetic ligands, Affinity Chromatography, GFP

RESUMO

O objetivo desta tese consistiu no desenvolvimento de pares de afinidade “etiqueta-recetor” para a purificação por afinidade de proteínas recombinantes. Cinco hexapéptidos com sequências RWRWRW, WFWFWF, NWNWNW, RKRKRK and NNNNNN foram selecionados como etiquetas de afinidade para serem utilizadas como proteínas de fusão com a “Green Fluorescent Protein” (GFP). Os respectivos fragmentos de ADN foram desenhados, clonados no vetor de expressão pET21c e produzidos em *Escherichia coli* sob a forma solúvel, à exceção de WFWFWF-GFP. A biblioteca combinatorial foi sintetizada em fase sólida, tendo por base a reação de Ugi, apresentando 64 ligandos de afinidade com funcionalidades complementares para as etiquetas desenhadas. A avaliação da interação da biblioteca para a GFP e proteínas de fusão com a GFP foi efetuada através de cromatografia de afinidade em micro-escala, resultando na seleção de quatro ligandos de afinidade A4C8, A3C4, A7C1 e A4C7 para as proteínas de fusão marcadas com WFWFWF, NWNWNW, RKRKRK e GFP não marcada, respetivamente. A proteína WFWFWF-GFP foi explorada para estudos de enrolamento utilizando a matriz de afinidade com o ligando A4C8. O par de afinidade NWNWNW- GFP e A3C4 revelou baixos rendimentos na recuperação da proteína de fusão. O par de afinidade RKRKRK -GFP e A7C1 emergiu como um par promissor, tendo o ligando permitido uma recuperação de 30% da proteína de fusão com uma pureza de 50%, utilizando arginina como agente competidor. Adicionalmente, o par de afinidade apresentou um K_a de $2.45 \times 10^5 \text{ M}^{-1}$ e um Q_{max} de 0.5 mg/g de resina. Uma nova estratégia para a purificação de GFP foi também desenvolvida, sendo que adsorventes com o ligando A4C7 permitiram uma recuperação seletiva de GFP com 92% de rendimento e 94% de pureza, apresentando um K_a de $2.3 \times 10^5 \text{ M}^{-1}$ e Q_{max} de 1 mg GFP/g resina

PALAVRAS CHAVE- Proteína de fusão, Etiqueta de afinidade, Ligando biomimético, Cromatografia de afinidade, GFP

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGEMENTS	V
RESUMO	IX
TABLE OF CONTENTS	XI
INDEX OF FIGURES	XV
INDEX OF TABLES	XXI
ABBREVIATIONS	XXIII
BACKGROUND	XXV
CHAPTER 1 : AFFINITY PAIRS TAG-LIGAND FOR THE PURIFICATION OF RECOMBINANT PROTEINS	1
1.1 AFFINITY LIGANDS FOR RECOMBINANT PROTEIN PURIFICATION	2
1.2. PURIFICATION OF FUSION PROTEINS BY TAG-AFFINITY LIGAND STRATEGIES	5
1.2.1. Biological ligands as binding partners of affinity tags	8
1.2.2. Structural ligands as binding partners of affinity tags	14
1.2.3. Affinity tags with versatile properties	20
1.3 TAG REMOVAL	22
1.4 CONCLUDING REMARKS AND FUTURE TRENDS	22
CHAPTER 2: MATERIALS AND METHODS	25
2.1 MATERIALS	26
2.1.1 Chemicals.....	26
2.1.2 Biochemicals	27
2.1.3 Chromatographic materials	27
2.2 INSTRUMENTATION	27
2.3 GENERAL METHODS	29
2.3.1 Solid phase combinatorial synthesis of the library of Ugi-based affinity ligands.....	29
2.3.1.1 Epoxyactivation of agarose beads	29
2.3.1.2 Aldehyde functionalization of agarose beads.....	30
2.3.1.3 Synthesis of the solid phase combinatorial library	30
2.3.2 Production of crude extracts containing GFP-tagged and non-tagged GFP	35
2.3.2.1 Molecular Cloning of RWRWRW-GFP encoded gene in a pET-21c expression vector	35
2.3.2.2 RWRWRW tagged GFP small scale expression studies	42
2.3.2.3 Large scale expression of RWRWRW tagged GFP	44
2.3.2.4 Large scale expression of the remaining fusion proteins and GFP	46
2.3.2.5 Antimicrobial assays with hexapeptide RWRWRW (RW) ₃	46
2.3.2.6. Production of RWRWRW tagged GFP in E.coli BL21(DE3) cells	47
2.3.2.7. Production Inclusion bodies (IBs) with the WFWFWF-GFP system	47
2.3.2.8 Avoiding the formation of inclusion bodies	49

2.3.3 Screening of library based on the Ugi reaction with crude extracts containing GFP-tagged and non-tagged	49
2.3.4 Synthesis and re-screening of the promising lead ligands and the respective crude extracts containing GFP-tagged and non-tagged	50
2.3.4.1 Synthesis of the promising lead ligands	50
2.3.4.2 Screening of the promising lead ligands and the respective crude extracts containing GFP-tagged and non-tagged	51
2.3.5 Optimization and characterization of the lead affinity pairs “tag-receptor”	52
2.3.5.1 Scale-up synthesis of lead affinity ligands	52
2.3.5.2 Screening of the lead ligands with the respective target in on-column format....	52
2.3.5.3 Lead ligands selectivity	53
2.3.5.4 Determination of the binding constants	53
2.3.5.5 Optimization of the elution profile for the lead affinity pairs	53
2.3.5.6 Study of the lead affinity pairs in the automated system ÄKTA avant 25	56
CHAPTER 3: PRODUCTION OF GFP-TAGGED PROTEINS	59
3.1 INTRODUCTION	60
3.2 FRAGMENT DNA DESIGN	61
3.3 MOLECULAR CLONING OF RWRWRW-GFP ENCODING GENE IN A PET-21C VECTOR	63
3.4. RWRWRW TAGGED GFP SMALL SCALE EXPRESSION STUDIES	65
3.5. LARGE SCALE EXPRESSION OF GFP TAGGED PROTEINS	69
3.5.1 Production of GFP tagged proteins	69
3.5.2. Antimicrobial properties of the RWRWRW-GFP system.....	76
3.5.2.1 Antimicrobial assays with (RW) ₃	77
3.5.2.2. Production of RWRWRW tagged GFP in E.coli BL21(DE3) cells	78
3.5. 3 Production of inclusion bodies with the WFWFWF tagged GFP system	79
3.5.3.1 Solubilization and Refolding Strategies	79
3.5.3.2 How to avoid inclusion bodies?	80
3.5.3.2.1 Inductor levels Influence.....	80
3.5.3.2.2 Temperature Influence	83
3.6. CONCLUDING REMARKS.....	86
CHAPTER 4: COMBINATORIAL LIBRARY SCREENING WITH CRUDE EXTRACTS CONTAINING GFP-TAGGED AND NON-TAGGED PROTEINS	89
4.1 INTRODUCTION	90
4.2. SYNTHESIS OF A COMBINATORIAL LIBRARY OF AFFINITY LIGANDS.....	92
4.3. SCREENING OF THE UGI BASED COMBINATORIAL LIBRARY	94
4.3.1. Screening with RWRWRW tagged GFP crude extract	99
4.3.2. Screening with WFWFWF tagged GFP crude extract	103
4.3.2.1 Screening under Non denaturing conditions	103
4.3.2.2 Screening under Denaturing Conditions	107

4.3.3. Screening with NWNWNW tagged GFP crude extract	111
4.3.4. Screening with RKRKRK tagged GFP crude extract	114
4.3.5. Screening with NNNNNN tagged GFP crude extract.....	117
4.3.6. Screening with non-tagged GFP extract	118
4.4 SELECTION AND RE-SCREENING OF THE PUTATIVE LEAD LIGANDS	122
4.4.1. Re-Screening with RWRWRW tagged GFP crude extract	123
4.4.2. Re-Screening with WFWFWF tagged GFP crude extract.....	125
4.4.2.1 Non-Denaturing Conditions	125
4.4.2.2 Denaturing Conditions	127
4.4.3 Re-Screening with NWNWNW tagged GFP crude extract	129
4.4.5 Re-Screening with non-tagged GFP crude extract	134
4.5. THE LEAD AFFINITY PAIRS “TAG-RECEPTOR”	137
4.6 CONCLUDING REMARKS	138
CHAPTER 5: OPTIMIZATION AND CHARACTERIZATION OF THE LEAD AFFINITY PAIRS “TAG-RECEPTOR”	141
5.1 INTRODUCTION	142
5.2. COMPARISON OF THE AFFINITY SYSTEMS PERFORMANCE ON 96-WELL FORMAT AND ON-COLUMN	143
5.3 LEAD LIGANDS SELECTIVITY	147
5.4 DETERMINATION OF BINDING CONSTANTS	149
5.5 OPTIMIZATION OF THE ELUTION PROFILE FOR THE LEAD AFFINITY PAIRS	153
5.5.1 Affinity pair NWNWNW tagged GFP-A3C4.....	157
5.5.2 Affinity pair RKRKRK tagged GFP - A7C1	159
5.5.3 Affinity pair GFP-A4C7	163
5.6 SCALE UP OF THE AFFINITY PAIRS.....	168
5.6.1 Affinity pair WFWFWF tagged GFP - A4C8 under denaturing conditions	168
5.6.2 Affinity pair NWNWNW tagged GFP-A3C4.....	170
5.6.3 Affinity pair RKRKRK tagged GFP-A7C1	174
5.6.4 Affinity pair non-tagged GFP- ligand A4C7	178
5.6.5 Purification of GFP fusion proteins with A4C7 ligand	182
5.8 CONCLUDING REMARKS	185
CHAPTER 6: CONCLUDING REMARKS	187
BIBLIOGRAPHY	193

INDEX OF FIGURES

	PAGE
Figure 1.1 Overview of the classes of affinity ligands employed on affinity chromatography.....	3
Figure 1.2 Overview of a recombinant fusion protein purification process through the use of affinity tags fused to the target protein.....	7
Figure 2.1 Aldehyde functionalization of agarose beads.....	29
Figure 2.2 Reactional scheme of the multicomponent Ugi reaction synthesized in the solid phase.....	31
Figure 2.3 Visualization of the reactional 96-well filtration block used for the synthesis of the combinatorial libraries based on the Ugi reaction.....	32
Figure 3.1 Production of affinity tags fused to GFP.....	61
Figure 3.2 Scheme of the (A) fragment DNA and (B) respective nucleotide sequence that encodes for RWRWRW tagged GFP.....	62
Figure 3.3 Hydrolysis with restriction enzymes <i>Nhe</i> I and <i>Eco</i> R I of (A) pAP001 plasmid that carries the gene encoding for RWRWW-GFPfusion protein (insert) and (B) pET-21c expression vector.....	63
Figure 3.4 DNA quantification of the insert (A) and the double digested expression vector (B) by agarose gel electrophoresis (0.8 % (w/v) agarose) analysis using the (C) NZYDNA Ladder III from Nzytech.....	64
Figure 3.5 Evaluation of the hydrolysis of pAP001 of twelve selected colonies obtained from the subcloning procedure with restriction enzymes <i>Nhe</i> I and <i>Eco</i> R I by agarose gel electrophoresis (0.8 % (w/v) agarose)	65
Figure 3.6 Quantification of RWRWRW tagged GFP by GFP fluorescence spectroscopy, produced by induction of gene expression at an optical density 0.6-0.8 in <i>E. coli</i> BL21(DE3) cultures.....	67
Figure 3.7 SDS-PAGE evaluation of the amount RWRWRW tagged GFP produced over time, when using different concentrations of IPTG (A) 0.1 mM, (B) 0.5 mM, (C) 1 mM, (D) 1.5 mM and (E) 2 mM at optical density 0.6-0.8.....	67
Figure 3.8 GFP fluorescence intensity monitoring of the expression of RWRWRW tagged GFP when inducing the gene expression at optical density of 2.....	68
Figure 3.9 SDS-PAGE evaluation of the amount RWRWRW tagged GFP produced over the time, when using different concentrations of IPTG (A) 0.1 mM, (B) 0.5 mM, (C) 1 mM, (D) 1.5 mM and (E) 2 mM at optical density 2.....	69
Figure 3.10 Correlation between the GFP fluorescence intensity and optical density measurements with induction time during the large-scale expression.....	71
Figure 3.11 Normalization of the GFP fluorescence intensity with the respective optical density measurements for all the affinity tags fused to GFP.....	72
Figure 3.12 SDS-PAGE time course of large scale production of GFP tagged proteins in <i>E. coli</i> BL21(DE3).....	72
Figure 3.13 Cellular fractionation analyzed by SDS-PAGE.....	74

Figure 3.14 Inclusion bodies obtained for the WFWFWF-GFP production system.....	76
Figure 3.15 Antimicrobial assays between different RWRWRW concentrations and <i>E. coli</i> K12.....	77
Figure 3.16 Monitoring RWRWRW tagged GFP production on <i>E. coli</i> BL21(DE3) cells through GFP fluorescence intensity and optical density measurements over induction time and growth curve for <i>E. coli</i> BL21(DE3) cells.....	78
Figure 3.17 Impact of IPTG concentration on the expression of WFWFWF tagged GFP at (A) 30 °C and (B) 37 °C	81
Figure 3.18 SDS-PAGE analysis of the soluble and insoluble WFWFWF tagged GFP produced over the induction time for different concentrations of IPTG at (A) 30°C and (B) 37°C.....	82
Figure 3.19 GFP fluorescence intensity of soluble and insoluble forms of WFWFWFW tagged GFP obtained after protein expression induced at 0.6-0.8 with different IPTG concentrations at (A) 30 °C and (B) 37 °C.....	83
Figure 3.20 Impact of expression temperatures on the overall productivity of WFWFWF tagged GFP.....	84
Figure 3.21 SDS-PAGE analysis of the soluble and insoluble WFWFWF tagged GFP produced over the induction time for different temperatures	84
Figure 3.22 GFP fluorescence intensity of soluble and insoluble forms of WFWFWFW tagged GFP obtained after protein expression induced at 0.6-0.8 with 1 mM of IPTG at different expression temperatures	85
Figure 3.23 Fluorescence microscopy analysis of the insoluble form of WFWFWF tagged GFP produced at different temperatures with 1 mM IPTG.....	85
Figure 4.1 Strategy for the selection of lead ligands.....	92
Figure 4.2 Diversity of the compounds used in the synthesis of the Ugi based combinatorial libraries.....	94
Figure 4.3 Peptide structures used as affinity tags.....	98
Figure 4.4 Screening of combinatorial library for RWRWRW tagged GFP.....	100
Figure 4.5 Putative lead ligands structures with an enrichment factor ~ 5 for the RWRWRW tagged GFP.....	100
Figure 4.6- Putative lead ligands structures with an enrichment factor ~3- 4 for the RWRWRW tagged GFP.....	101
Figure 4.7 Putative lead ligands structures with an enrichment factor ~2- 3 for the RWRWRW tagged GFP.....	102
Figure 4.8 Correlation between (A) the isoelectric point and (B) Log P values of the selected putative lead ligands and the respective value of the enrichment factor obtained from the screening with the RWRWRW tagged GFP.....	103
Figure 4.9 Screening of combinatorial library for WFWFWF tagged GFP under non denaturing conditions.....	104

Figure 4.10 Structures of the putative lead ligands A6C1 and A1C5 with enrichment factors ~10 and 5, respectively for the WFWFWF tagged GFP under non-denaturing conditions.....	105
Figure 4.11 Structures of the putative lead ligands with enrichment factors ~2-4 that contain the same amine A2 for the WFWFWF tagged GFP under non-denaturing conditions	105
Figure 4.12 Structures of the putative lead ligands with enrichment factors ~2-4 that contain the same amine A3 for the WFWFWF tagged GFP under non-denaturing conditions.....	106
Figure 4.13 Correlation between (A) the isoelectric point and (B) Log P values of the selected putative lead ligands and the respective value of the enrichment factor obtained from the screening with the WFWFWF tagged GFP.....	106
Figure 4.14 Screening of combinatorial library for WFWFWF tagged GFP under denaturing conditions.....	109
Figure 4.15 Structures of the selected putative lead ligands for the WFWFWF tagged GFP under denaturing conditions.....	110
Figure 4.16 Screening of combinatorial library for NWNWNW tagged GFP.....	111
Figure 4.17 Putative lead ligands for NWNWNW tagged GFP, where ligand A7C8 presents an enrichment factor ~6 and ligands A6C8 and A3C5 ~5.....	112
Figure 4.18 Promising lead ligands for NWNWNW tagged GFP, where ligand A2C5 presents an enrichment factor ~4 and ligands A5C1 and A5C6 ~3.....	112
Figure 4.19 Lead ligands for NWNWNW tagged GFP with a common A2 amine and enrichment factor ~2.....	113
Figure 4.20 Lead ligands for NWNWNW tagged GFP with a common A3 amine and enrichment factor ~2.....	113
Figure 4.21 Correlation between (A) the isoelectric point and (B) Log P values of the selected putative lead ligands and the respective value of the enrichment factor obtained from the screening with the NWNWNW tagged GFP.....	114
Figure 4.22 Screening of combinatorial library for RKRKRK tagged GFP	115
Figure 4.23 Lead ligands A7C7 and A7C8 for RKRKRK tagged GFP.....	115
Figure 4.24. Lead ligands with A7 amine, with enrichment factor ~2	116
Figure 4.25 Ligands A2C4 and A5C4, selected as putative lead ligands for the RKRKRK tagged GFP.....	116
Figure 4.26 Correlation between (A) the isoelectric point and (B) Log P values of the selected putative lead ligands and the respective value of the enrichment factor obtained from the screening with the RKRKRK tagged GFP.....	117
Figure 4.27 Screening of combinatorial library for NNNNNN tagged GFP.....	118
Figure 4.28 Screening of combinatorial library for GFP.....	119
Figure 4.29 Lead ligand for GFP with the highest enrichment factor ~3.....	119

Figure 4.30 Ligands selected also as putative lead ligands with enrichment factor ~ 1.4 -2 for GFP.....	120
Figure 4.31 Lead ligands for GFP with an enrichment factor ~ 1	121
Figure 4.32 Correlation between (A) the isoelectric point and (B) Log P values of the selected putative lead ligands and the respective value of the enrichment factor obtained from the screening with the GFP.....	122
Figure 4.33 Estimation of GFP proteins retention onto the affinity adsorbents by SDS-PAGE analysis.....	123
Figure 4.34 (A) Binding of GFP and total protein for the putative lead ligands and RWRWRW tagged GFP and (B) comparison of enrichment factor values obtained from the two screening assays.....	124
Figure 4.35 SDS-PAGE evaluation of the screening of putative ligands and RWRWRW-GFP	125
Figure 4.36 (A) Binding of GFP and total protein for the putative lead ligands and WFWFWF tagged GFP under non-denaturing conditions and (B) comparison of enrichment factor values obtained from the two screening assays.....	126
Figure 4.37 SDS-PAGE evaluation of the screening of putative ligands for WFWFWF tagged GFP.....	127
Figure 4.38 Binding of GFP for the putative lead ligands and WFWFWF tagged GFP under denaturing conditions.....	128
Figure 4.39 SDS-PAGE evaluation of the screening of putative ligands for WFWFWF tagged GFP under denaturing conditions.....	128
Figure 4.40 (A) Binding of GFP and total protein for the putative lead ligands and NWNWN tagged GFP and (B) comparison of enrichment factor values obtained from the screening assays.....	129
Figure 4.41 SDS-PAGE evaluation of the screening of putative ligands and NWNWNW tagged GFP.....	130
Figure 4.42 (A) Binding of GFP and total protein for the putative lead ligands and RKRKRK tagged GFP and (B) comparison of enrichment factor values obtained from the screening assays.....	132
Figure 4.43 SDS-PAGE evaluation of the screening of putative ligands and RKRKRK tagged GFP.....	133
Figure 4.44 (A) Binding of GFP and total protein for the putative lead ligands and GFP and (B) comparison of enrichment factor values obtained from the screening assays.....	135
Figure 4.45 SDS-PAGE evaluation of the screening of putative ligands and GFP.....	136
Figure 4.46 Lead ligands selected for GFP tagged proteins and non-tagged GFP.....	137
Figure 4.47 Binding evaluation ($\mu\text{g bound/ g resin}$) of each one of the GFP tagged proteins and non-tagged GFP and the respective affinity ligand.....	138
Figure 5.1 Different stages required for the optimization of the lead affinity systems.....	143

Figure 5.2 Comparison of the percentage of binding of (A) GFP protein and (B) total protein and (C) enrichment factor for the 96-well format and on-column assays for the lead ligands	145
Figure 5.3 Comparison of the percentage of binding of (A) GFP protein and (B) total protein for the 96 well format and on-column for blank agarose.....	146
Figure 5.4 Selectivity of the ligands A3C4 and A7C1.....	147
Figure 5.5 SDS-PAGE analysis of the binding A3C4 to (A) RKRKRK tagged GFP and (B) GFP.....	148
Figure 5.6. SDS-PAGE analysis of the binding A7C1 to (A) NWNWNW tagged GFP and (B) GFP.....	148
Figure 5.7 Adsorption and desorption scheme between an immobilized affinity ligand and a target protein.....	149
Figure 5.8 Binding isotherms and resulting static binding capacities for the affinity pairs (A) NWNWNW tagged GFP and A3C4, (B) RKRKRK tagged GFP and A7C1 and (C) GFP and A4C7.....	152
Figure 5.9 Strategy for the optimization of the elution conditions.....	156
Figure 5.10 Microspecies distribution (%) for the ligand A3C4 over the pH values between 3 and 11.....	158
Figure 5.11 Influence of pH on the charge of the NWNWNW peptide without free C-terminal.....	158
Figure 5.12 Microspecies distribution (%) for ligand A7C1 over pH values between 3 and 11.....	160
Figure 5.13 Influence of pH on the charge of RKRKRK peptide without free C-terminal.....	160
Figure 5.14 Influence of the different elution buffers on the recovery of RKRKRK tagged GFP.....	161
Figure 5.15 Second approach for the optimization of the elution conditions for the recovery of RKRKRK tagged GFP.....	163
Figure 5.16 Influence of the different elution buffers on the recovery of GFP.....	165
Figure 5.17 Microspecies distribution (%) for the ligand A4C7 over pH values between 3 and 11.....	165
Figure 5.18 Crystal structure of GFP protein (1QYO).....	166
Figure 5.19 Second approach for the optimization of the elution conditions for the recovery of GFP.....	167
Figure 5.20 Strategy used for the refolding of the WFWFWF tagged GFP on A4C8 functionalized matrix performed on the AKTA™ avant 25 with the expected chromatogram	169
Figure 5.21 SDS-PAGE analysis of the samples from the purification of WFWFWF tagged GFP and the ligand A4C8 under denaturing conditions.....	170

Figure 5.22 Percentages of NWNWNW tagged GFP bound, eluted and regenerated from affinity adsorbent containing A3C4.....	171
Figure 5.23 Chromatograms obtained for the purification of NWNWNW tagged GFP with affinity ligand A3C4.....	172
Figure 5.24 SDS-PAGE analysis of the samples from the purification of NWNWNW tagged GFP and the ligand A3C4.....	172
Figure 5.25 Percentages of RKRKRK tagged GFP bound, eluted and regenerated from affinity adsorbent containing A7C1.....	175
Figure 5.26 Chromatograms obtained for the purification of RKRKRK tagged GFP with affinity ligand A7C1.....	176
Figure 5.27 SDS-PAGE analysis of the samples from the purification of RKRKRK tagged GFP and the ligand A7C1	176
Figure 5.28 Percentages of GFP bound, eluted and regenerated from affinity adsorbent containing A4C7.....	178
Figure 5.29 Chromatograms obtained for the purification of GFP with affinity ligand A4C7.....	179
Figure 5.30 SDS-PAGE analysis of the samples from the purification of GFP and the ligand A4C7.....	180
Figure 5.31 Percentages of GFP fusion protein bound, eluted and regenerated from affinity adsorbent containing A4C7.....	182
Figure 5.32 Chromatograms obtained for the purification of GFP fusion protein with affinity ligand A4C7.....	183
Figure 5.33 SDS-PAGE analysis of the samples from the purification of GFP fusion protein and the ligand A4C7.....	183
Figure 5.34 SDS-PAGE analysis obtained from the chromatographic assays from the purification of GFP fusion protein and the ligand A4C7 when using 0.1M Glycine-NaOH pH 9 50% ethylene glycol.....	184

INDEX OF TABLES

	PAGE
Table 1.1 Comparison the affinity ligands employed on purification of proteins through affinity chromatography.....	5
Table 1.2 Comparison of the general properties of proteins and peptides used as affinity tags	8
Table 1.3 Overview of biological ligands involved as binding partners of affinity tags.....	10
Table 1.4 Examples of a peptide and protein affinity tags and their respective biological ligands employed on the purification of fusion proteins based on affinity chromatography.....	11
Table 1.5 Overview of the structural ligands involved as binding partners of affinity tag.....	16
Table 1.6 Examples of commercially available affinity tags and respective purification supports based on biological ligands.....	18
Table 1.7 Examples of commercially available affinity tags and respective purification supports based on structural ligands.....	19
Table 2.1 List of the amines compounds used on the Ugi reaction for the solid phase combinatorial library.....	33
Table 2.2 List of the carboxylic acid and isocyanide compounds used on the Ugi reaction for the solid phase combinatorial library.....	34
Table 2.3 Promising lead ligands selected for each respective target from the first screening between the combinatorial library based on the Ugi reaction and the crude extracts containing GFP-tagged and non-tagged.....	51
Table 2.4 Elution conditions employed on the first approach of the optimization.....	54
Table 2.5 Elution conditions employed on the second approach of the optimization.....	55
Table 2.6 Best elution conditions employed on the screening between the affinity pairs on the ÄKTA avant 25 automated system.....	57
Table 3.1 Percentage of tagged and non-tagged GFP produced over the total amount of total protein produced by <i>E.coli</i> cells.....	75
Table 3.2 Expression profiles summary of each affinity tag fused to GFP.....	87
Table 4.1 Maximum values of the enrichment factor expected with GFP-tagged and non-tagged proteins.....	97
Figure 4.2 Properties of the peptides used as affinity tags.....	99
Table 4.3 Putative lead ligands selected for each of the GFP-tagged and non-tagged proteins.....	122
Table 4.4 Elution of RKRKRK tagged GFP at pH 11.....	133
Table 5.1 Kinetic parameters obtained for each of the affinity pairs after fitting with Langmuir model.....	153
Table 5.2 Comparison of the results obtained for the affinity pair NWNWNW tagged GFP/A3C4	173
Table 5.3 Comparison of the results obtained for the affinity pair RKRKRK tagged GFP/A7C1.....	177

Table 5.4 Comparison of the results obtained for the affinity pair GFP/A4C7.....181
Table 6.1 Comparison of the RKRKRK affinity tag with similar tags commercial available.....190
Table 6.2 Purification methods used on the purification of Green Fluorescent Protein (GFP).191

ABBREVIATIONS

AMP – Antimicrobial peptide
ATPS - Aqueous two-phase solvent
BSA – Bovine Serum Albumin
BCA - Bichinchoninic Acid
CBP – Calmodulin binding protein
CIP- Cleaning-in-place
CMA - Carboxymethylaspartate
DDDDK – Peptide with four Aspartic acid residues and one Lysine residue
DNA - Deoxyribonucleic acid
DTT- DL-Dithiothreitol
E.coli – *Escherichia coli*
EDTA - Ethylenediaminetetraacetic acid
EGTA – Ethylene glycol tetraacetic acid
ELPs Elastin-like polypeptides
FITC - Fluorescein isothiocyanate
GFP - Green Fluorescent Protein
GST- Glutathion-S-transferase
HIC – Hydrophobic interaction chromatography
His-tag –Histidine hexapeptide
HSA- Human serum albumin
HTS - High Throughput Screening
IDA - Iminodiacetic acid
IEX – Ion-exchange chromatography
IgG - Immunoglobulin G
IMAC - Immobilized metal affinity chromatography
IPTG - β -D-1-thiogalactopyranoside
LB - Luria Broth
MCR – Multicomponent reaction
MBP – Maltose binding protein
NNNNNN - Peptide with six asparagine residues
NTA - Nitrilotriacetic acid
NWNWNW: Peptide with asparagine and tryptophan residues
NusA- N-utilization substance A
OD₆₀₀ – Optical density at 600 nm
PBS – Phosphate buffered saline buffer
PDB: Protein data bank
PolyArg - Polyarginine-tag
PVDF – polyvinylidene difluoride
Pyrene: 1-Pyrene methylamine hydrochloride

RWRWRW: Peptide with arginine and tryptophan residues
RKRKRK: Peptide with arginine and lysine residues
R1: Substitution of the first chloride of triazine immobilized on to agarose
R2: Substitution of the second chloride of triazine immobilized on to agarose
SBP - Streptavidin-binding peptide
SDS-PAGE - Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIP- Sterilization-in-place
SpA - Staphylococcal protein A
SpG - Staphylococcal protein G
TAG- Triacylglycerol
TAP- Tandem affinity chromatography
SPR – Surface Plasmon resonance
TEMED - Tetramethylethylenediamine
Tris -Tris(hydroxymethyl)aminomethane
Trx - thioredoxin
WFWFWF: Peptide with tryptophan and phenylalanine residues
TACN - 1,4,7-triazacyclononane

BACKGROUND

The market of therapeutic proteins represents one of the fastest-growing segments of pharmaceutical industry and the most prominent best-selling classes include monoclonal antibodies, hormones and growth factors [1-2]. With the advances on genetic engineering it became possible the production of the therapeutic proteins in different microorganisms where *Escherichia coli* represents the most widely used platform, covering 30% of the total production of recombinant proteins among mammalian, insect and yeast cells [2]. Along with these advances, regulatory demands on purity became stricter, and the cost-effectiveness of the end product has been a driving force for the development of more efficient downstream processes [3-5].

Currently, there is not a universal strategy for the purification of recombinant proteins, as the individual biochemical and biophysical properties of each target protein dictate the production and the purification process. Affinity chromatography is considered the most widely used chromatographic method on the downstream processing due to the high yields of recovered protein with purity over 90% [3, 6], but there is a lack of suitable affinity ligands for a wide range of proteins. Moreover, the diversity of proteins makes the development of general purification and capturing strategies difficult. Therefore, the incorporation of a genetically encoded purification affinity tag is an attractive choice for the purification of recombinant proteins, while contributing for a reduction on the number of chromatographic processes involved in a purification scheme [7-8].

The use of purification affinity tags requires a respective binding receptor immobilized onto a chromatographic matrix and their use is dependent on the molecular recognition mediated by the non-covalent interactions and the complementarity between the pair tag-receptor [9]. To date, the purification affinity tags commercially available and those referred in the literature, are not specifically designed to the receptor. The common receptors comprise biological and structural ligands that are able to recognize peptides or proteins used as purification affinity tags. One group of tags includes native or engineered proteins: enzymes (e.g. Glutathione-S-transferase), protein binding domains (e.g. protein A and Z domain) and carbohydrate binding domains (e.g. maltose-binding protein) [7-8, 10-11]. The other group of purification tags encloses peptides: antigenic epitopes (e.g. FLAG), protein-binding peptides (e.g. calmodulin-binding peptide), streptavidin-binding peptides (e.g. Strep-tag), charged peptides (e.g. poly-arginines), and metallo-binding peptides (e.g. poly-histidine) [7-8, 10-11]. Short peptide tags present several advantages over larger tags (as proteins) as they are less likely to interfere with the expression, structure and function of the target recombinant protein and the cloning procedure is facilitated [8].

The corresponding capturing affinity ligands comprise biological ligands (e.g. complementary monoclonal antibodies or haptens) and structural ligands (e.g. ion-exchange resins, metal-chelate) [7-8, 10-11]. In general, biological ligands tend to have high binding and specificity capacities, but also high associated costs, low stability and re-usability [12-14]. Available structural ligands are usually inexpensive to produce but display low selectivities and as a consequence the purity of the final protein can be compromised, in particular when using metal chelator ligands which are more prone to leaching [15-17].

A class of synthetic affinity ligands denominated as biomimetic ligands have been employed in the affinity purification of specific proteins such as immunoglobulins, [18-20], glycoproteins [21-22], prion proteins [23], human recombinant factor VII, [24] and elastases [25]. These can be *de novo* rationally designed for a specific target to better mimic the structure and binding of natural biological ligands [12-14, 26], and can present high selectivity combined with low production costs. Additionally, the biomimetic ligands are highly robust and can operate over a wide range of chemical conditions (e.g. pH, buffer composition, ionic strength) making cleaning in place and matrix re-utilization possible [12-14].

To date, there is no purification process based on the use of an affinity pair that combines a biomimetic ligand as a binding partner for a specific affinity tag.

Therefore, this project aims the development of a new “tag-receptor” to be used as generic platform for recombinant protein purification. The novelty of this work is the *de novo* creation of a tailor-made affinity pair, where the tag will be a small peptide and the receptor a biomimetic synthetic affinity ligand designed target the tag.

CHAPTER 1

AFFINITY PAIRS TAG-LIGAND FOR THE PURIFICATION OF RECOMBINANT PROTEINS

SUMMARY

The purification of recombinant proteins based on affinity chromatography is one of the most efficient purification strategies that allows high recovery yields of and purity of the target protein. However, this strategy is dependent on the availability of affinity adsorbents. Adsorbents comprising synthetic affinity ligands which combine the selectivity of biological ligands with the cost-effectiveness of structural ligands are promising viable options. The diversity of proteins requires a generic platform for the purification of recombinant proteins. Therefore, genetically encoded affinity tags are an attractive and viable option, which only requires specific receptors for capturing the fusion protein through the affinity tag tail. There is a wide range of available affinity pairs "tag-receptor" combining biological or structural ligands with respective binding partners such as peptide or proteins tags

Although current affinity pairs are well established for purification purposes there is still a lack of compromise between high selectivity and affordable costs within the affinity pairs so far developed.

1.1 AFFINITY LIGANDS FOR RECOMBINANT PROTEIN PURIFICATION

Recombinant proteins produced through recombinant DNA technology present one of the major classes of biopharmaceuticals and include erythropoietin, insulin, interferons, colony-stimulating factors, blood factors and growth hormones, employed in the treatment of cancer, diabetes, auto-immune and growth disturbances, among many others [3, 27]. In 2009, therapeutic proteins were the leaders of the global biopharmaceutical industry with a market size of 50 billion Euros, and by 2016, the market is estimated to reach 120 billion Euros [27]. The manufacturing of biopharmaceuticals is constantly evolving to reach higher productivity, quality, and lower production costs [4-5, 28]. Upstream and downstream processing are intimately linked, where the former deals with the production and the latter with the separation steps to ensure a highly purified final product [5]. Downstream processing costs can be as high as 80% of the total production cost of a therapeutic recombinant protein [29-30]. This is mainly a consequence of the several purification steps needed, combined with the strict regulatory demands on purity, and the required optimization processes for each particular biologic target [3-5, 29]. Chromatography is a well established platform for downstream bioprocesses, being currently economically feasible and yielding recovery of bioproducts with purities as high as 97-99% [3, 5, 31-32]. Among the several chromatographic techniques available for bioseparation purposes, affinity chromatography is extremely popular [6] due to its high selectivity and reduction of process steps [5, 29, 33].

The concept of affinity chromatography emerged in the 1970's and exploits the reversible and selective interaction between two entities: an affinity ligand immobilized on a solid support, and a target biomolecule to purify [14, 29-30, 34-35]. The solid supports most widely used in affinity purification are composed of cross-linked agarose, polymethacrylate, polycrylamide, cellulose, and silica, and include porous and non-porous particulates, membranes and monoliths [14, 34-37]. The properties required for an efficient and effective affinity separation matrix include chemical inertness, mechanical and chemical stability, proper pore and particle size [14, 34, 36-37]. The chemical inertness of the matrix is related with the minimization of nonspecific interactions between the matrix material and the target protein to be purified [34, 37]. As the support should interact in a minimal way with the mobile phase, the matrix support should be hydrophilic and non-charged to minimize the formation of ionic interactions [37]. Concerning the stability of the material, the matrix must be physically and chemically stable under a wide range of conditions including those employed during surface modification, protein adsorption, elution, regeneration, matrix sterilization in place (SIP) and cleaning in place (CIP). These conditions include extreme pH values, high and low temperatures and the use of organic solvents, enzymes in the sample, detergents and disruptive eluent conditions [14, 34, 36-37].

Matrix selectivity can be introduced through the chemical functionalization of the solid support with desired affinity ligands, which can be divided into three main categories: biological, structural and synthetic ligands (Table 1.1 and Fig 1.1)

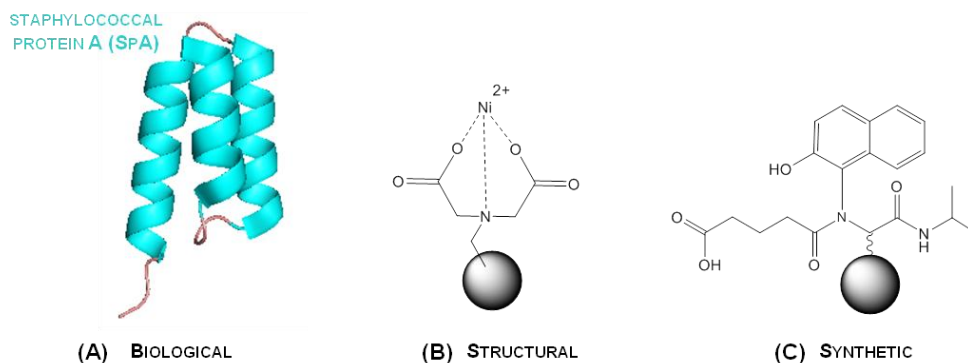


Figure 1.1 Overview of the classes of affinity ligands employed on affinity chromatography, where (A) is a biological ligand(staphylococcal protein A, PDB: 1DEE), (B) a structural ligand (metal chelate such as iminodiacetic acid chelated to Ni^{2+}) and (C) a synthetic biomimetic ligand (ligand A3C1 specific for immunoglobulins [38]). The solid support is representing agarose beads (●).



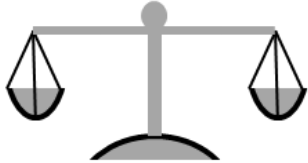
Biological ligands comprise of biomolecules obtained from natural biological sources and from *in vitro* selection techniques. The former include natural receptors with high selectivity and affinity for their target, such as peptides, antibodies, antigens and receptor proteins [13, 39-40], but usually associated with high costs of production and purification, poor stabilization under SIP and CIP conditions, as well as potential leakage and end-product contamination. Examples of common affinity biological ligands include immunoglobulins against a target protein (antigen), bacterial immunoglobulin-binding domains such as staphylococcal proteins A, G and L [41-45], natural lectins targeting glycoproteins [46]. Novel biological affinity ligands can be obtained through *in vitro* selection techniques as phage, ribosome or yeast display [47], being phage display the most popular for these purposes.

Structural ligands are produced at affordable prices with relatively easy chemistries, and possess high resistance to harsh conditions. These ligands comprise ion-exchange [48-50], hydrophobic [51], metal chelate [16], thiophilic ligands [13-14, 52], boronate ligands [53-57], and mixed-mode affinity ligands [58-62]. The multimodal affinity chromatography relies on multiple forms of interactions and the respective mixed-mode affinity ligands usually combine a hydrophobic moiety with ionic and hydrophilic groups [58]. However, their limited selectivity and low affinity for specific targets usually obliges the design of multi-step purification schemes.

Synthetic affinity ligands have been developed in an attempt to overcome disadvantages of natural and structural ligands, by combining the best of two worlds: affinity molecular recognition features associated with high resistance to chemical and biological degradation, high scalability, as well as low costs and low toxicity [30, 39-40]. Synthetic ligands have been well established since the 90's being mostly represented by biomimetic ligands or *de novo* designed ligands

tailor-made for a specific target protein [13]. The research strategy followed for the development of these ligands involves ligand design using *in silico* molecular modelling tools as a first step [13, 63], followed by on-bead combinatorial synthesis and high-throughput screening (HTS) of ligand libraries, with the aim of testing enough diversity of compounds while reducing costs and time [64], and further ligand optimization. Different scaffolds were already used for the development of synthetic ligands. The triazine scaffold is a well-established technique by Lowe and co-workers. This scaffold was in the basis of alternatives of proteins A, G and L for the purification of antibodies [18, 20, 65], on the development of an artificial lectin for purification of glycoproteins [21-22], as well as other protein targets such as prion proteins [23], human recombinant factor VII, [24] and elastases [25]. This strategy was also considered by other authors for the purification of alkaline phosphatase [66] and human tissue plasminogen activator [67]. More recently, a different scaffold based on the Ugi reaction was exploited by the Lowe group [38]. The Ugi reaction is a multicomponent reaction that involves four main compounds, an aldehyde, an amine, a carboxylic acid and an isocyanine. So far, this strategy has been mainly used for the development of antibody-binding ligands [38, 68-69].

Table 1.1 Comparison the affinity ligands employed on purification of proteins through affinity chromatography. The affinity ligands can be divided as biological, structural, and synthetic (biomimetic).

		AFFINITY LIGANDS		
		BIOLOGICAL	STRUCTURAL	SYNTHETIC – BIOMIMETIC
ADVANTAGES				
		<ul style="list-style-type: none"> - High Selectivity - Well-established - Optimized 	<ul style="list-style-type: none"> - Robustness - Cost-effectiveness - Stability to sterilization and cleaning in place - Scalability - Reusability 	<ul style="list-style-type: none"> - High Selectivity - Robustness - Cost-effectiveness - Scalability - Resistance to chemical and biological degradation - Stability to sterilization and cleaning in place - Reusability
DISADVANTAGES		<ul style="list-style-type: none"> - Poor stability - Leakage - End product contamination - Limited lifespan - Low scale-up potential - High costs 	<ul style="list-style-type: none"> - Reduced Selectivity - Leaching (e.g. Metal chelate) - Toxicity (e.g. Metal chelate) 	<ul style="list-style-type: none"> - Potential toxicity - Limited availability

1.2. PURIFICATION OF FUSION PROTEINS BY TAG-AFFINITY LIGAND STRATEGIES

The diversity of proteins and their biochemical properties makes the development of universal purification and capturing strategies difficult. Most proteins of interest lack a suitable, specific and robust affinity ligand for capture on a solid matrix. Genetically encoded affinity tags are a viable and common option for the purification of recombinant fusion proteins and also represent important tools for structural and functional proteomics initiatives. This approach requires the existence and availability of specific receptors for capturing the fusion protein through an encoded affinity tag tail (Fig. 1.2), denominated as affinity pairs tag-receptor. Currently available affinity pairs tag-receptor fall within one of the categories: protein-protein, protein-small biological ligands, peptide-protein or peptide-metal chelating ligands.

Affinity tags display different size ranges from a single amino acid to entire proteins, and can be genetically fused to the N- or C-terminal of the target biomolecule [7-8, 10-11, 70-71]. Apart from facilitating the purification process, affinity tags can also enhance protein solubility and stability, increase expression levels [72-73] and allow labelling for cellular localization and imaging studies [70]. An overview of the main advantages and disadvantages associated with different affinity tags are presented in Table 1.2., and have been already thoroughly discussed on literature [7-8, 10-11, 70, 72-75]. In general, shorter affinity tags (peptides) are more attractive as they are less likely to interfere with the expression, structure and function of the target protein. Due to their smaller size, the peptides represent a smaller burden for the host during fusion protein production and might interfere less with the tertiary structure and biological activity of the target fused protein [8, 74-75]. Therefore, the removal of the peptide tags can be exempt, decreasing thus the overall costs of the purification process based on affinity peptide tags [8, 75]. For certain proteins, mainly those with applications in therapy, the presence of the affinity tag can compromise protein properties and tag removal is usually compulsory [8, 10-11, 70, 76], which will be further discussed.

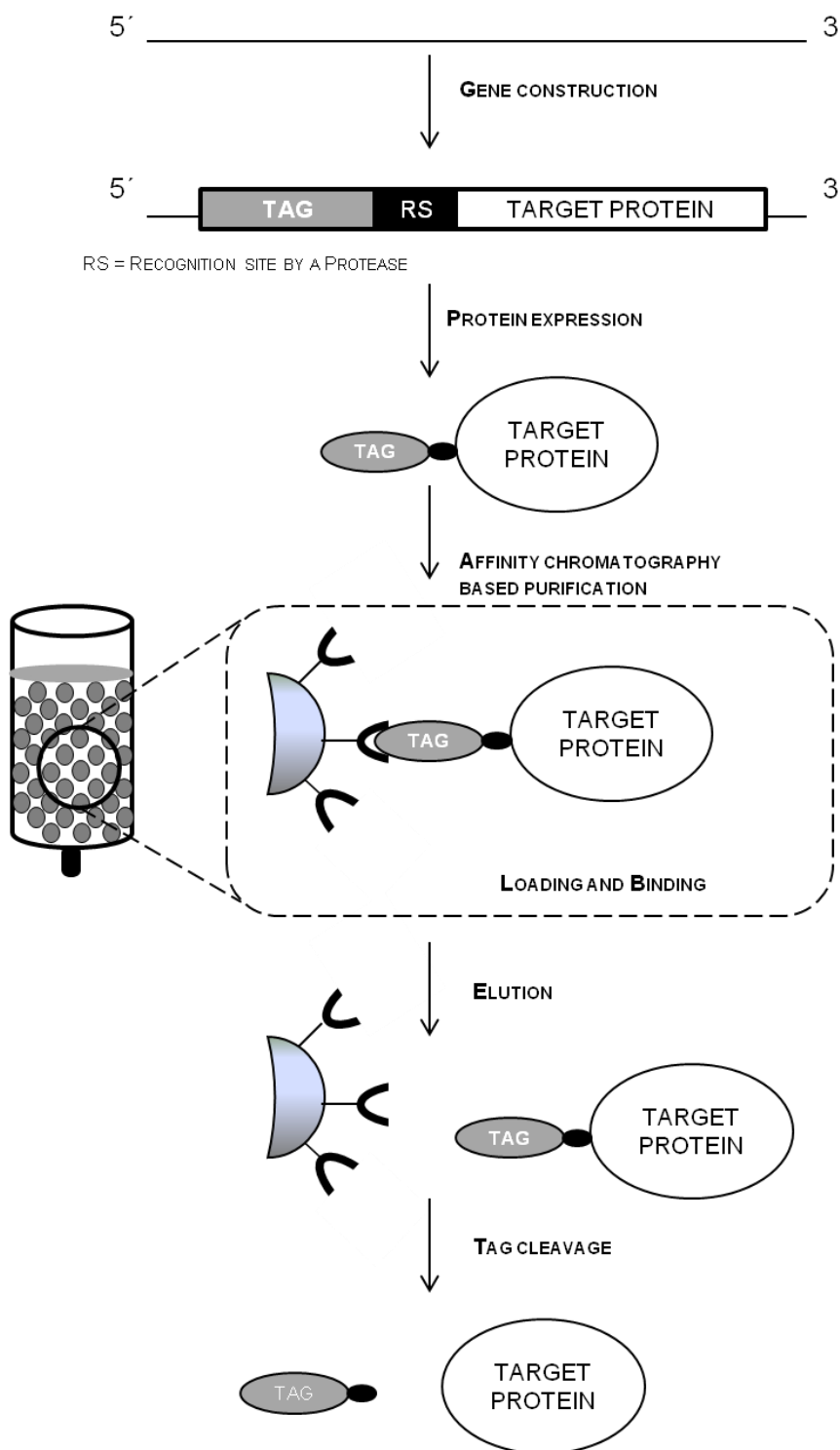


Figure 1.2 Overview of a recombinant fusion protein purification process through the use of affinity tags fused to the target protein. The process comprises several stages as gene construction of the fusion protein, where the gene that encodes the affinity tag is fused to the nucleotide sequence of the target protein. Then the fusion protein is expressed in the selected host and purified by affinity chromatography, through the molecular recognition between the tag and the ligand. After protein recovery, the tag might be removed leading to a pure target protein.

Table 1.2 Comparison of the general properties of proteins and peptides used as affinity tags

AFFINITY TAGS		
	PROTEIN	PEPTIDE
ADVANTAGES	<ul style="list-style-type: none"> - Enhance protein solubility and stability - Genetically fused to N- or C-terminal of the target protein - Facilitate protein purification - Molecular diversity 	<ul style="list-style-type: none"> - Easy cloning - Genetically fused to N- or C-terminal of the target protein - Minimal effect on tertiary structure and biological activity - Less burden for the host - Enhance protein solubility and yield - Facilitate protein purification - Useful for matrix-assisted protein refolding - Tag removal might not be required - Allows accessibility to tag removal
DISADVANTAGES	<ul style="list-style-type: none"> - Large size and complex structure - Cysteine content - Interference with tertiary structure and biological activity of the target protein - High energy consumption by the host - Low scale-up potential 	<ul style="list-style-type: none"> - Can be buried within the internal structure of the protein - Mask functional binding sites - A wide range of the respective binding partners are biological ligands - High costs of the matrix resin

1.2.1. Biological ligands as binding partners of affinity tags

Biological ligands involved on the molecular recognition with a partner affinity tag include peptides, protein and carbohydrates (Table 1.3 and 1.4).

Immunoglobulin-based adsorbents for affinity chromatography purposes are usually very selective for the target proteins but the costs incurred are very high. The first affinity tag developed was based on the intrinsic selectivity and affinity between the bacterial immunoglobulin-binding domain staphylococcal protein A (SpA) and the Fc region of mammalian IgG [77]. SpA is well-known as a biological adsorbent for the purification of immunoglobulins, but it can be also used as an affinity tag for fusion to target proteins, as alkaline phosphatase [78]. In these situations, the purification of the fusion target protein was conducted on a single step of IgG based affinity chromatography with protein recovery at acidic pH [78]. According to Nilsson *et al* [77], SpA presents five homologous domains (E, D, A, B and C) where IgG binds preferentially to B domain. A mutated version of B domain was developed and denominated as Z domain to improve resistance towards tag removal by chemical methods

[79]. The bacterial immunoglobulin-binding domain staphylococcal protein G (SpG) has been also studied as a fusion partner due to its bifunctional behaviour – SpG is composed of different domains (A, B, C and D) with affinity for both IgG and human serum albumin (HSA) [80-81] – allowing the purification of SpG tagged proteins through HSA and IgG affinity chromatography [8, 71, 80-81].

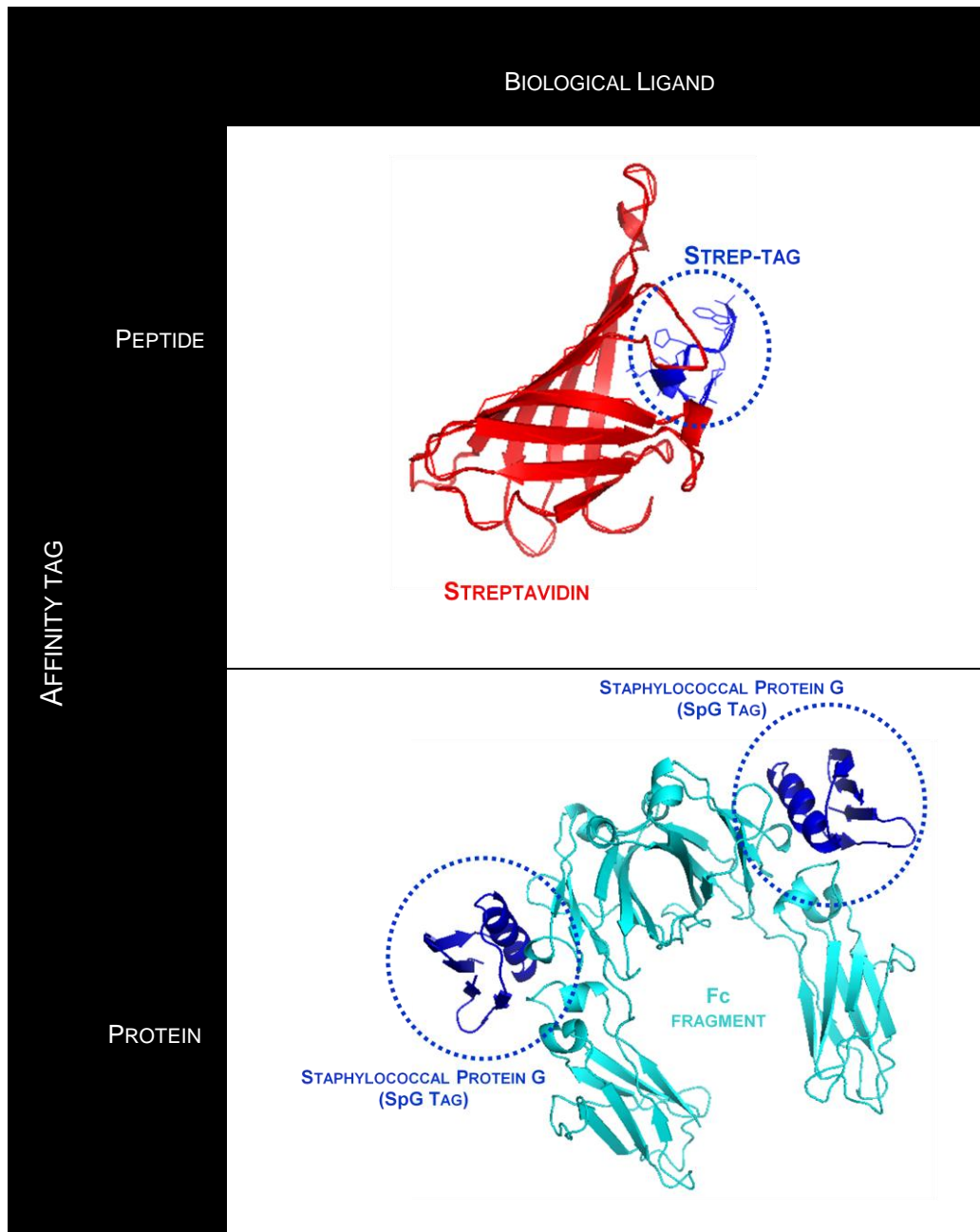
Other affinity tags recognizing immunoglobulins based adsorbents include the peptide epitopes FLAG, c-myc, T7, hemagglutinin antigen (HA) and Softags [8, 11]. The FLAG tag is the most popular, and has been employed on the purification of several recombinant proteins such as immunoglobulins, cytokines, gene-regulatory proteins [82]. It is a hydrophilic peptide with the sequence DYKDDDDK with high affinity for the monoclonal antibodies M1 and M2, being M1 binding calcium dependent [82-85]. Therefore, the elution of FLAG fusion proteins from M1 resins can be carried out under mild conditions in the presence of metal chelators such as EDTA, whereas elution from M2 resins requires a decrease in pH [82-85]. An attractive feature of FLAG tag is the possibility of simultaneous purification and removal by enterokinase which recognises the DDDDK sequence [82, 84]. The c-myc is a product of a proto-oncogene which epitope presents high affinity for the monoclonal antibody 9E10 [86]. This affinity pair has been mostly used as a tool for the detection of recombinant proteins through immunoblotting assays rather than for purification processes [75, 86-87]. The affinity tags T7 and HA are mostly used on immune-detection assays [11, 73], where the T7-tag is a leader peptide of phage T7 with eleven amino acid with affinity for the anti-T7 monoclonal antibody [88-89], and the HA tag is a peptide epitope of the influenza virus hemagglutinin [90] recognized by the monoclonal antibody 12 CA5 [91-92]. The epitope tags denominated as Softags are also employed on immunoaffinity chromatography by using polyols responsive monoclonal antibodies as biological adsorbents [93-94]. According to their designation, these tags allow a soft competitive elution with polyols which are low molecular weight compounds [93-94]. There are three Softags, where Softag 1 corresponds to a thirteen amino acid sequence near the C-terminal of the β' subunit of *E.coli* RNA polymerase [95]; Softag 2 is a repeat heptapeptide found on C-terminal of RNA polymerase I [96]; and Softag3 is an epitope near the N-terminal of human transcription factor [94].

Table 1.3 Overview of biological ligands involved as binding partners of affinity tags

	BIOLOGICAL LIGAND	TAG	TAG SIZE	APPLICATION	REFERENCE
PEPTIDES AND PROTEINS	Glutathione	GST	26 kDa	Purification and solubility	[97]
	Calmodulin	CBP	26 aa	Purification	[98]
	S protein (S-fragment RNase A)	S-tag	15 aa	Purification	[99]
	Streptavidin	Strep-tag, Strep-tag II	9 aa	Detection and purification	[100-101]
		Nano tag	9-15 aa		[102]
		SBP	38 aa		[103]
	Strep-Tactin (modified streptavidin)	Strep tag II	8 aa		[104]
	NeutraAvidin	AviD-tag	6 aa		[105]
	IgG	Protein A (SpA)	14-31 kDa	Purification	[78]
		Z domain	7 kDa		[79]
	IgG / HSA	Protein G (SpG)	28 kDa	Purification	[81]
	Monoclonal antibody M1, M2	FLAG	8 aa	Purification	[85]
	Monoclonal antibody 9E10	c-myc	10 aa	Purification	[86]
	Anti-T7 monoclonal antibody	T7 tag	11 aa	Detection and Purification	[106]
	Monoclonal antibody 12 CA5	HA tag	9 aa	Detection and Purification	[91]
Polyol responsive monoclonal antibodies	Softag 1/2/3	13/7/8 aa	Detection and Purification	[94, 96, 107-108]	
CARBOHYDRATES	Cross-linked amylose	MBP	42 kDa	Purification and solubility	[109]
	Cellulose	Cellulose binding protein	100 kDa	Purification	[110]
	Chitin	Chitin binding domain	5 kDa	Purification	[111]
	Starch	Starch binding domain	133 aa	Purification	[112]

GST- Glutathione-S-transferase; CBP- calmodulin binding peptide; SPB- streptavidin-binding peptide; IgG –immunoglobulin G; HSA- human serum albumin; HA- hemagglutinin antigen; MPB- maltose binding protein

Table 1.4 Examples of a peptide and protein affinity tags and their respective biological ligands employed on the purification of fusion proteins based on affinity chromatography. The peptide tag is the Strep-tag, an eight amino acid sequence, with the affinity for streptavidin protein (PDB: 1RST), whilst the example of a protein used as an affinity tag is related with the staphylococcal protein G and the respective biological ligand, the Fc fragment of immunoglobulin G (PDB:1FCC).



Another important group of biological affinity ligands includes streptavidin, where the tag-receptor system is based on the selective natural binding pair avidin-biotin with a dissociation constant of 10^{-15} M [113]. The first tag with affinity for streptavidin was Strep-tag, a nine amino acid peptide selected from a random peptide library [100], generally fused to the C-terminal of the target protein. According to structural studies, this affinity tag binds to the same pocket of

biotin, the natural partner of streptavidin [101]. The main advantages of Strep-tag are related to the resistance to proteolysis *in vivo*, the lack of interference with expression in *E.coli* and the mild elution conditions employed which include competitive elution with a biotin or analogue compound such as iminobiotin [100]. However, a limitation of the Strep-tag is the restriction to C-terminal fusion and an improved version of Strep-tag was developed and denominated as Strep-tag II with equilibrium dissociation constant of 37 μM [101]. Also, at the same time, an engineered streptavidin chromatographic support (Strep-Tactin) was developed to improve the binding capacity towards Strep-tag II. This new version of streptavidin was generated from random mutagenesis on the amino acids 44-47 located at the flexible loop region near the binding site [114]. The affinity pair Strep-tag II – Strep tactin has been extensively used on a variety of applications [114], still affinity tags with higher affinity for streptavidin matrices were searched: The streptavidin-binding peptide (SPB) tag comprising a 38 amino acid sequence with an equilibrium dissociation constant of 2.5 nM was selected from a peptide library screened against immobilized streptavidin [103]; The Nano-tag combining the small size of Strep-tag II and a nanomolar dissociation constant of 4 nM [102]. This tag is a 15 amino acid peptide and was selected from a synthetic library based on the heart fatty acid-binding protein by ribosome display [115]. The AviD-tag is another affinity tag based on Strep-tag displaying affinity for neutravidin, a neutral form of avidin rather for streptavidin [105]. The AviD-tag is composed by a 6-amino acid cyclic peptide that was selected through phage display technique with a dissociation constant of 12 μM for both Neutravidin and avidin [105].

Calmodulin is a small (17 kDa) and stable peptide employed as an affinity ligand on the purification of calmodulin binding peptide (CBP) fusion proteins. CBP is a 26 amino acid sequence derived from the carboxyl-terminal of rabbit skeletal muscle myosin light chain kinase [98, 116], displaying a dissociation constant of 10^{-9} M for calmodulin, calcium dependent [98, 116-117]. CBP has been fused on N-terminal of a wide range of recombinant proteins that were further produced with high levels of expression. The size and stability of calmodulin ligands allows the production of calmodulin affinity resins at affordable prices. In addition, protein elution can be carried out at mild conditions with calcium chelating agents such as ethylenediaminetetraacetic acid (EDTA) and ethylene glycol tetraacetic acid (EGTA) [98, 116-117].

Another example of tag-biological ligand includes biological ligand for tag-ligand pairs include the S peptide tag and S-protein derived from the enzymatic cleavage of ribonuclease A by the protease subtilisin [118]. The S-peptide, a fifteen amino acid sequence, is used as an affinity tag recognized by the S-protein immobilized on resin [99]. The major advantages are related with the small size of the affinity tag and the nanomolar dissociation affinity constant of the affinity pair [99].

Glutathione-S-transferase (GST) is a monomeric protein (26 kDa) originated from *Schistosoma japonicum* and belongs to a family of enzymes that catalyze the reaction between a nucleophile, reduced glutathione and electrophilic compounds [119-120]. The complementary GST tag acts both as a purification anchoring point and as solubility/stability enhancer [97], which justifies the popularity of this tag-ligand system [121-122]. In addition, GST resins are cost effective and the elution conditions are mild as reduced glutathione can be employed as a competitive agent [120]. However, if reducing conditions are not guaranteed, the fusion protein can undergo oxidative aggregation due to the existence of four cysteine residues exposed at the surface of GST tag [123]. Also, fusion proteins with higher molecular weights than 100 kDa can lead to partially or completely insoluble proteins [122].

The other class of biological ligands that recognize affinity ligands include carbohydrate based ligands, including cross-linked amylose, cellulose, and chitin, among others. The greatest advantages of these affinity ligands is that carbohydrates display high affinity for their respective affinity tags, the affinity matrices are inexpensive and prepared in a simple manner, which contributes for the ease of scaling-up the process.

The maltose-binding protein (MBP) is a 42 kDa periplasmic protein engaged with the transport of maltose and maltodextrins across the bacterial cytoplasmic membrane [124-125]. The MBP tag fused on the N-terminal of target proteins can be purified in one-single step with a cross-linked amylose affinity resin [126]. The MBP tag presents a high affinity for amylose, a maltose analogue, with a similar dissociation constant (10^{-6} M) to maltose, the competitive elution agent [124, 127]. This presents one of the greatest advantages of this affinity pair because the elution of the target fusion protein can be carried out under mild conditions with 10 mM of maltose [126]. In addition, the MBP tag is an extremely efficient solubilizing agent when compared to GST and thioredoxin (Trx) [128], and does not present any cysteine residue, thus not interfering with disulfide bonds of the target protein [10, 126].

The cellulose and chitin binding proteins present high affinity for the polysaccharides cellulose and chitin, respectively, which act as affinity ligands for purification purposes. Polysaccharides represent extremely interesting adsorbents for affinity based purification as they tend to be non-toxic, inert and stable [110, 112, 129]. Cellulose binding proteins can be found in a wide range of carbohydrases, display considerable dissociation constants in the low micromolar range for cellulosic supports, and vary on size from small (33-36 aa) to large domains with more than 180 amino acids [110]. For the particular application of fusion protein expression and purification, smaller domains are preferred. Cellulosic supports are commercially available in a wide range of geometries including fibers, beads, membranes and hydrogels, with a diverse extent of porosity, polymerization and functionalization. Chitin binding domains present high affinity for chitin, the second most abundant polymer after cellulose, and are part of the enzymes involved on the

biodegradation of chitin and chitosan [129]. Starch binding domains have also been explored as affinity tags, where these domains were fused to N-terminal of β -galactosidase which was then purified by using starch granules [112].

1.2.2. Structural ligands as binding partners of affinity tags

Structural affinity ligands, mainly cation and anion-exchange, metal chelate and hydrophobic ligands are extremely popular as binding partners of affinity tags and widely employed on the affinity purification of fusion proteins (Table 1.5). So far, synthetic ligands have been only tailor-made for the purification of specific biomolecules and not regarded as universal purification adsorbents for fusion proteins.

The more implemented are the metal chelate ligands complementary to the affinity tag His-tag (6xHis), where purification of the fusion protein is performed by immobilized metal affinity chromatography (IMAC) and has been comprehensively reviewed in the literature [15-17, 130-131]. The concept of IMAC was developed by Porath and co-workers in 1975, through the exploitation of the affinity between proteins and heavy metal ions (Zn^{2+} , Cu^{2+} , Ni^{2+}), in particularly zinc and copper metal which strongly adsorb polypeptides containing histidine and cysteine residues in aqueous solutions [132]. Based on this principle, the same authors coupled iminodiacetic acid (IDA) on to agarose to bind metal ions and further chelate proteins through histidine and cysteine residues [133]. Later, nitrilotriacetic acid (NTA), another highly stable quadridentate chelating adsorbent, was found to bind strongly to Ni^{2+} and Cu^{2+} and to further chelate adjacent histidine residues [134]. NTA technology was then reported for the production and purification of a recombinant protein. The hexapeptide 6xHis was genetically fused to the N-terminal of mouse dihydrofolate reductase protein, produced *in E. coli* and then purified by IMAC using a Ni^{2+} -NTA adsorbent. Subsequently, the fusion protein was eluted by different attempts such as a pH gradient (pH 8-5) in presence of denaturant agents (e.g. guanidine-HCl), and the tag was removed by carboxypeptidase A cleavage [134]. However, the major drawback on this pioneer purification study was the employment of denaturing conditions, which triggered the improvement of the procedure by conducting binding under native conditions and elution with low concentrations of competitor imidazole, or EDTA [135-136]. In addition to the advantages associated to structural ligands and to small affinity tags already discussed, the compatibility of His-tag with denaturing agents can be a benefit for recombinant proteins produced as inclusion bodies, as they require steps of solubilization with chaotropic agents prior binding to column [134]. The main disadvantages of IMAC adsorbents are metal leaching [137], that contribute for an end-product contamination and additional steps of purification with increased costs. In addition, IMAC adsorbents can bind to native histidine patches in proteins or to other amino acids (Glutamic and Aspartic acid, Tyrosine, Cysteine, Arginine, and Methionine)

resulting in a low selectivity. Other disadvantages are related with the evidences that the imidazolyl side chain groups of histidine can interfere with protein expression and folding and also mask the biological activity of the protein [74]. Nonetheless, the universe of application for the IMAC/His-tag system has surmounted protein purification and is extended to matrix-assisted refolding [138], detection and immobilization, protein microarrays [139-141], and to the development of other affinity tag technologies. The HAT tag is one of these examples and comprises a polyhistidine peptide sequence that can be fused on both ends of the target protein. The purification can be carried out under the same conditions as for His-tag, however the adsorbent used is based on carboxymethylaspartate (CMA) complexed with Co^{2+} , presenting high capacity [142]. Novel metal chelating agents based on 1,4,7-triazacyclononane (TACN) complexed with Ni^{2+} were developed in an attempt to overcome the limitations such as leaching, stability and lack of selectivity non-specific [143]. The respective affinity tags correspond to random heptapeptide sequences displaying multiple histidine, tryptophan, and/or tyrosine residues selected from a phage display library [143]. The selected peptide tag (HHHNSWD) was fused to N—terminal of Green Fluorescent Protein (GFP), produced in *E.coli* and then purified under physiological conditions with high salt concentration. Recovery was done with imidazole due to the presence of three histidine residues [143].

Ion-exchange chromatographic resins are also used as structural ligands with affinity for several tags, which comprise charged amino acids to promote electrostatic interactions with opposite charges on the matrices. The polyarginine-tag (polyArg) is an example that consists of six arginine residues and correspondent fusion proteins have been purified on cationic exchange resins. A great advantage on using cation-exchange chromatography for purification of recombinant proteins produced in *E.coli* is that most host proteins are neutral or acidic, which can contribute for the reduction of non-specific interactions as these proteins do not bind to cationic resins at physiological and alkaline pH [48, 74]. The first reported example was the production and two-step purification of human urogastrone fused to the polyArg tail, with protein eluted under a NaCl linear gradient at alkaline pH, with 44% yield and purity higher than 95% [144]. The polyArg-tag has also been successfully employed on refolding studies of a target protein upon immobilization, [145-146]. This affinity tag system shows low selectivity and promotes aggregation due to charge-charge interactions [8]. Poly-anionic affinity tails from one to eight glutamic residues were used in alternative to purify functional human growth hormone [147] and virus-like particles [148] produced in *E.coli*. Negatively charged affinity tags based on aspartates residues also yielded to stable fusion proteins with unaltered activity [149].

Table 1.5 Overview of the structural ligands involved as binding partners of affinity tags

	STRUCTURAL LIGANDS	TAG	TAG SIZE	TAG TYPE	REFERENCE
METAL CHELATE	Ni ²⁺ -NTA / Ni ²⁺ -IDA	Histidine residues (His tag)	6 aa	Purification	[132, 134]
	Co ²⁺ -CMA	HAT	18 aa		[142]
	Ni ²⁺ -TACN	peptide phage display library	7 aa		[143]
ION EXCHANGE (IEX)	Cation-exchange	Arginine residues	5-15 aa	Purification And	[144]
		Z _{basic}	7 kDa	Refolding	[49]
	Anion-exchange	Aspartic acid residues	5-16 aa	Purification	[149]
		Glutamic acid residues	1-8 aa	Purification	[148]
		Z _{acidic}	7 kDa	Purification	[50]
MULTIPURPOSE – IEX, IMAC AND HIC	Ni ²⁺ -NTA (IMAC) Phenyl agarose (HIC)	Multifunctional tag	6 aa	Purification	[150]

The Z domain, a variant of the B domain of SpA, has been engineered to create a highly charged Z domain - Z_{basic} - through the introduction of charged residues at specific positions, to allow interaction with negatively charged matrices at physiological conditions [48-49]. The Z domain was proved to be an excellent protein scaffold for engineering as it is soluble, does not display disulfide bonds and presents a reversible folding after exposure to chaotropic agents [49]. Due to its excellent properties, the Z_{basic} was explored for matrix assisted refolding of fusion proteins solubilised with chaotropic agents after being produced as inclusion bodies. This successful strategy is more advantageous over the common solubilization and refolding procedures based on dialysis, contributing for an effective, less expensive and time-consuming procedure. The Z domain was also on the basis of Z_{acidic}, engineered to have negatively charged and therefore to be purified by anion-exchange chromatography [50].

In order to combine several functionalities, a multipurpose peptide tag with the sequence HYDHYD that comprises a bi-repetition of the tripeptide constituted by a histidine, tyrosine and aspartate residues was developed [150]. Due to these functionalities, this peptide can be fused

on N-or C-terminal of the target proteins enabling the purification by a combination of several techniques such as IMAC, IEX, aqueous two phase system (ATPS) and hydrophobic interaction chromatography (HIC). This versatile affinity tag has proven to have multifunctional purification capabilities and was employed for the purification of GFP, lactate dehydrogenase and human haemoglobin [150].

Overall, the purification of recombinant proteins through affinity pairs “tag-receptor” have been extensively used and there is a broad range of commercially available purification supports, both biological (Table 1.6) and structural (Table 1.7).

Table 1.6 Examples of commercially available affinity tags and respective purification supports based on biological ligands

BIOLOGICAL LIGANDS	AFFINITY SOLID SUPPORT AND SUPPLIER	AFFINITY TAG	CAPACITY (MG/ML)
PEPTIDES AND PROTEINS	- Glutathione Sepharose (GE Healthcare) - Glutathione Superflow (Clontech) - Glutathione HiCap matrix (Qiagen) - Pierce Glutathione agarose / magnetic beads (Thermo Scientific) - Glutathione agarose and magnetic agarose beads (Sigma)	GST tag	10 10 20 40 / 5-10 5-10/ 15
	- Calmodulin resin (VWR) - MagneZoom™ – CAM (bioWorld) - Calmodulin resin (G Biosciences)	CMP	3 na 3
	- S protein Agarose (Novagen-EMD Millipore)	S-tag	na
	- StrepTactin™ (GE Healthcare) - Strep-Tactin Superflow / Magnetic Beads (Qiagen) - Strep-Tactin Superflow (Novagen- EMD Millipore)	Strep-tag II	6 9 / 0.3 3
	- IgG Agarose (GE Healthcare) - IgG Agarose (Sigma)	Protein A/G tag	2 1
	- Anti-FLAG M1, M2 resin (Sigma-Aldrich) - Anti-FLAG antibody agarose/ magnetic beads (Clontech)	Flag-tag	0.6 na/ 1.3
	- EZview™ red Anti-myc resin (Sigma) - Anti-c-myc agarose (Sigma-Aldrich) - Anti c-myc antibody agarose (Clontech) - Anti-Myc-tag mAb-Magnetic beads (MBL) - Pierce c-Myc-Tag IP (Thermo Scientific)	c-myc tag	0.05 0.2 0.5 na na
	- T7 Tag® antibody agarose (EMD Millipore) - T7 tag antibody agarose (Genetex) - Anti-T7 tag® antibody agarose (Abcam)	T7-tag	0.3 0.5 na
	- Pierce Anti-HA Agarose (Thermo Scientific) - Pierce Anti-HA Magnetic Beads (Thermo Scientific) - Ezview™ anti-HA affinity gel (Sigma) - Anti-HA Affinity Matrix (3F10) (Roche) - Anti-HA-tag mAb-Magnetic beads (MBL)	HA tag	150nmol/ml 10 µg/g 0.4 9 nmol/ml na
	- Softag immune affinity resin (Neoclone)	Softag 1/2/3	0.3

na- data not available

Table 1.6 Examples of commercially available affinity tags and respective purification supports based on biological ligands – Continuation

BIOLOGICAL LIGANDS	AFFINITY SOLID SUPPORT AND SUPPLIER	AFFINITY TAG	CAPACITY (MG/ML)
CARBOHYDRATES	- Amylose resin/ magnetic beads (New England Biolabs)	MBP	7-10 / 0.01
	- Cellulose (Sigma)	Cellulose Binding domain	15
	- Chitin agarose (New England Biolabs)	Chitin Binding domain	3

Table 1.7 Examples of commercially available affinity tags and respective purification supports based on structural ligands

STRUCTURAL LIGANDS	AFFINITY SOLID SUPPORT AND SUPPLIER	AFFINITY TAG	CAPACITY (MG/ML)
METAL CHELATE	- Ni-sepharose high performance (GE Healthcare)	His-tag	40
	- His Mag Sepharose (GE Healthcare)		50
	- TALON® Superflow™ (GE Healthcare)		20
	- HisPur Ni-NTA resin/ magnetic beads (Thermo Scientific)		20 / 0.5
	- HisPur Cobalt Superflow agarose (Thermo Scientific)		20
	- His60 Ni resin / magnetic beads (Clontech)		60 / n.a
	- TALLON® resin / magnetic beads (Clontech)		5-18/ n.a
	- Ni-NTA agarose (Invitrogen)		5-10
	- Dynabeads® His-Tag (Invitrogen)		0.04
	- NTA-Agarose / Magnetic Agarose Beads (Qiagen)		50 / 2
- Profinity™ IMAC Nickel Charged resin (BIORAD)	15		
- IMAC HyperCel™ (PALL)	na		
ION-EXCHANGE (IEX)	-TALLON® metal affinity resin (Clontech)	HAT tag	18
	-Anion or cation exchange resins(several sources)	Charged peptides	50-120

na- data not available

1.2.3. Affinity tags with versatile properties

Affinity tags that have been described so far are developed to be the anchoring point for purification by affinity chromatography, where the existence of a complementary affinity ligand is required. However, affinity tags alone can be auto-sufficient for enhanced fusion protein solubility and purification. Examples include the employment of hydrophobic tags based on tryptophan, phenylalanine and tyrosine which allow direct purification of the fusion protein through aqueous two-phase systems [151]; the Halo-tag, a 34 kDa protein derived from bacterial haloalkane dehalogenase that is able to yield pure proteins through the covalent binding to chloroalkane [152];

An alternative affinity system relies on the interaction between oil bodies and oleosin fusion proteins. Oil bodies are spherical structures that contain a triacylglycerol (TAG) matrix surrounded by a phospholipid mono-layer and several proteins, mainly oleosins, caleosins and steroleosins [153-155]. Oleosin being the main structural protein of the oil bodies's membrane will work as an affinity tag. The target fusion protein can be easily purified by using a flotation centrifugation step, and then the tag can be cleaved by enzymatic methods [154-155].

Elastin-like polypeptides (ELPs) are composed by repetitions of the pentapeptide VPGXG (where X is any amino acid, with the exception of proline) and present as an attractive property such as a thermally responsive reversible phase transition [156-158]. Below transition temperature (T_t), ELPs are highly water soluble whereas at temperatures higher than T_t , they aggregate and become insoluble [156-158]. Due to their unique properties, ELPs have been used as tags, and therefore successfully employed on purification of recombinant fusion proteins [158-159]. The purification of ELP fusion proteins can be carried out by several cycles of salt addition, heating and centrifugation until precipitation [160]. ELP technology eliminates the use of chromatographic resins, is a simple methodology with scale-up potential [158].

Self-cleaving tags have also been developed to possess inducible proteolytic activity representing an extremely interesting class of affinity tags [161-162]. The intein-based system is a well established example where the fusion protein cleavage can be induced by the presence of compounds such as β -mercaptoethanol or 1,4-dithiothreitol (DTT), or by a shift on pH or temperature [161-162]. Recently, a self-cleaving aggregation tag was constructed by combining the attractive properties of ELPs and inteins [160]. The major drawback of using this system is related with premature cleaving that can lead to a loss of the target protein and also the incompatibility with cleaving conditions [163].

The production recombinant proteins in bacterial hosts can lead to the formation of protein aggregates with unfolded or partially folded structures [164-165], denominated as inclusion

bodies (IBs) [166]. When these proteins are fused to solubility affinity tags, solubility is improved during protein expression, although the mechanism is not fully understood [10-11, 73]. Thioredoxin A (TrxA)[167], small ubiquitins-related modifier (SUMO) [168-169], and N-utilization substance A (NusA) [170-171] are examples of widely employed affinity tags with the purpose of increasing the solubility of the fusion partner rather than purifying it [10, 73]. As a consequence, a purification tag must be also added to the fusion partner to facilitate purification through affinity chromatography. The TrxA is a small (12 kDa) oxido-reductase protein with intrinsic thermal stability, solubility and robust folding properties that when fused to a target protein contributes for high yields of soluble protein [172-173]. The SUMO protein (11kDa) is involved in post-translational modification in eukaryotic cells, enhances solubility and stability of the binding partner when fused to the N-terminal [169], and allows simultaneous cleavage as it is recognized by the SUMO protease (*S. cerevisiae* Ulp1), at the conserved dipeptide glycine-glycine [168]. However, when used in eukaryotic hosts, the affinity tag can be cleaved due to the presence of SUMO proteases in vivo [173-174]. The NusA is a 55 kDa transcription elongation and antitermination factor, also known to increase the solubility of fusion proteins, however due to their large size can contribute for a high metabolic burn for the host cell [175].

Most affinity tags developed so far present different properties and the choice of selecting the best affinity pair is ambiguous and also dependent on the target protein and production system employed.

Novel dual-tagging methods have been developed and the concept of tandem affinity chromatography (TAP) has emerged as a generic platform for protein complex purification, combined with mass spectrometry that allows the isolation, identification and characterization of protein partners in multi-protein complexes [176-177].

The original C-terminal TAP tag was developed by *Puig et al* by using *Saccharomyces cerevisiae* hosts and consists of two IgG binding domains of SpA and a calmodulin binding peptide, separated by TEV protease cleavage site [177]. Also the N-terminal version was developed, being an inverse version of C-terminal as the SpA needs to be at N-or C-terminal end of the fusion protein to facilitate the purification process [177]. The main advantage related with this strategy is the highly efficient recovery of target protein [176-178]. Nowadays, the TAP strategy have been successfully employed in main application studies of protein-protein interaction studies in prokaryotic and eukaryotic cells, and different TAP tags also have been developed [176, 178-181].

1.3 TAG REMOVAL

For most applications, affinity tags are innocuous and do not interfere with the structure and biological function of the target protein. However, for therapeutic proteins there is a demand for tag cleavage to avoid immunological responses and to guarantee authenticity of the protein structure. The removal of the affinity tag can be performed by enzymatic cleavage or harsh chemical treatments and have been extensively reviewed [7, 74-76]. The chemical methods mostly used involve cyanogen bromide or hydroxylamine treatments, which are an inexpensive option but with low specificity and stringent conditions deleterious to the protein [8, 74]. Therefore, enzymatic methods are usually preferred as they operate under milder conditions and present higher selectivity as endoproteases recognize specific amino acid sequences or motifs [8, 74, 76]. The main endoproteases used on tag removal have been extensively reviewed and comprise enterokinase [182], tobacco etch virus (TEV) [183] Factor Xa [184], thrombin [184-185] and SUMO protease [168]. An attractive feature of enterokinase and factor Xa is that after cleavage there is no additional amino acid residues on the protein structure, contributing for an intact end product [74, 182, 184]. The main limitations in the use of proteases are related with the tag incompatibility under certain tag cleavage conditions (e.g. buffers, temperature), thus influencing its operation, and accessibility to the tag cleavage site [76]. The high costs associated with proteases limit the scalability of the purification process [161]. In order to overcome the higher costs, strategies on the reusability of these proteases without losing their efficiency have been studied, as recently reported for the immobilization of enterokinase on magnetic supports [186].

1.4 CONCLUDING REMARKS AND FUTURE TRENDS

Nowadays there is still a demand to develop processes for recombinant protein purification that can combine selectivity at affordable prices. Affinity chromatography is the most frequent choice for protein purification; however this technique is dependent on the adsorbents availability. Within the range of available affinity ligands, those of biological origin are still the preferred option due to the high selectivity. The use of structural ligands is dependent on the properties of the target protein and contaminants; however they could be extended to the purification of a wide range of proteins at low costs. Nevertheless, the main limitation of these ligands is the lack of specificity and the need to employ multi-steps purification protocols. Synthetic ligands raised to overcome the high costs associated with biological ligands have been developed to mimic naturally occurring interactions and are tailor-made to specific target proteins, fulfilling the lack of selectivity of structural ligands. However, most synthetic ligands developed so far have focused on high value proteins, as immunoglobulins. As each protein presents different biochemical and physical properties, a universal strategy for the purification of recombinant

proteins based on the use of affinity ligands tends to be difficult due to the fact that one ligand is required to purify one target protein, limiting the purification strategies. As a result, the use of affinity pairs “tag-receptor” emerged as an universal purification strategy.

Currently there is a wide range of affinity pairs “tag-receptor” that can be used for the production and purification of fusion proteins. Commercial purification supports possess either biological or structural ligands are presented in Table 1.6 and 1.7, respectively. The choice of the affinity tag should take into account the target protein, the host and production conditions, as well as the respective binding partner.

Therefore there is still a need to progress on the development of affinity pairs “tag-receptor” that can combine robustness, cost effectiveness and selectivity of the affinity ligand (receptor) and a peptide tag and also promotes the solubility and then the proper folding of the target protein. Therefore, there is a window of opportunity for the implementation of synthetic ligands *de novo* designed for affinity tags as viable alternatives for recombinant protein purification.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS**2.1.1 Chemicals**

All chemicals used were at least 98% pure and the solvents were pro-analysis.

1-Pyrene methylamine hydrochloride, 1,4 Diaminobutane, 3-Indoleacetic acid, 3-(4-Hydroxybenzoic acid), 4-Aminobenzamide, 4-Hydroxybenzoic acid, 5-Aminoisophthalic acid, 9-Anthracenecarboxylic acid, Arabic Acid, β -glutamic acid, β -Mercaptoethanol, Bichinchoninic Acid (BCA) kit, Citric acid, Coomassie Brilliant Blue R250, cyanuric chloride, DL-Dithiothreitol, Ethylenediaminetetraacetic acid (EDTA), epichlorohydrin, Glycerol 99%, Glutaric acid, isopropyl isocyanide, L-Arginine \geq 98.5%, L-tryptophanol, N_{α},N_{α} -Bis(carboxymethyl)-L-lysine hydrate, N-acetyl-D-Phenylalanine, N,N-dimethylformamide, Phenylacetic acid, Sodium Bicarbonate, Sodium Periodate, Sodium phosphate monobasic monohydrate, Sodium phosphate dibasic heptahydrate, Sodium thiosulfate, Succinamic acid, Triton-X100, Tryptamine, Urea were purchased from Sigma- Aldrich.

Ethanol absolute PA, Ethanol 96%, Ethylene glycol, Hydrochloric acid 37%, Sodium Chloride, Sodium-di-hydrogen Phosphate 1-hydrate, Sodium Hydroxide, (Di) Sodium-hydrogen Phosphate 2-hydrate were from Panreac.

β -D-1-thiogalactopyranoside (IPTG), Agar, Luria Broth, Agarose, Ampicillin, Glycine ultra pure for Molecular biology, NZYMiniprep kit, Tris(hydroxymethyl)aminomethane (Tris) Base ultra pure for Molecular Biology, SDS Micropellets (Sodium dodecyl sulphate) were acquired from NzyTech. The plasmid midi kit was purchased from Omega Bio-Tek

2-propanol, Acetone \geq 99.5%, Ammonium peroxydisulphate, Bromphenol blue sodium salt, Methanol, Tetramethylethylenediamine (TEMED) were obtained from Roth.

Ammonium hydroxide solution and Glycine 98% were obtained from Fluka and ACROS, respectively. Silver stain kit which includes fixative enhancer concentrate, silver complex solution, reduction moderator solution, image development reagent and development accelerator reagent, 30% acrylamide and bis-acrylamide solution 37:5:1 and Sodium dodecyl sulphate Solution 10% were purchased from Bio-Rad. The Coomassie Plus (Bradford) assay kit and polyvinylidene difluoride (PVDF) membranes were acquired from Thermo Scientific and Life Technologies Ltd., respectively.

2.1.2 Biochemicals

BSA protein (Albumin from bovine serum) min. 98% and recombinant green fluorescent protein rTurboGFP (FP552-Evrogen) were purchased from Sigma and Biocat GmbH respectively. The labelled peptides FITC-RWRWRW peptide (98.31%), FITC-RKRKRK peptide (98.56%), FITC-NNNNNN (98.48%), FITC-WFWFWF (98.51%) were obtained from Antagen Inc (USA).

The plasmids pET21c containing the DNA fragment respective for the hexapeptides WFWFWF, NWNWNW, RKRKRK, NNNNNN fused to GFP and non-tagged GFP were synthesized by Geneart (Invitrogen), and the pET21c was kindly provided by Prof. Alice Pereira (Requimte, Portugal). The restriction enzymes NheI and EcoRI were obtained from TAKARA. T4 DNA ligase, competent cells BL21 (DE3), NZY5 α and NZYSTAR, NZYDNA II and III ladders, low molecular weight protein marker were purchased from NZYTech. DNase I was obtained from Roche. The *E. coli* K-12 was kindly provided by Dr Pedro Vidinha and Diana Garcia.

The E-gel® CloneWell agarose gel, Novex® Sharp protein standard marker were obtained from Invitrogen. The fusion protein GFP - *B. megaterium* spore cortex lytic enzyme SleL was kindly provided by Dr Isik Ustok, University of Cambridge. The Rabbit PAb to GFP and Goat PAb to Rabbit IgG were acquired from Abcam. The Novex HRP Chemiluminescent substrate reagent kit was from Life Technologies, Paisley, UK

2.1.3 Chromatographic materials

Cross-linked agarose (Sephacrose™ CL-6B), PD-10 Columns – Sephadex™ G – 25M and HiScale16/20 columns were obtained from GE Healthcare. Captiva 96-well filtration block and empty columns (0.8 x 6.5 cm) were purchased from Varian. Brand Black immunograde 96-well microplates and 96-well UV half area (Greiner) was supplied from VWR International. The 96-well transparent microplates were acquired from Sarstedt.

2.2 INSTRUMENTATION

The synthesis of the combinatorial library of ligands based on the Ugi reaction was carried out in IKA KS 4000 ic control (VWR) incubator. The Fluorescence Microscope Olympus BX 51 with an objective U-RFL-T (40x amplification), U-MWB (λ_{exc} = 460-490 nm; λ_{em} = 515 -570 nm), an Olympus U-RFL-T lamp, an objective Uplam FLN, and Cell F software were employed for monitoring the formation of WFWFWF tagged GFP inclusion bodies. Image J was utilised for analysis of data obtained from fluorescence microscopy.

The measurement of the fluorescence intensity and the absorbance from BCA assay on a 96-well plate format were conducted on a Microplate Titre Infinite F200 with the respective Tecan filters (λ_{exc} = 485 – 505 nm; λ_{em} = 535 -560 nm and 560 nm, respectively).

The isolation of the plasmids pET21c and pAP001 was carried out in Sigma 3-18K centrifuge and the respective hydrolysis of the plasmids was performed on a GRANT dry block. The plasmid DNA was analyzed by spectrophotometric analysis on Thermo scientific Evolution 300UV-VIS and evaluated through gel agarose electrophoresis, by using a Horizon 11.14 electrophoresis chamber with GE EPS 301 power supply and Safe Imager™ Blue-Light Transilluminator from Invitrogen through KODAK 1D 3.6 software for visualization. The DNA fragment of pAP001 and double-digested pET21c were purified through E-gel® system using E-gel® CloneWell agarose gel on E-gel® iBase™ from Invitrogen. The ligation products resultant from the molecular cloning between DNA fragment of pAP001 and double-digested pET21c were carried out on a GRANT dry block. The sequencing analysis of the positive clones was performed by StabVida. The pH adjustment was performed with a pH meter from Hanna Instruments.

The expression of GFP-tagged proteins was conducted in OVAN incubator. Cells were collected with a Sigma 3-18K centrifuge, lysed with French press from Thermo Scientific, centrifuged with the same Sigma 3-18K and then ultracentrifuged in a Beckman *Optima* LE-80 (rotor 45TI). The Mini-Protean Tetra System from BIO-RAD was used for SDS-PAGE gels electrophoresis, and revealed with UVITEC Transilluminator through KODAK 1D 3.6 software.

The antimicrobial assays were carried out on Digital Thermo Mixer Vortemp 56, S2056-A, from Labnet. The static partition equilibration studies were conducted in a 96-well format in a Enduro™ MiniMix™ Nutating Mixe from Labnet.

The scale-up screening between the lead ligands and the respective GFP-tagged and non-tagged proteins was carried out in the automated system ÄKTA avant 25 from GE Healthcare. The western blot was carried out in a Semi Dry Transfer Cell unit from Biorad and revealed in X-ray film using a developer from GE, Amersham Hyperfilm ECL.

The structures of the lead ligands were drawn in ChemDraw Ultra 11.0 from CambridgeSoft. The isoelectric point and LogP properties of the affinity tags were estimated by Marvin Sketch 4.1.6 from ChemAxon. The GFP structure (PDB: 1QYO) was visualized with Pymol from DeLano Scientific LLC.

2.3 GENERAL METHODS

2.3.1 Solid phase combinatorial synthesis of the library of Ugi-based affinity ligands

The synthesis of the Ugi-based ligands begins with aldehyde functionalization of agarose beads. This functionalization encloses the following stages (i) epoxyactivation of SepharoseTM CL-6B, (ii) cis-diol activation and (iii) generation of aldehyde groups on the agarose beads (Fig. 2.1).

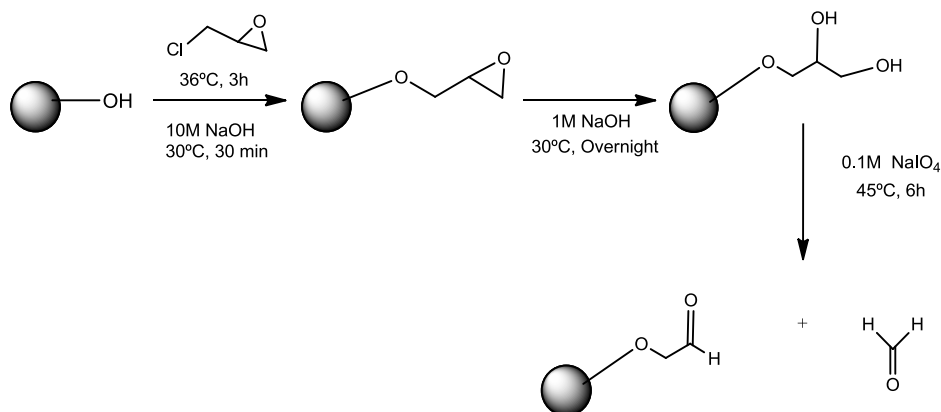


Figure 2.1 Aldehyde functionalization of agarose beads.

2.3.1.1 Epoxyactivation of agarose beads

The required amount of SepharoseTM CL-6B (100g) was washed with distilled water (10x volume of resin's weight) and then filtered under vacuum, being the last washing filtered under gravity. Then the resin was resuspended in 100 ml of distilled water and 4 ml of 10M NaOH (40 ml per 1 Kg of moist gel) were added and the slurry was incubated at 30°C during 30 min with orbital shaking at 200 rpm. Afterwards, 7.2 ml of epichlorohydrin (72 ml per 1kg of moist gel) was added to the moist gel and incubated for 3h at 36°C at 200 rpm. Agarose was then extensively washed with distilled water, and the first litter was discarded to halogens waste disposal.

2.3.1.1.1 Quantification of epoxyactivation extension

The extension of epoxyactivation was determined through a titration between OH⁻ groups released from epoxy ring opening and hydrochloric acid. Therefore, 3ml of 1.3M of Na₂S₂O₃ were added to 1g of epoxyactivated agarose and incubated for 20 min at room temperature. Released OH⁻ groups were further titrated with 0.1M HCl, until pH reached 7. The volume used corresponded to the number of moles of OH⁻ released (10 μmoles per 100 μl added). Typically, epoxyactivation of SepharoseTM CL-6B yielded 20 μmol of epoxy/g of gel.

2.3.1.2 Aldehyde functionalization of agarose beads

The cis-diol activation was carried out by adding 97 ml 1M NaOH (1 ml per gram of moist gel) to 97g of epoxyactivated moist gel and then incubated overnight at 30°C under orbital agitation at 200 rpm. Afterwards, the gel was washed as described in 2.3.1.1.

Aldehyde functionalization was performed by adding 97 ml of 0.1 M NaIO₄ (with the proportion of 1 ml of solvent / g of moist gel) and then incubated during 6h at 45°C at 200 rpm. After the functionalization the washing is performed as described in 2.3.1.1.

2.3.1.2.1 Qualitative evaluation of aldehyde functionalization

To 1g of aldehyde-activated agarose 1 ml of freshly prepared Tollens reagent was added. The presence of a silver mirror or a black precipitate revealed the presence of aldehyde. The positive control was performed with 1 ml of glutaraldehyde and the negative control with 1 g of Sepharose™ CL-6B unmodified. The Tollens reagent was freshly prepared by adding into a test tube, (previously cleaned with 3 M NaOH) 2 ml of 0.2 M AgNO₃ and a drop of 3 M NaOH. After this, the base 2.8 % NH₄OH was added dropwise with a constant shaking until almost all precipitate of silver oxides has dissolved. In order to remove all the precipitate, a 8.8% NH₄OH was added dropwise. The Tollens reagent was ready to use.

2.3.1.3 Synthesis of the solid phase combinatorial library

The preparation of the combinatorial library based on the Ugi reaction is described in Fig 2.2. The Ugi reaction is a multicomponent reaction that involves four different groups: an aldehyde, an amine, an isocyanide and a carboxylic acid.

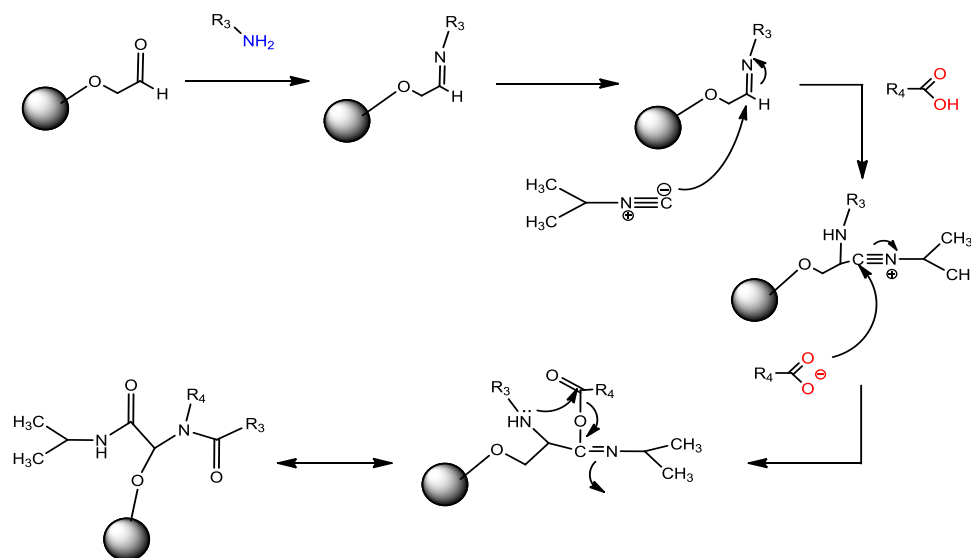


Figure 2.2 Reaction scheme of the multicomponent Ugi reaction synthesized in the solid phase, with isopropyl isocyanide being kept constant, whereas amine (R₃) and carboxylic acid compounds (R₄) are variable groups. The aldehyde component was attached to the solid phase.

In order to perform the Ugi reaction as shown in Fig. 2.2., the agarose gel functionalized with aldehyde was washed with 20% increments of methanol, from 0% (v/v) to 100 % (v/v) methanol using gravity without applying any vacuum. For the generation of the first combinatorial library comprising 56 affinity ligands, 16 g of the methanol-saturated resin was resuspended in 16 ml of 100% methanol (v/v). Prior the addition of the aldehyde beads in each well of the 96-well filtration block, 1-ml pipette tip was cut 4 mm so that the resin slurry was not entrapped in the tip, enhancing resin pipetting reproducibility. Therefore, 400 μ l of resin slurry aliquots (~250 mg of resin) were pipetted in each well of a total of 56 wells (8x7). At this point, the flexible end cap of the reaction block was removed in order to drain the solvent and allow resin settling in each well. After this, the end cap was again placed for the synthesis of the Ugi based ligands. The combinatorial library based on the Ugi reaction presents diversity within all the affinity ligands due to the variability introduced by the use of different amines (A3 to A10 columns) (Table 2.1 and Fig.2.3) and carboxylic acids (C1 to C7 rows) (Table 2.2). The isocyanide and aldehyde functionality was kept constant in all affinity ligands by using isopropyl isocyanide and the aldehyde groups functionalized in agarose beads.

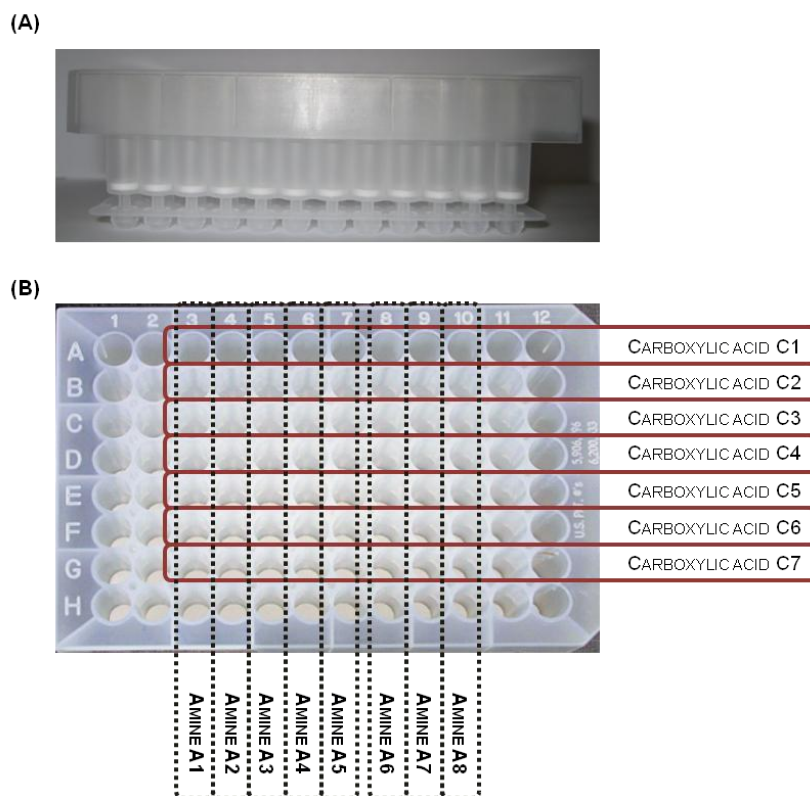


Figure 2.3 Visualization of the 96-well filtration block used for the synthesis of the combinatorial libraries based on the Ugi reaction and triazine scaffold (A) from the front and (B) from the top with the respective addition of the compounds used on the Ugi reaction. Regarding the triazine library, amines were employed in both rows and columns.

Each compound used presented a 5 molar excess of the epoxide-activated resin (§2.3.1.1).

The synthesis of the Ugi based ligands begins with the addition of the amine compound (Fig. 2.2), being the addition (250 μ l/well) done in each column of the reaction block from A3 (first column) until A10 (last column). All the amines were dissolved in 2.5 ml of 100% methanol (v/v), however for of amines A2, A3 and A7 that possess both amine and carboxylic acid functionalities, the carboxylic acids were also neutralized with 1M NaOH (370 μ l for A3 and A7, and 555 μ l for A2). The neutralization was carried out taking into account also the 5x molar excess and the respective number of COOH groups. The amine A4 was also neutralized with 180 μ l of 1M NaOH. The 96-well block was then sealed with the upper cap and left incubating for 2h at 60°C at 180 rpm, in order to guarantee the formation of the imine compound. The next compound added was isopropyl isocyanide (2.2 μ l per each well). Finally, the last compound (carboxylic acid) was added (250 μ l/well). The respective amounts of each carboxylic acid were dissolved in 2.5 ml of 100% methanol (v/v) and added according to the following order: C1 in first row and C7 in last row of the 96-well filtration block.

Table 2.1 List of the amines compounds used on the Ugi reaction for the solid phase combinatorial library

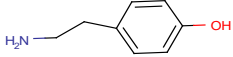
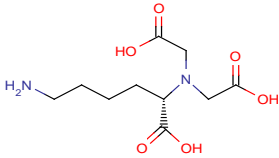
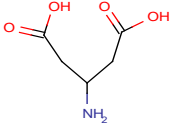
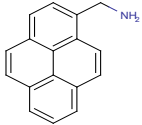
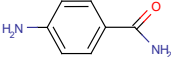
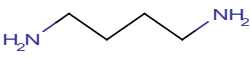
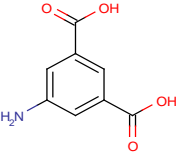
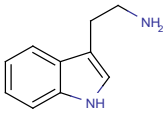
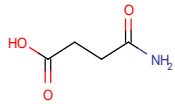
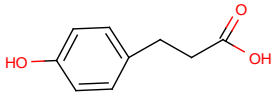
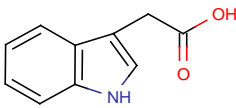
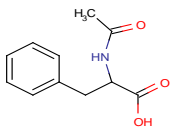
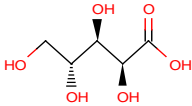
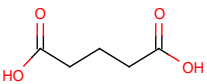
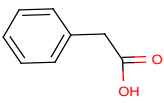
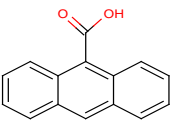
NUMBER	STRUCTURE	CHEMICAL NAME
A1		Tyramine
A2		N _α ,N _α -Bis(carboxymethyl)- L-lysine hydrate
A3		β - Glutamic Acid
A4		1-Pyrenemethylamine hydrochloride
A5		4 - Aminobenzamide
A6		1,4 - Diaminobutane
A7		5 - Aminoisophthalic acid
A8		Tryptamine

Table 2.2 List of the carboxylic acid and isocyanide compounds used on the Ugi reaction for the solid phase combinatorial library

NUMBER	STRUCTURE	CHEMICAL NAME
C1		Succinamic acid
C2		3-(4-Hydroxyphenyl) propionic acid
C3		3-Indoleacetic acid
C4		N-acetyl-D-phenylalanine
C5		Arabic acid
C6		Glutaric Acid
C7		Phenylacetic acid
C8		9-Anthracenecarboxylic acid

After the addition of all components, the upper cap was again fixed on the top of the block, and incubating for 48h at 60°C at 180 rpm

At the end of this reaction, the end cap was carefully removed to drain each well and then a washing procedure was performed to guarantee that all unreacted reagents were removed. Each well was washed with 1 ml of the following reagents: (1) 100 % (v/v) methanol (2) 50% DMF : 50 % methanol (v/v) (3) 50% (v/v) DMF (4) water (5) 0,1 M HCl (6) water (7) 0,2 M NaOH in 50 % (v/v) isopropanol (8) 2x water and (9) 20% (v/v) ethanol. The washed beads were stored with 250 µl of 20% (v/v) ethanol at 4°C.

Seven identical combinatorial libraries were synthesized to be further screened against the respective target: five different fusion proteins with different affinity tags (RWRWRW, WFWFWF, NWNWNW, RKRKRK and NNNNN) under physiological conditions, WFWFWF tagged GFP in presence of 8M urea and with the negative control, non-tagged GFP.

2.3.2 Production of crude extracts containing GFP-tagged and non-tagged GFP

2.3.2.1 Molecular Cloning of RWRWRW-GFP encoded gene in a pET-21c expression vector

2.3.2.1.1 Preparation of Luria Broth (LB) medium and LB agar plates with ampicillin preparation

The LB medium was prepared for the bacterial culture, where 25 g were added to 1 L of distilled water. The LB agar plates were prepared by dissolving 7.5 g of LB and 4.5 g agar (15 g for 1 L distilled water) in 500 mL of distilled water in a Schott flask. Then, the LB and LB agar were autoclaved at 120 °C for 20 min. Afterwards, the LB agar plates were cooled down without medium solidification at room temperature, for the addition of 500 µl of 100 µg/ml ampicillin under sterile conditions. The importance of cooling down the LB agar medium before antibiotic addition is relevant to avoid degradation of the antibiotic at high temperatures. Once the antibiotic was added, the medium was spread in the sterile Petri dishes (~20 ml of medium per Petri dish) also under sterile conditions. After the solidification of LB agar, the plated were kept at 4 °C. The LB liquid medium was kept at room temperature.

2.3.2.1.2 Transformation of the plasmids pAP001 and pET-21C in NZY5α competent cells

The plasmid pAP001 carries the gene that encodes RWRWRW tagged GFP and was synthesized by Genearth™ (Germany). The plasmid (5 µg) was resuspended in 10 µl autoclaved H₂O MiliQ. The transformation of the NZY5α competent cells with pAP001 was performed according with supplier instructions, NZYTech (Portugal). The NZY5α competent cells (60 µl) were thawed on ice. After addition of 2 µl of pAP001, the cells were incubated on ice during 30 min. After the cells were subjected to a heat-shock during 30 seconds in a 42 °C

water bath and then immediately placed on ice for 2 minutes. Subsequently, 940 μ l of LB medium were added to the cells, up to a final volume of 1 ml, at room temperature and were shaken at 210 rpm for 1 hour at 37 °C. Afterward, 50 μ l, 100 μ l and 200 μ l of the transformed cells with pAP001 were spread on LB agar plates containing 100 μ g/ml ampicilin. Then to concentrate the cells, the remaining volume (700 μ l) of cell culture was centrifuged for 3 min at 1850xg, and 600 μ l of the supernatant medium was discarded. The concentrated cells were resuspended in the remaining volume and spread on the LB agar plates. The plates were incubated overnight at 37 °C. Negative and positive controls were performed without adding any plasmid and using 1 μ l of pUC19 plasmid, respectively, and 20 μ l of NZY5 α competence cells. This procedure was also followed for the transformation of the NZY5 α competent cells with pET-21c plasmid, kindly provided by Prof. Alice Pereira (Requimte, Portugal).

2.3.2.1.3 Isolation and purification of plasmids DNA pAP001 and pET-21c

The isolation and purification of plasmid DNA from pAP001 and pET-21c were carried out using a standard procedure from plasmid midi kit according with the supplier instructions (Omega Bio-Tek). Five pre-inoculum test tubes were prepared: 2x pAP001, 2x pET-21c and a negative control. Each tubes containing 6 μ l of 100 μ g/ml ampicilin, 6 ml of LB medium and one single isolated colony from the respective freshly transformed plate (§2.3.2.1.2). In the negative control test tube was not added any colony. Then all the test tubes were incubated overnight at 37 °C at 210 rpm. After the incubation, 2 ml of each pre-inoculum was used to inoculate 300 ml of LB medium with 300 μ l of 100 μ g/ml ampicilin. This was repeated twice for pAP001 and pET-21c. Then, all flasks were incubated overnight at 37 °C with 210 rpm. The culture was split in Falcon tubes of 50 ml and centrifuged during 10 minutes at 11000xg until the total volume has been used. After centrifugation, the supernatant was removed and the cell pellet was kept for plasmid DNA isolation using the plasmid midi kit (Omega Bio-Tek). The procedure and solution volumes were maintained equal as described in the kit instructions for the high-copy plasmid pAP001. The isolation of plasmid DNA from low-copy number plasmids requires double of cells and volumes of lysis buffers for a more effective isolation. The plasmid pET-21c is considered as low-copy plasmid [187]. This kit simplifies the extraction and purification of nucleic acids through the binding DNA or RNA into a HiBind® matrix under certain optimal conditions. After cell harvest, 2.25 ml of solution I/RNase A was added to bacterial cell pellet and the cells were resuspended by vortexing. After, the cell suspensions were transferred to 50 ml centrifuge tubes, and 2.25 ml of solution II was added, and the centrifuge tubes were covered and mixed gently by inverting and rotating the tube 8-10 times to obtain a cleared lysate. The tubes were then incubated at room temperature with occasional mixing for 2-3 minutes. Subsequently, 3 ml of solution III was added into the tubes, and then gently mixed until a flocculent white precipitation has formed. The tubes were incubated at room temperature for 5 minutes and afterwards centrifuged at 11000xg for 25 minutes at room temperature, and the cleared

supernatant was kept. Meanwhile, a HiBind® DNA Midi column was pre-inserted in a 15 ml collection tube and 1 ml of equilibration buffer was added into the column and let sit for 4 minutes at room temperature to equilibrate the membrane. The columns were centrifuged at 11000xg for 3 minutes. Then, the cleared supernatant was added by carefully aspirated it into the HiBind® DNA Midi column and after centrifuged at 11000xg for 5 minutes until all the supernatant has pass through the column. The flow-through was discarded. Subsequently, 3 ml of Buffer HB was added to the Midi column and centrifuged as above. The flow-through was again discarded, and 3.5 ml of DNA wash buffer (previously warmed at room temperature) and then the columns were centrifuged at 11000xg for 5 minutes. The flow-through was again discarded and the procedure was repeated. The empty column was centrifuged at 11000xg for 15 minutes to dry the column matrix and remove the ethanol from the column. The columns were placed into a clean 15 ml centrifuge tube. Then, the elution step was carried out by adding 500 µl of MiliQ water (pre-heated at 65 °C) onto the column matrix and let it sit at room temperature for 2-3 minutes. Then, the columns were centrifuged at 11000xg for 5 minutes. A second elution step with 600 µl MiliQ water was performed with the same conditions as first elution. Finally, the plasmid DNA concentration was determined.

2.3.2.1.4 Spectrophotometric quantification of plasmid DNA pAP001 and pET-21c

The spectrophotometric analysis allows the quantification of plasmid DNA and evaluation of its purity. The eluted fractions (pAP001 and pET-21c) obtained from §2.3.2.1.3 were diluted 1:100 (10 µl of sample dissolved in 990 µl of MiliQ water) and then a spectrum scan between 220 and 280 nm was performed in a 1-cm spectrophotometric cuvette. The determination of DNA concentration is related with the optical density (OD) at 260 nm, where for 1-cm pathlength, the optical density at 260 nm equals to 1.0 for 50 µg/ml of double stranded DNA (dsDNA) solution [188]. This relation is given by the following equation:

$$[\text{dsDNA}] = 50 \mu\text{g/ml} \times \text{OD}_{260 \text{ nm}} \times \text{Dilution factor} \quad [1]$$

The DNA purity is given by the ratio between $\text{OD}_{260 \text{ nm}}$ and $\text{OD}_{280 \text{ nm}}$ and for a pure sample of DNA, this ratio should be ~1.8 [188].

2.3.2.1.5 Evaluation of plasmid DNA through agarose gel electrophoresis

The eluted fractions of plasmids DNA pAP001 and pET-21c were analysed by agarose gel electrophoresis in order to evaluate the three possible conformations of plasmid DNA: superhelical circular (form I), nicked circular (form II) and linear (form III). These three forms have different electrophoretic mobilities, in which form I DNA migrates faster than form III DNA, however the form III is a result from DNA manipulation and therefore should not appear on the

gel [189]. A 0.8% agarose gel was prepared by adding 0.48 g of agarose to 60 ml of TAE 1x buffer pH 8 (buffer denominated Tris-acetate-EDTA composed by 40mM tris(hydroxymethyl)aminomethane base (Tris), 20mM glacial acetic acid and 1Mm ethylenediaminetetraacetic acid (EDTA) and then heated 2 minutes under microwave until complete dissolution of agarose. Afterwards, the solution was poured in the proper container at room temperature for the complete solidification of agarose into a gel and with the proper comb with the desired number of wells. The samples preparation begins with the addition of 2 μ l of each eluted fraction of both pAP001 and pET-21c in different microtubes and 5 μ l of the sample buffer blue juice (buffer composed by 65% (w/v) sucrose, 10 mM Tris-HCl (pH 7.5), 10 mM EDTA, and 0.3% (w/v) bromophenol blue) and then well homogeneized. The same treatment was performed with the DNA marker (NZYDNA ladder II). Afterwards the samples were spin down and added to each well of the agarose gel. Once the samples were pipetted in each well, the agarose gel was run at 100 V during 60 minutes. Then, the agarose gel was staining during 30 minutes within a container with SyBr® Safe (Invitrogen). A photograph of the gel was taken in an UV transilluminator by using the KODAK 1D 3.6 software.

2.3.2.1.6 Hydrolysis of plasmids DNA pAP001 and pET-21c with restriction enzymes *Nhe* I and *EcoR* I

Plasmids DNA pAP001 and pET-21c were subjected to sequential hydrolysis with two different restriction enzymes *Nhe* I and *EcoR* I. The enzymatic hydrolysis of plasmid pAP001 released a fragment with 843 bp that corresponds to RWRWRW tagged GFP encoding gene. This gene was used for further directional cloning into plasmid pET-21c that becomes linear (5443 bp) after hydrolysis with the same restriction enzymes. The double sequential hydrolysis with the two restriction enzymes should be carried out in a buffer where both enzymes are close to optimal activity. According with TAKARA supplier instructions, the buffer in which *Nhe* I enzyme presents 100% of relative activity is buffer M (100 mM Tris-HCl pH 7.5, 100 mM MgCl₂, 10 mM Dithiothreitol (DTT) and 500 mM NaCl), while for *EcoR* I is buffer H (500 mM Tris-HCl pH 7.5, 100 mM MgCl₂, 10 mM DTT and 1 M NaCl). The *EcoR* I also presents an optimal activity with buffer M; however in this buffer the enzyme may display star activity, according with supplier instructions. In order to improve the compatibility between the reactional conditions of both enzymes activity, the buffer M was used and after complete digestion the enzymes were further inactivated. Two different microcentrifuge tubes were prepared, one for pAP001 and other for pET-21c. In case of pAP001, the hydrolysis with *Nhe* I was carried out by assemble in a microcentrifuge tube 13 μ g substrate DNA (stock solution 425 ng/ μ l), 5 μ l 10x restriction enzyme buffer M, 2 μ l *Nhe* I (10 units/ μ l, Takara) and 13 μ l sterilized distilled water for a final volume of 50 μ l. The restriction enzyme was the last component added to the mixture and its volume should not exceed 10% of the reactional volume to avoid high glycerol concentrations. Then the reactional mixture was incubated at 37 °C during 2 hours. The *Nhe* I hydrolysis with

pET-21c was carried out with the same reactional conditions as pAP001, however 43 μl of substrate DNA (240 ng/ μl) was used and no sterilized was added for the final volume of 50 μl . After the first hydrolysis, an aliquot of 5 μl of both reactions was taken for further agarose gel electrophoresis. The restriction enzyme was then inactivated for 15 minutes at 70 °C. Afterwards, 2 μl of *EcoR* I (12 units/ μl , Takara) was added to each one of the reactional mixtures and then incubated at 37 °C for 1:30 hours. Once the hydrolysis was finished, 5 μl -aliquots of both reactions were collected and then the enzyme was also inactivated by incubation at 60 °C for 15 minutes.

2.3.2.1.7 Hydrolysis evaluation by agarose gel electrophoresis

The success of sequential hydrolysis was analysed through agarose gel electrophoresis. The 0.8% agarose gel was prepared as described in § 2.3.2.1.5. The samples taken in the first hydrolysis corresponds to the digested DNA and those obtained from the second hydrolysis corresponds to double digest DNA. To each 2 μl of each sample 5 μl of the sample buffer blue juice was added, and then well homogeneized. Samples (2 μl) of plasmid DNA pAP001 and pET-21c without any enzymatic treatment were also analysed as controls. The same treatment was performed with 5 μl of DNA marker (NZYDNA ladder III). The gel was run and stained according with § 2.3.2.1.5. The double digested pET-21c and the insert released from the sequential hydrolysis of pAP001 were then purified by E-gel® CloneWell agarose gels (Invitrogen) for further cloning.

2.3.2.1.8 Purification of the insert fragment from pAP001 and double digested pET-21c using the E-gel system

At the end of sequential double hydrolysis the final concentration of pAP001 and pET21c was 255 ng/ μl and 206 ng/ μl , respectively. In order to proceed with fragments purification through E-gel® CloneWell agarose gel, the gel was placed in E-gel® iBase™ and then a pre-run was performed with the comb in place by selecting the program PRE-RUN for 2 minutes. At the end of the run, the combs were removed and the samples were loaded. According with supplier instructions, the loading sample should present 20-400 ng per band in a total of 50-700 ng in a total sample volume of 20-25 μl per well. Regarding the hydrolysis samples, 10 μl of pAP001 double digested was added in each well of four total wells, and then diluted with 10 μl of sterilized water. 10 μl of the DNA marker (NZYDNA ladder III) was added in the following well. Then, 10 μl of the double digested pET-21c was added in the following four well, and then 10 μl of sterilized water was added to each one of the wells. In the remaining empty wells, 25 μl of sterilized water was applied in the each one of the wells. Then E-E-gel® iBase™ Power System was placed over a blue light transilluminator and a run CloneWell program was selected according with the size of band that was purified. According with the size, there is an optimal

time that the band takes to reach the reference line. The fragment released after the sequential hydrolysis of pAP001 corresponds to 843 bp and the double digested pET-21c presented 5443 bp. This corresponds to 17-21 and 32-36 minutes, respectively. After, the time was set, the gel was monitorized through E-Gel® Safe Imager™ Real-Time Transilluminator. Once the band of interest reached the reference line, 25 µl of sterile water was added to in each empty well because some pre-filled water was lost during the run. From the reference line until the collection well, the sample took 2 and 3 minutes for fragment released from pAP001 and double digested pET-21c, respectively. Then, after the band of interest migrated into the well, the DNA was collected by pipetting. The samples were further analysed through agarose gel electrophoresis § 2.3.2.1.5. The DNA concentration was then estimated by comparing the fluorescence intensity of the bands obtained with the respective bands of the DNA marker. The fluorescence intensity of the marker bands corresponds to a specific amount of DNA (ng). The fluorescence intensity of the bands of the fragments of interest and from the marker was further evaluated by ImageJ software. Then by comparison it was possible to estimate the final concentration of the obtained fragments.

2.3.2.1.9 Ligation of gene encoding RWRWRW tagged GFP (insert) and double digested pET-21c with T4 DNA ligase

The ligation reaction with T4 DNA ligase was carried out according with supplier instructions (NZYTech). The reaction was performed by assembling in a microcentrifuge tube, 3.8 µl of double digested plasmid pET-21c (8 ng/µl), 13.2 µl of 3x excess of insert (1.03 ng/µl), 2 µl of 10x Reaction Buffer (660 mM Tris-HCl, pH 7.6, 66 mM MgCl₂, 100 mM DTT and 660 µM ATP), 1 µl T4 DNA ligase (5 Units/µl, NZYtech) and autoclaved Milli-Q water up to final volume of 20 µl. The T4 DNA ligase was the last component added. The amount of insert was determined according with the following equation, given by suppliers:

$$\text{insert (ng)} = \frac{\text{ng of vector} \times \text{kb size insert}}{\text{kb size vector}} \times \text{molar ratio} \frac{\text{insert}}{\text{vector}}$$

Then the reactional mixture was incubated overnight at 16 °C.

2.3.2.1.10 Transformation of the ligation reaction product *in* NZYstar cells

The ligation product was used to transform NZYstar cells according with procedure previously described in § 2.3.2.1.2 with the difference that the LB/ agar plates also contained 100 mg/ml of X-gal (5-bromo-4-chloro-indolyl-β-D-galactopyranoside) and 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). A negative control was also performed, without transforming cells

with any ligation product or plasmid. After the overnight incubation at 37 °C, twelve colonies were randomly selected and plasmid DNA was isolated.

2.3.2.1.11 Isolation of plasmid DNA

Twelve test tubes were prepared with 6 µl of 100 µg/ml ampicillin, 6 ml of LB medium and one single isolated colony from the respective freshly transformed plate (§2.3.2.1.10). In the negative control test tube no colony was added. Then all test tubes were incubated overnight at 37 °C with 210 rpm.

The isolation and purification of plasmid DNA was performed with NZYMiniprep kit according with the supplier instructions, NZYTech (Portugal). The procedure and solution volumes were maintained equal as described in the kit instructions for the high-copy plasmid pAP001. The isolation of plasmid DNA from low-copy number plasmids requires double of cells and volumes of lysis buffers for a more effective isolation. This kit also allows a highly recovery of plasmid DNA through the selective adsorption of the plasmid DNA into a silica gel-based NZYTech plasmid spin column. This procedure involves sequential stages such as (1) bacterial cell harvest, (2) cell lysis, (3) clarification of lysate, (4) DNA binding, (5) silica membrane washing and drying and (6) elution of highly pure DNA. The procedure begins with the cell harvest after overnight growth, by centrifugation for 2 minutes at 7400xg in separate tubes depending on the plasmid used, until the end of total cells suspension. The supernatant was discarded in each step of centrifugation. Afterwards, already in cell lysis stage, the bacterial cell pellet obtained was resuspended with 250 µl of buffer A1 (RNase A) by vigorous vortexing. 250 µl of buffer A2 was added, mixed gently by inverting the tube several times and incubated at room temperature for 4 minutes. After incubation, 300 µl of buffer A3 was added and mixed softly by inverting the tubes several times. At the end of this stage, the tubes were centrifuged for 10 minutes at 7400xg. The supernatant was then loaded onto a NZYTech column already placed in a 2-ml collection tube for the plasmid DNA binding. The column was centrifuged for 2 minutes at 7400xg and the flow-through was discarded. The silica membrane was then washed by adding 250 µl of buffer AY, pre warmed, onto the column and afterwards centrifuged for 2 minutes at 7400xg. The resultant flow-through was discarded. As a last step of washing, 300 µl of buffer A4 (with ethanol already added) was added onto the column and once again centrifuged for 2 minutes at 7400xg. The flow-through was once again discarded. In order to proceed with elution of plasmid DNA, the NZYTech column was re-inserted into an empty 2-ml collecting tube and centrifuged for 2 minutes at 11000 rpm. The dried NZYTech column was placed into a clean 1.5 ml centrifuge tube for the first elution step. This step was carried out by adding 30 µl of MilliQ water, pre-warmed at 65 °C, into the centre part of the tube. The column was then incubated 1 minute in a 37 °C water bath, and centrifuged for 2 minutes at room temperature. The first fraction of the eluted plasmid DNA was then kept at 4°C. For the second elution step, the procedure was repeated as in the first elution step however 50 µl of MilliQ water added. The

second eluted fraction was also kept at 4 °C. After elution, the plasmid DNA isolated from each colony was subjected to sequential hydrolysis for the evaluation of success of cloning procedure. After this, the eluted plasmid DNA fractions were kept at -20 °C for further use.

2.3.2.1.12 Confirmation of ligation reaction by agarose gel electrophoresis and DNA sequencing

The success of the cloning procedure was validated by the presence of product obtained from the sequential hydrolysis and evaluated by agarose gel electrophoresis. The sequential hydrolysis with *Nhe* I and *Eco*R I following the strategy described in § 2.3.2.1.6. However in this case, the total volume of hydrolysis was 20 µl, with 10 µl of the pDNA, 2 µl 10x restriction enzyme buffer M, 7 µl of autoclaved Mill-Q water and 1 µl of *Nhe* I. After incubation and enzyme inactivation, 1 µl of *Eco*R I was added to reactional mixture, and after hydrolysis, the enzyme was inactivated and then the sequential hydrolysis product was evaluated by agarose gel electrophoresis. The agarose gel electrophoresis was performed according with §2.3.2.1.5. The wells that present only the presence of double digested DNA and the insert fragment without star activity or any unexpected fragment were selected for further DNA sequencing by StabVida (Portugal).

2.3.2.2 *RWRWRW* tagged GFP small scale expression studies

A positive clone, plasmid pET-21c harbouring the gene that encodes *RWRWRW* tagged GFP, was selected for small scale expression studies. These studies involved the transformation of cells suitable for protein expression such as *E. coli* BL21(DE3) with the positive clone. Afterwards, the effect of inducer concentration at different OD_{600 nm} on the amount of *RWRWRW* tagged GFP produced was studied by monitorization of GFP fluorescence and SDS-PAGE analysis.

2.3.2.2.1 Transformation of *E. coli* BL21(DE3) competent cells with the positive clone

E. coli competent BL21(DE3) cells were transformed with the positive clone as described in § 2.3.2.1.2. A negative and positive controls were performed without adding any plasmid and using 1 µl of pUC19 plasmid, respectively, and 20 µl of BL21(DE3) competent cells.

2.3.2.2.2 Small scale expression tests

The small scale expression studies were divided in two stages: (i) *RWRWRW* tagged GFP expression at OD_{600 nm} 0.6-0.8 with 0.2, 0.5, 1, 1.5 and 2 mM IPTG and (ii) the protein expression using a constant IPTG concentration at higher OD_{600 nm} ~2. In both cases, a pre-inoculum was prepared by adding 6 ml LB, 6 µl of 100 µg/ml ampicilin and 1 colony of the

transformants (§ 2.3.2.2.1) in a sterile test tube. The culture was incubated overnight at 37 °C with orbital shaking of 220 rpm. In the first studies, 100 ml LB, 100 µl of 100 µg/ml ampicilin and 1 ml of pre-inoculum were added in 250 ml shaking flasks and incubated at 37 °C at 220 rpm. The culture growth was monitorized by optical density measurements, and once reached an $OD_{600\text{ nm}}$ 0.6-0.8, the protein expression was induced with IPTG. The culture was kept at 37 °C with 220 rpm overnight. Each shaking flask corresponded to a different IPTG concentration. After IPTG induction, the protein expression was monitored by GFP fluorescence, optical density measurements and SDS-PAGE analysis of the aliquots taken every two hours until 6 h of induction and then after overnight protein expression (22 hours). The second stage studies began by assembling 50 ml LB medium, 50 µl of 100 µg/ml ampicilin and 1 ml of pre-inoculum in a 250 ml shaking flask and then incubated at 37 °C with 220 rpm. Once again, the culture growth was measured by $OD_{600\text{ nm}}$ and once reached a value ~2, the culture suspension was ready for induction. Prior induction, 50 ml of fresh LB medium and 50 µl of 100 µg/ml ampicilin were previously added to the culture suspension and then the IPTG was subsequently added. Also in this case, the protein expression was monitorized as in the first stage of studies

2.3.2.2.3 Evaluation of the amount of RWRWRW tagged GFP produced by SDS-PAGE Analysis

Sample preparation

In order to evaluate the amount of protein expression time course, sample volume was normalized. The normalization was performed according with the ratio between 1.2 and the respective optical density measurement of each sample. The respective sample volumes were pipetted in different microcentrifuge tubes, and then centrifuged at 6800xg for 5 minutes. The supernatant was discarded and the pellet was resuspended in 50 µl of sample buffer. The sample buffer is composed by 5 ml of 0.5 M Tris-HCl pH 6.8, 2 ml of 100% glycerol, 4 mg of blue bromophenol, 8 ml of 10% Sodium dodecyl sulfate (SDS), 1 ml of β-mercaptoethanol and distilled water up to a final volume of 20 ml. Then, the samples were spin down, and boiled for 2 minutes. The 5 µl of low molecular weight protein marker (NZYTech) was also subjected to the same treatment as the samples. After this, the samples were ready to be applied in each well of the 12.5% acrylamide gel.

12.5% acrylamide gel preparation and blue-coomassie staining

The acrylamide gel is composed by the stacking gel and running gel. The 12.5% acrylamide running gel was prepared by assemble in a tube 0.75 ml solution I (3M Tris Base pH 8.8), 2.08 ml of 30% acrylamide and bis-acrylamide solution 19:1, 0.05 ml 10% SDS, 2.1 ml distilled water, 38 µl of 10% ammonia persulfate and 2.5 µl of 99% tetramethylethylenediamine (TEMED). Then the mixture was transferred to the glass plates of the casting frame. Afterwards, ~1 ml of

2-butanol 99% was added on the top of the gel solution to promote the formation of a homogeneous surface and consequently the gel was polymerized for 30 minutes. After this, the 2-butanol solution was completely removed and the running gel was washed with distilled water. The 5% stacking gel was then prepared by adding 0.450 ml solution II (0.5M Tris Base, pH 7.0), 0.3 ml of 30% acrylamide and bis-acrylamide solution 19:1, 0.018 ml 10% SDS, 0.94 ml distilled water, 13.5 μ l 10% sodium persulfate and 2 μ l of 99% TEMED in a tube. Then, the mixture was added on the top of the glass plates and the comb with wells was inserted. Then the gel was polymerized for other 30 minutes. Afterwards, the electrophoresis buffer (0.25 M Tris Base, 1.92M Glycine, 0.1% SDS pH 8.3) was added to the tank, the glass plates containing the polymerized gel were introduced in the running module and 15 μ l of sample were applied on each well. The 12.5% acrylamide gel ran at 150 V, 250 mA for 1 hour, and then transferred into the staining solution (1 g Coomassie blue R-250, 15 ml glacial acetic acid, 90 ml of methanol and distilled water up to 200 ml) for 30 minutes. Then, the gel was overnight distained with the distaining solution (75 ml glacial acetic acid, 450 ml of methanol and distilled water up to 1 L).

2.3.2.2.4 Evaluation of the amount of RWRWRW tagged GFP produced by GFP fluorescence

The fluorescence of the RWRWRW tagged GFP of each aliquot of a respective induction time was measured in a microplate format by adding 200 μ l of each sample in each well of the black 96-well plates (VWR, Portugal). The fluorescence was measured by using an $\lambda_{\text{excitation}} = 485$ nm and $\lambda_{\text{emission}} = 535$ nm) in microplate reader (Tecan Infinite F200).

2.3.2.2.3 Large scale expression of RWRWRW tagged GFP

According with small scale studies, the optimal condition for the production of RWRWRW tagged GFP was using 1 mM IPTG for the protein production at $OD_{600 \text{ nm}} 0.6-0.8$ during 22 h at 37 °C with orbital shaking of 220 rpm. Therefore, the large scale expression was conducted with these optimal conditions and for this was required the performance of two inocula. The first pre-inoculum was performed in a test tube by adding 6 ml of LB medium, 6 μ l of 100 ug/ml ampicilin and 1 colony from the transformed BL21(DE3) cells (§ 2.3.2.2.1) and then incubated during 6 hours at 37 °C at 220 rpm. Afterwards, the inoculum was performed in a 100 ml shaking flask, where 100 ml LB medium, 100 μ l of 100 ug/ml of ampicilin and 1 ml of pre-inoculum were added and overnight incubated at 37 °C with 220 rpm. After this, the large scale expression begins through the addition of 1 L of LB medium, 1 ml of 100 μ g/ml of ampicilin and 10 ml of inoculum in a 2 L shaking flask and incubation at 37 °C with orbital shaking of 220 rpm. The cellular growth was monitored by optical density measurements and once reached an $OD_{600 \text{ nm}} 0.6-0.8$, the protein expression was induced. During protein expression, 5 ml aliquots of the 2, 4, 6 and 22 hours of induction were taken and evaluated by SDS-PAGE (§ 2.3.2.2.3)and fluorimetric analysis according with described in § 2.3.2.2.4.

2.3.2.3.1 Cellular fractionation

In order to obtain the soluble crude extract containing RWRWRW tagged GFP from the large scale protein production, the cells were lysed and then subjected to multiple steps of centrifugation and a final dialysis with the binding buffer required for further screening assays. The amount of GFP-tagged and total protein were quantified through GFP fluorescence and colorimetric Bicinchoninic acid (BCA) assay, respectively.

At the end of protein expression, the cell culture was centrifuged for 25 minutes at 4200xg at 4 °C. The supernatant was discarded and the pellet was resuspended with 20 ml of Buffer A (10 mM Tris-HCl pH 7.6) and kept in ice. The resuspended cells were subjected to three freeze/thaw cycles in order to fragilize the cells to improve the cell lysis efficiency. The cell lysis was carried out by a mechanical process (French Press), by passing four times the sample 800psi. After this process, DNase I was added to the lysate during 15 minutes, in order to reduce viscosity of the sample. Afterwards, the lysate was centrifuged for 15 minutes at 4 °C. The resultant pellet was resuspended in 20 ml of Buffer A and the supernatant was further ultracentrifuged at 180000xg at 4 °C for 1 h. Then, the pellet obtained was also resuspended with 20 ml of Buffer A. The supernatant was then dialysed in the buffer required for further screening assays, PBS (10 mM sodium phosphate, 150 mM NaCl, pH 7.4). The supernatant was introduced in a dialysis membrane (MWCO of 10 kDa) and dialysed against 5 L of PBS during 6 hours at 4 °C with gentle magnetic stirring. Then, a second dialysis was set overnight with the same operational conditions. After this, the soluble crude extract was kept at -20 °C. During the fractionation, aliquots of each step were taken and evaluated through SDS-PAGE (§2.3.2.2.3) and fluorimetric analysis (§ 2.3.2.2.4). However, in this case the aliquots volumes used in SDS-PAGE analysis were normalized according to the final volume of the sample obtained after each step. Afterwards, 5 µl of the sample buffer was added to the normalized aliquots volume from each step of fractionation and then the total volume was pipetted in each well of the 12.5% acrylamide gel.

2.3.2.3.2 Quantification of GFP and total protein of the soluble crude extract

The quantification of GFP-tagged and non-tagged and total protein was performed through GFP fluorescence and BCA colorimetric assay. The GFP fluorescence allowed the quantification of GFP proteins by measuring the intensity of fluorescence in the microplate reader using the respective filters $\lambda_{\text{excitation}} = 485 \text{ nm}$ and $\lambda_{\text{emission}} = 535 \text{ nm}$ with a gain of 41. Therefore a calibration curve was assessed with different purer GFP concentrations within the range $10^{-6} - 10^{-1} \text{ mg/ml}$. 200 µl of crude extract obtained from cellular fractionation was pipetted in triplicates in a black 96-well microplate and the fluorescence intensity was measured. In all cases, the fluorescence intensity of 200 µl of blank sample (PBS buffer (§ 2.3.2.3.1)) was also measured in triplicates. The total protein was quantified through the colorimetric test, BCA assay. A

calibration curve was made with BSA as a standard protein within the range 0.2 - 1 mg/ml. In order to perform the BCA assay, 25 μ l of each sample (calibration curve and crude extract) was added in each well of a transparent 96-well microplate. Then, the BCA working reagent was prepared by mixing 50 parts of reagent A with 1 part of reagent B until the formation of a light green color. After, 200 μ l of BCA working reagent was added to all wells. The microplates were incubated at 37 °C for 30 minutes and the absorbance at 560 nm was measured in a microplate reader.

2.3.2.4 Large scale expression of the remaining fusion proteins and GFP

The affinity tags with the sequence WFWFWF, NWNWNW, RKRKRK and NNNNNN fused to GFP and non-tagged GFP were also expressed in bacterial hosts in large scale as well as GFP without tagging. The expression of the GFP-tagged and non-tagged followed the same procedures as those already described for the expression of RWRWRW tagged GFP. The molecular cloning procedure of the DNA fragment that encodes the specific affinity tag fused to GFP and the plasmid pET-21c was carried out by GenearthTM (Germany). The plasmid was also resuspended in 10 μ l of sterile MilliQ water. Prior to large scale expression, the cloning product was amplified in NZY α competent cells for further isolation and purification of plasmid DNA. Afterwards, competent cells BL21 (DE3) were transformed with purified plasmid DNA and the large scale protein expression was performed by using the optimal conditions: the protein expression was induced at OD_{600 nm} 0.6-0.8 with 1 mM IPTG at 37 °C with orbital shaking of 220 rpm. After protein expression, cellular fractionation was also performed and the final soluble crude extract was quantified by GFP fluorescence and BCA colorimetric test. During the protein expression and the cellular fractionation, aliquots of each stage were taken and evaluated by SDS-PAGE and fluorimetric analysis. However, in the case of WFWFWF tagged GFP, the fusion protein was found in an insoluble form. At this point, no fluorescence was visualized in the supernatant. Therefore the pellet sample (15 μ l) was analysed by fluorescence microscopy (objective of 100x of amplification). Three different photographs of the sample field were taken under bright field (to visualise the refractile bodies) and filtered light appropriate for excitation and emission ($\lambda_{\text{excitation}} = 460-490$ nm and $\lambda_{\text{emission}} = 520$ IF). The images were recorded with Cell F software and the size of the refractile bonds was determined with Image J software.

2.3.2.5 Antimicrobial assays with hexapeptide RWRWRW (RW)₃

Frozen *E. coli* K-12 (kindly provided by Dr Pedro Vidinha and Diana Garcia) at -80°C was thawed on ice. From the top of the tube, a stab was taken using a sterile pipette tip. The cells were spread on pre-warmed LB/agar plate. The LB/Agar plate was incubated overnight at 37 °C. Afterwards, in a sterile test tube was added 6 mL of LB and a colony of *E. coli* K-12 and left incubated overnight at 37 °C with 220 rpm.

The OD_{600 nm} of the overnight culture was measured in a sterile transparent 96-well microplate and diluted to an OD_{600 nm} of 0.045 with LB medium. Afterwards, 1.8 ml of the diluted culture was pipetted in each well of an autoclaved 2 ml 96-well block, and incubated in a microplate incubator at 37 °C with 500 rpm until OD_{600 nm} reached 0.15. At this point, 200 µL of different concentration of (RW)₃ were added to the wells, in triplicates for each condition, and then incubated for 7 hours at 37 °C with 500 rpm. Each row of three wells represent a different (RW)₃ concentration. The hexapeptide concentrations used were 0, 10, 100, 500, 1000, 2500 and 5000 µM. Samples were taken every 30 minutes, pipetted into a sterile 96-well transparent microplate and the OD_{600 nm} was measured in the microplate reader to monitor cell growth.

2.3.2.6. Production of RWRWRW tagged GFP in *E.coli* BL21(DE3) cells

The production of RWRWRW tagged GFP in *E. coli* BL21(DE3) competent cells was performed according with already described in § 2.3.2.2, where after the transformation of the positive clone of RWRWRW tagged GFP in pET-21c, the cell growth was carried out in a 500 ml shaking flask with 200 ml LB, 200 µl of 100 µg/ml ampicilin and 1 ml pre-inoculum at 37 °C at 220 rpm. Then the protein expression was induced with 200 µl of 1 mM IPTG at OD_{600 nm} 0.6-0.8 and incubated at 37 °C at 220 rpm. In this particular case, the monitorization was performed in a more extensive way because the aliquots were taken every two hours until 43 hours of induction. The aliquots were analyzed through optical density measurements at 600 nm and GFP fluorescence intensity in a 96 well microplate format (§2.3.2.2.4). Also, a cell growth assay was performed with only competent cells BL21(DE3). In this case, from the top of the tube of -80°C frozen BL21(DE3) cells, a stab was taken using a sterile pipette tip and then spread on pre-warmed LB/agar and 100 µg/ml ampicilin and afterwards incubated overnight at 37 °C, in order to obtained individualized colonies. Then, a pre-inoculum was performed, and after this, the BL21(DE2) cell growth was also conducted in a 500 ml shaking flask with 200 ml LB, 200 µl of ampicilin 100 µg/ml and then incubated at 37 °C with 220 rpm. In this situation, there is no induction with IPTG. The monitorization of the cell growth was also performed in an extensive way because aliquots were taken every 2 hours until 50 hours of cell growth and then evaluated through optical density measurements at 600 nm.

2.3.2.7. Production Inclusion bodies (IBs) with the WFWFWF-GFP system

The expression of the WFWFWF tagged GFP has lead to the formation of inclusion bodies when using the protein expression conditions of 37 °C and 1 mM IPTG. Due to this, the sample containing the protein as inclusion bodies were solubilised by using a standard procedure [190] and then different attempts of refolding were used. Moreover, in order to overcome the

formation of inclusion bodies, different culture medium conditions were employed (different temperatures and IPTG concentrations).

2.3.2.7.1 Solubilization and Refolding Strategies

The solubilization procedure was performed with the pellet sample obtained from the centrifugation step during cellular fractionation after cell harvest. The pellet was resuspended with the buffer A (§ 2.3.2.3.1) and then the sample was centrifuged for 15 minutes at 13700 g at 4 °C. The supernatant was kept at -20°C and then the resultant pellet was resuspended in 30 ml lysis buffer (50 mM Tris-HCl pH 7.9, 0.1 mM EDTA, 5% glycerol, 0.1 mM DTT and 0.1 M NaCl). Then 1% (v/v) of Triton X-100 was added to the lysate and incubated on ice for 10 minutes, and afterwards centrifuged at 13700 g , for 15 minutes at 4 °C. The supernatant was kept at -20°C and the IB pellet was again resuspended in 30 ml of lysis buffer with 1% (v/v) of Triton X-100 and incubated on ice for other 10 minutes. Again, the lysate was centrifuged at 13700 g for 15 minutes at 4 °C. Then the drained IB pellet was resuspended in 30 ml of lysis buffer without addition of 1% (v/v) of Triton X-100 to remove the Triton and then recentrifuged with the same conditions. The IB pellet obtained was designated as washed IB fraction and the supernatant was kept at -20 °C. The washed IB pellet was resuspended with 30 ml denaturant buffer, incubated 1 hour at room temperature with gentle stirring in order to allow the solubilization and denaturation of the IBs. Afterwards, the mixture was centrifuged at 13700 g for 15 minutes to remove any insoluble material. The supernatant was then used for further refolding of the WFWFWF tagged GFP through multisteps of dialysis.

Different attempts were performed using different denaturant buffers and denaturant decrements during dialysis. The first attempt was using the denaturant buffer 10 mM sodium phosphate, 150 mM NaCl pH 7.4 with 8 M urea and then the dialysis was performed with 2 M decrements from 8 M to 0 M at 4 °C with gentle magnetic stirring. The second attempt was prior to addition of the denaturant buffer addition, in the washing IBs step, the resuspended pellet was divided in two 15 ml lysate. After centrifugation, the two washed pellets were resuspended with the same denaturant buffer as above and after solubilization and centrifugation, the respective supernatants, 15 ml each, followed different routes. The first route included the dialysis of the 15 ml by decreasing the urea concentration from 8 M to 2 M with 2 M decrements, following dialysis with 1 M, 0.5 M and 0 M at 4 °C with gentle magnetic stirring. The second route was performed by using the 15 ml IBs fraction resuspended in denaturant buffer (10 mM sodium phosphate, 150 mM NaCl pH 7.4 with 8 M urea) in each well of 96-well block, where each well present a different affinity ligand and select a lead ligand that could bind to WFWFWF tagged GFP in denaturant conditions, so that the refolding on column could be performed afterwards. A third attempt was performed as the second one with dialysis as refolding strategy, however the denaturant buffer was 30 mM PBS, 150 mM NaCl pH 8 with 8 M

of urea. The two dialysis procedures were carried out in the same way, however at two different temperatures (4 °C and room temperature). A last attempt was performed by resuspending the IB washed pellet with 15 ml of denaturant buffer 10 mM PBS, 150 mM NaCl pH 7.4 with 8 M urea and 1% (v/v) of Triton X-100 and then incubated 1 hour at room temperature and recentrifuged at 13700xg for 15 minutes. Afterwards, the refolding strategy was the dialysis carried out at 4 °C with gentle magnetic stirring in the same way as the second and third attempts.

2.3.2.8 Avoiding the formation of inclusion bodies

The formation of inclusion bodies can be minimized by employing different protein expression conditions. For this, the first parameter studied on the impact of WFWFWF tagged GFP expression was the IPTG concentration, (0.1, 0.5 and 1 mM) added at OD_{600 nm} 0.6-0.8 in cellular cultures grown at two different temperatures, 30 and 37 °C. Then, the other parameter considered was the temperature of WFWFWF tagged GFP expression and four different temperatures were studied: 27, 30, 33 and 37 °C and induction with 1 mM of IPTG at OD_{600 nm} 0.6-0.8. These experiments were performed according with what was described in § 2.3.2.2. However, one particularity of these studies was the fact that cell growth was always conducted at 37 °C, and the temperature was only decreased after the addition of IPTG. Moreover, the final volume of cell growth was 150 ml. Each condition was tested in triplicates. The monitorization of protein expression was performed with the aliquots taken every hour after IPTG addition until 19 and 20 hours of induction. The aliquots were evaluated through optical density measurements at 600 nm, SDS-PAGE (§ 2.3.2.2.3) and GFP fluorimetric analysis (§ 2.3.2.2.4). Afterwards, in order to evaluate if the WFWFWF tagged GFP was produced as soluble or insoluble form, all the stages of cellular fractionation were performed until the centrifugation step after cellular lysis (§ 2.3.2.3.1). The pellet and supernatant were then evaluated by SDS-PAGE (§ 2.3.2.2.3) and GFP fluorimetric analysis (§ 2.3.2.2.4). The pellet obtained was also observed under fluorescence microscopy, according with the procedure described in (§ 2.3.2.4).

2.3.3 Screening of library based on the Ugi reaction with crude extracts containing GFP-tagged and non-tagged

Each round of screening involved regeneration, equilibration, loading, washing and elution in each well of the 96-well filtration block, where each well corresponds to a resin functionalized with a different affinity ligand in a total of 64 ligands. The steps of regeneration and equilibration were performed under gravitational force instead of centrifugal force as in the washing step. The seven targets were five different fusion proteins with different affinity tags (RWRWRW, WFWFWF, NWNWNW, RKRKRK and NNNNN) under physiological conditions, WFWFWF

tagged GFP in presence of 8M urea and with non-tagged GFP. The regeneration was conducted by adding 250 μ l of regeneration buffer (0.1M NaOH, 30% (v/v) isopropanol) alternated with 250 μ l of distilled water in a total of three times. Afterwards, the equilibration step was carried out by adding 12x 250 μ l of the equilibration buffer (PBS buffer § 2.3.2.3.1). In the last step of equilibration, the samples were collected on a 96-well microplate UV-half area (VWR) and then the absorbance was measured at 280 nm, to confirm that the $A_{280\text{ nm}}$ reached ≤ 0.005 . Then 250 μ l of the respective target (crude extract containing GFP-tagged and non-tagged) was loaded on each well and incubated at 4°C for 1 hour with manual shaking to ensure a good mixing between the resin and the target. After incubation, the flow-through was collected in a black 96-well microplate (Brand, VWR) by placing the microplate below the 96-well block and then centrifuged together at 170 xg for 1 minute. Afterwards, each well was washed with 12x 250 μ l of binding buffer and between the washes the samples were collected in a black 96-well microplate by centrifugation at 170 xg for 1 minute. The same procedure was carried out for the screening with target WFWFWF tagged GFP in presence of 8M urea, however during the equilibration step the last 8 x250 μ l were performed with PBS buffer (§ 2.3.2.3.1) with 8M urea. Additionally, during the washing steps after the loading, a completely urea removal was conducted with the following urea decrements 3x 250 μ l PBS 8M urea, 3x 250 μ l PBS 4M urea, 3x 250 μ l PBS 2M urea and 3x 250 μ l PBS without urea in order to promote protein refolding. After the screening, the resins in all 96-well blocks were regenerated and stored at 0-4°C in 20% (v/v) ethanol. The fractions loaded and the samples collected in the black 96-well microplates in all washing steps were quantified by the BCA assay and GFP fluorescence (§2.3.2.3.2.). In case of the screening with WFWFWF tagged, the loaded fractions and the collected washes were only possible to be quantified by GFP fluorescence as the urea concentrations above 3M interfered with BCA assay according to supplier instructions.

2.3.4 Synthesis and re-screening of the promising lead ligands and the respective crude extracts containing GFP-tagged and non-tagged

2.3.4.1 Synthesis of the promising lead ligands

The synthesis of the promising lead ligands (Table 2.3) was also carried out on in duplicates on the 96-well filtration block according to 2.3.1.

Table 2.3 Promising lead ligands selected t from the screening between combinatorial library based on the Ugi reaction and the crude extracts containing GFP-tagged and non-tagged.

PROMISING LEAD LIGANDS	TARGET	96-WELL BLOCK Row
A1C6; A2C4; A2C5; A2C8; A3C5; A3C6; A3C7; A3C8; A5C6; A6C6; A6C7; A6C8	RWRWRW tagged GFP	A
A2C2; A2C3; A2C4; A2C5; A2C6; A3C1; A3C4; A3C5; A5C1; A5C6; A6C8; A7C8	NWNWNW tagged GFP	B
A2C4; A5C4; A7C1; A7C2; A7C3; A7C4; A7C5; A7C6; A7C7; A7C8	RKRKRK tagged GFP	C
A1C5; A2C4; A2C5; A3C6; A2C7; A2C8; A3C1; A3C6; A3C7; A3C8: A6C1	WFWFWF tagged GFP (Non-denaturant conditions)	D
A4C5; A4C8; A8C8	WFWFWF tagged GFP (denaturant conditions)	E
A3C7; A3C8; A4C7; A4C8; A6C1; A6C4; A6C5; A6C6; A6C7	Non-tagged GFP	G

2.3.4.2 Screening of the promising lead ligands and the respective crude extracts containing GFP-tagged and non-tagged

The screening of the promising lead ligands and the respective targets was carried out as in 2.3.3. However, in this case the elution step was added on the screening after the washing step by using a different elution condition in each 96-well block. The bound target was eluted by adding 5x250 µl of elution buffer, and between each addition, the samples were also collected in a black 96-well microplate by centrifugation at 170 *xg* during 1 minute. Afterwards, all the black 96-well microplates were also covered with aluminum foil and kept at 4°C. The elution conditions used were 0.1M Glycine-HCl, pH 3 and 0.1M Glycine-NaOH, pH 11, except for the .WFWFWF tagged GFP screening under denaturant condition, where the last condition was substituted by 400 mM NaCl. After elution, the resins in each well regenerated and stored at 4°C in in 20% (v/v) ethanol. The fractions loaded and the samples collected in the black 96-well microplates in all washing and elution steps were quantified by the BCA assay and GFP fluorescence §2.3.2.3.2. In case of the screening with WFWFWF tagged, the loaded fractions and the collected washes were only possible to be quantified by GFP fluorescence as the urea concentrations above 3M interferes with BCA assay according to supplier instructions. The enrichment of the respective target by the promising lead ligands was also evaluated by SDS-PAGE analysis. The same evaluation was carried out with the blank agarose against all the fusion proteins and non-tagged GFP.

SDS-PAGE analysis

The enrichment of the respective target was evaluated by SDS-PAGE analysis according to § 2.3.2.2.3. The 12.5% acrylamide gel and sample preparation was carried out in the same way, however the volume pipetted from each sample and sample buffer was 5 µl and 5 µl of the low molecular weight protein marker (NZYTech) used. The samples tested were loading sample, flow-through and the first elution. Gels were stained with Silver Stain Plus (Bio-Rad). The fixative step was carried out by placing the gel on 200 ml fixative enhancer solution (100 ml 100% (v/v) methanol, 20 ml 100% (v/v) glacial acetic acid, 20 ml fixative enhancer concentrated solution and 60 ml distilled water) for 20 minutes under gentle agitation. The solution was disposed off on hazardous waste. The gels were rinsed with 200 ml of distilled water for 10 minutes under gentle agitation. This step was repeated twice and after each rinse, the water also was decanted for hazardous waste. Gels were placed on the staining solution for a maximum time of 20 minutes also under gentle agitation, and then the solution was also discarded for hazardous waste. The staining solution was prepared immediately before using, by adding the components in the following order: 35 ml of distilled water, 5 ml of silver complex solution, 5 ml of reduction moderator solution, 5 ml of image development reagent and 50 ml of the room temperature development accelerator solution. After this step, the staining reaction was stopped, by placing the gels in a 5% (v/v) acetic acid solution for 15 minutes, also under mild agitation. Once stopping the reaction, the gels were rinsed with high purity water for 5 minutes. Then, the gels were photographed in a UVITEC transilluminator through KODAK 1D 3.6 software.

2.3.5 Optimization and characterization of the lead affinity pairs “tag-receptor”*2.3.5.1 Scale-up synthesis of lead affinity ligands*

The required amount for the scale-up synthesis of the resins functionalized with lead ligands was carried out as described in § 2.3.1. However, each ligand was synthesized in 60-ml Nalgene bottle with a total of 20 g of moist gel (60 ml 50% (v/v)) slurry instead of 0.25 g in a 96-well block. The lead ligands synthesized were A4C8, A3C4, A7C1 and A4C7.

2.3.5.2 Screening of the lead ligands with the respective target in on-column format

The screening by using the on-column format was performed in the same manner as described in §2.3.4.2, however in this case, 250 mg of the moist gel with the lead ligand immobilized was packed in 4 ml columns (0.8 x 6 cm) and the assay ran under gravitational flow, and the resultant fractions were collected in 1.5-ml microcentrifuge tubes, covered with aluminum foil to protect from the light and kept at 4°C. The elution conditions were the same as in § 2.3.4.2.

The screening was conducted by using the affinity pairs: WFWFWF tagged GFP-A4C7 (under denaturant conditions), NWNWNW tagged GFP-A3C4, RKRKRK tagged GFP-A7C1 and GFP-

A4C8. This assay was also extended to blank agarose without any functionalization and the target crude extracts containing the target biomolecules. The fractions obtained were analysed by BCA assay, GFP fluorescence §2.3.2.3.2 and SDS-PAGE analysis (§2.3.4.2).

2.3.5.3 Lead ligands selectivity

The selectivity of the affinity ligands A3C4 and A7C1 was evaluated through the screening on column according to § 2.3.5.2 against NWNWNW tagged GFP, RKRKRK tagged GFP and GFP.

2.3.5.4 Determination of the binding constants

Binding affinity constants (K_a) and theoretical maximum capacity (Q_{max}) were estimated by static partition equilibration studies on a 96-well filtration block. For this, 400 μ l (~ 250 mg resin) of each functionalized resin with the lead ligand (A3C4, A7C1 and A4C7) was added in each well of two entire rows of the 96-well block in total of 24 wells per ligand. Then, the 72 wells were regenerated and equilibrated (§2.3.3) and 250 μ l of the respective target loaded in each well and incubated overnight at 4°C with orbital shaking at 200 rpm. The crude extracts were tested with several dilution factors from 1 to 10000 and this assay was run in duplicates.

At end of incubation, a black 96-well microplate was placed underneath the 96-well block and centrifuged at 170 xg for 1 minute. Then, the GFP fluorescence of the black 96-well microplate with the collected flow-through of all the samples was evaluated on the microplate reader (§2.3.2.3 .2).

2.3.5.5 Optimization of the elution profile for the lead affinity pairs

The optimization of the elution conditions for each affinity pair was carried in a 96-well block format according to § 2.3.4.4 by using two different approaches as shown in Table 2.4 and 2.5 with a few modifications. The eluted fractions were quantified only by GFP fluorescence and BCA assay. The respective elution buffer was added in each well until no GFP fluorescence was detectable. All the elution conditions were used in duplicates.

Table 2.4 Elution conditions employed on the first approach of the optimization

ELUTION CONDITIONS OPTIMIZATION – FIRST APPROACH			
LEAD AFFINITY PAIRS	PH INFLUENCE	0.1 M Citrate pH 3	
		0.1 M Citrate pH 4	
		0.1 M Citrate pH 5	
		0.1 M Citrate pH 6	
		0.1 M Sodium phosphate pH 7	
		0.1 M Sodium phosphate pH 8	
		0.1 M Glycine-NaOH pH 9	
		0.1 M Glycine-NaOH pH 10	
	SALT INFLUENCE	IONIC STRENGTH (0.010 M Sodium phosphate pH 7.4)	0.15M NaCl
			0.25 M NaCl
			0.5 M NaCl
			1 M NaCl
		PHOSPHATE CONCENTRATION (pH 7.4, 0.15M NaCl)	0.010 M Sodium phosphate
			0.1 M Sodium phosphate
			0.5 M Sodium phosphate
			1 M Sodium phosphate

Table 2.5 Elution conditions employed on the second approach of the optimization

ELUTION CONDITIONS OPTIMIZATION – SECOND APPROACH					
LEAD AFFINITY PAIRS	TAGGED SYSTEMS	NON-TAGGED GFP	BEST ELUTION CONDITIONS (0.1 M Glycine-NaOH pH 9 and pH 11)	ADDITIVES	0% (v/v) Ethylene glycol
					5 % (v/v) Ethylene glycol
					10 % (v/v) Ethylene glycol
					25 % (v/v) Ethylene glycol
					50 % (v/v) Ethylene glycol
		IONIC STRENGTH	0 M NaCl		
			0.15 M NaCl		
			0.25 M NaCl		
			0.5 M NaCl		
			1 M NaCl		
TAGGED SYSTEMS	TAGGED SYSTEMS	NWNWNW TAGGED GFP	COMPETITIVE ELUTION WITH TRYPTOPHAN	MILD CONDITIONS (0.01 M sodium phosphate pH 7.4 0.15M NaCl)	0 M Tryptophan
					0.001 M Tryptophan
					0.005 M Tryptophan
					0.010 M Tryptophan
					0.015 M Tryptophan
		BEST ELUTION CONDITION (0.1 M Glycine-NaOH pH 11)	0 M Tryptophan		
			0.001 M Tryptophan		
			0.005 M Tryptophan		
			0.010 M Tryptophan		
			0.015 M Tryptophan		
NWNWNW TAGGED GFP AND RKRKRK TAGGED GFP	BEST ELUTION CONDITION (0.1 M Glycine-NaOH pH 11)	ADDITIVES	0% (v/v) Ethylene glycol		
			5 % (v/v) Ethylene glycol		
			10 % (v/v) Ethylene glycol		
			25 % (v/v) Ethylene glycol		
			50 % (v/v) Ethylene glycol		
IONIC STRENGTH	0 M NaCl				
	0.15 M NaCl				
	0.25 M NaCl				
	0.5 M NaCl				
	1 M NaCl				
TAGGED SYSTEMS	TAGGED SYSTEMS	RKRKRK TAGGED GFP	COMPETITIVE ELUTION WITH ARGININE	MILD CONDITIONS (0.01 M sodium phosphate pH 7.4 0.15M NaCl)	0 M Arginine
					0.01 M Arginine
					0.1 M Arginine
					0.5 M Arginine
					0.750 M Arginine
		BEST ELUTION CONDITION (0.1 M Glycine-NaOH pH 11)	0 M Arginine		
			0.01 M Arginine		
			0.1 M Arginine		
			0.5 M Arginine		
			0.750 M Arginine		

2.3.5.6 Study of the lead affinity pairs in the automated system ÄKTA avant 25

The screening was carried out on the ÄKTA avant 25 automated system with the lead ligands packed on the HiScale16/20 columns (dimensions: 16 cm inner diameter and 40 cm length). A volume of 50% (v/v) resin slurry in 20% (v/v) ethanol used to pack 1g of functionalized resin with lead ligands into the columns was determined according to:

$$V \text{ (ml)} = \frac{A_C \times L_{\text{packed}} \times Cf}{C_{\text{slurry}}}$$

Where A_C corresponds to the cross sectional area of the column, L_{packed} to packed bed height (0.5 cm), Cf to compression factor that in case of Sepharose CL-6B is 1.15, according to supplier instructions and C_{slurry} to the concentration of the slurry (50% (w/v)). The volume used to pack 1g column was 2.3 ml of 50% (w/v) slurry functionalized with lead ligands.

2.3.5.6.1 Tests with NWNWNW tagged GFP-A3C4, RKRKRK tagged GFP-A7C1 and non-tagged GFP-A4C7

The adsorbents were packed (1g) on column with a $0.3 \text{ ml}\cdot\text{min}^{-1}$ flow and prepared for analysis using an ÄKTA avant 25. Afterwards the solid support was subjected to the following steps with a flow rate of $1 \text{ ml}\cdot\text{min}^{-1}$: regeneration (regeneration buffer alternated with H_2O , 10 c.v. each in a total of 30 c.v.), equilibration (PBS 15 c.v.) and loading (1 c.v. of 1ml of the respective target). Then the column was washed with binding buffer (PBS 20 c.v.), and then a step of elution with the respective buffer (10 c.v.) was performed and at the end other regeneration step was done as already described with a total of 20 c.v. (regeneration buffer 10 c.v.) alternated with H_2O (5 c.v.). All the fractions were collected from the loading in 1 ml 96-well block and kept at 4°C . The fractions were then quantified by GFP fluorescence (§ 2.3.2.3.2), Bradford assay and evaluated by SDS-PAGE (§2.3.4.2). On SDS-PAGE analysis, the molecular weight used was Novex® Sharp protein standard (Invitrogen). Two different novel runs were performed for each affinity pair by using the two best elution conditions (Table 2.6) selected on the optimization of elution conditions (§2.3.5.5).

Table 2.6 Best elution conditions employed on the screening between the affinity pairs on the ÄKTA avant 25 automated system.

BEST ELUTION CONDITIONS	LEAD AFFINITY PAIRS		
	NWNWNW TAGGED GFP AND A3C4	RKRKRK TAGGED GFP AND A7C1	NON-TAGGED GFP AND A4C7
	0.1M Glycine-NaOH pH 11	PBS with 500 mM Arginine	0.1M Glycine-NaOH pH 9
0.1M Glycine-NaOH pH 11 50% Ethylene glycol	0.1M Glycine-NaOH pH 11 150 mM NaCl	0.1M Glycine-NaOH pH 9 50% Ethylene glycol	

Coomassie Plus (Bradford) assay kit

The quantification of total protein was performed with the standard Bradford assay (Coomassie Plus™ assay kit), a colorimetric assay that relies on a spectral shift from red (free dye) to blue due to the formation of the complex coomassie dye-protein. Samples to be quantified were added to each well of a 96-well transparent microplate (10 µl /well). Coomassie Plus™ protein assay reagent (300µl) was added to each well and mixed in an orbital shaker (30 sec, 50 rpm) at room temperature. The plate was left to incubate for 10 minutes at room temperature, after which the absorbance was read at 595nm using a microplate reader Synergy HT (BioTek). Also a calibration curve with standard BSA concentrations (100-15000 µg/ml) was prepared by using the same procedure. Controls containing the regeneration, binding and elution buffers were also quantified.

2.3.5.6.2 Purification of GFP fusion proteins with A4C7 ligand

The ligand A4C7 was used for the purification of a GFP fusion protein, in particular the fusion between GFP and *B. megaterium* spore cortex lytic enzyme SleL produced in *Lactococcus lactis*, which was kindly provided by Dr Isik Ustok, University of Cambridge. The procedure was carried out according to § 2.3.5.6.1. Moreover, the purification profile was also evaluated by Western blot analysis.

Western Blot analysis

Protein samples were initially resolved by SDS-PAGE, and then transferred onto PVDF Membranes (Life Technologies Ltd., Paisley, UK) which was prewashed with methanol for 30 seconds and transfer buffer for 5 minutes. Transfer was performed at room temperature using a semi-dry blotting unit (Trans-Blot SD, Semi Dry Transfer Cell, Biorad) at 15 V for 15 minutes.

Following transfer and washing in water twice for 5 minutes, the membrane was blocked with Tris buffered saline (TBS) containing 5% BSA and 0.5% Tween-20 overnight at 4°C. After blocking, the membrane was washed with buffer composed of TBS containing 0.1% Tween 20 (TBST) for 1 minute prior to incubation with appropriate antibodies. For the detection of GFP fusion proteins, primary antibody (Rabbit PAb to GFP) was diluted in blocking buffer (1:2,500 dilution) and then added to the membrane at room temperature for 1 hour with gentle shaking. After extensive washing, a secondary antibody (Goat PAb to Rabbit IgG;) conjugated with horseradish peroxidase (HRP) was diluted in TBST buffer (1:10,000 dilution) and the membrane was incubated with secondary antibody at room temperature for 30 minutes with gentle shaking. Following the final washing, excessive water was removed from the membrane which was then placed on an acetate paper facing upwards. The chemiluminescence reaction was conducted for 1 minute (Novex, HRP Chemiluminescent substrate reagent kit, Life Technologies, Paisley, UK) according to the manufacturer's instructions, and signal was detected on an X-ray film (GE, Amersham Hyperfilm ECL) using the X-ray film developer.

2.3.5.6.2 WFWFWF tagged GFP-A4C8 – Matrix-assisted refolding strategy

A4C8 adsorbent was packed (1g) on column with a flow of 0.3 ml.min⁻¹ and subjected to the following steps with a flow rate of 1 ml.min⁻¹: regeneration (0.1M NaOH, 30% isopropanol (v/v) alternated with H₂O, 10 c.v. each in a total of 30 c.v.), equilibration (10 mM sodium phosphate, 150 mM NaCl, pH 7.4 8M urea 15 c.v.) and loading (1 c.v. of 1ml of WFWFWF tagged GFP in presence of 8M urea as a result of solubilization of inclusion bodies described in § 2.3.4.7.1). In this particular case, after the loading, a refolding step was performed by using a gradient of 1M NaCl in binding buffer in order to dilute the urea (250 c.v.). Afterwards, elution was proceeded with 0.1M Glycine-NaOH pH 11. (30 c.v.). The regeneration was done as already described with a total of 20 c.v. (0.1M NaOH, 30% isopropanol (v/v) (10 c.v.) alternated with H₂O (5 c.v.). All the fractions were collected from the loading in 1 ml 96-well block in all the steps except for the refolding step, where the samples were collected in 50ml centrifuge tubes. The fractions were then quantified by GFP fluorescence (§2.3.2.3.2) and SDS-PAGE analysis (§2.3.4.2). Also, in this last analysis, the molecular weight used was Novex® Sharp protein standard (Invitrogen)

CHAPTER 3

PRODUCTION OF GFP-TAGGED PROTEINS

SUMMARY

Five different expression systems were constructed each containing a different tag sequence, namely RWRWRW, WFWFWF, NWNWNW, RKRKRK and NNNNN. These affinity tags were fused to GFP as a reporter protein. Since these affinity tags present novelty on the sequence, the gene was designed and subcloned into an expression vector for protein expression using *E. coli* BL21(DE3) as host cells. As a proof of concept, the gene design and subcloning steps were accomplished for RWRWRW tagged GFP. For the remaining affinity tags, the expression vectors were constructed by Genart™. The expression conditions were optimized for RWRWRW tagged GFP and then extended to all other GFP-tagged proteins. The production of the affinity tags fused to GFP in bacterial cells revealed to be straightforward except for RWRWRW and WFWFWF tagged GFPs. In the first case, the RWRWRW peptide exhibited antimicrobial properties, which influenced the production of the fusion protein. The expression of WFWFWF tagged GFP resulted in the formation of inclusion bodies, which were subsequently solubilised with chaotropic agents and different attempts of protein refolding were performed. Moreover, different cell growth conditions, such as temperature and inducer levels, were studied in order to minimize the formation of inclusion bodies. The temperature has revealed to have a higher impact on the overall production of soluble WFWFWF tagged GFP rather than the inducer levels. Overall, it was possible to produce sufficient amounts of soluble GFP tagged proteins to be further used in affinity purification systems.

3.1 INTRODUCTION

Recombinant proteins find various applications, in therapy and diagnostics as well as in industry. Depending on the final application, their production can be carried out in different host cells and with different levels of purity. Bacterial host cells are widely used for the expression of heterologous proteins and *Escherichia coli* is the most popular system. This host presents several advantages, such as simplicity, well established methods for genetic manipulation, high products yields, rapid expression and cost effectiveness [164, 191-192]. The main disadvantages are related with the limitation to undertake pos-translational modifications [164, 191-193] and the formation of insoluble aggregates. In addition, genes presenting rare or low-usage codons, that lead to translational errors and incorrect expression of the target recombinant protein, cannot be expressed by *E. coli* [194-195]. Alternative host cells, usually more costly and less scalable but overcoming some of bacterial limitations, include eukaryotic organisms (e.g. mammalian cells, yeast, insect) and transgenic plants and animals [164, 196-203].

The production of recombinant proteins can be facilitated through the fusion with reporter proteins. The Green Fluorescent Protein (GFP) from jellyfish *Aequorea victoria* is the most extensively studied reporter protein and is a well-developed marker in bioprocess control and analysis, applied to monitor dynamic processes in cells, organisms and organs (e.g. gene expression, intracellular localization and protein-protein interactions studies) [164, 204-205]. GFP-like proteins do not require cofactors, enzymes or substrates other than molecular oxygen for the formation of the chromophore [166, 206]. This property makes the formation of the chromophore easier in live organisms, tissues and cells. Other advantages of these proteins include the low-toxicity and stability in a wide range of pH and solvents, the easy detection in a bulk cell suspension without cell disruption, and relatively small molecular weight which contributes to a low burden to host cells [206].

In order to develop an affinity pair “tag-receptor” for the purification of fusion proteins, five hexapeptides, RWRWRW, WFWFWF, NWNWNW, RKRKRK and NNNNNN, were previously considered as potential affinity tags and chosen by presenting different properties of hydrophilicity/hydrophobicity and potential to establish hydrogen bonds and hydrophobic interactions with receptors [207-208]. This chapter focus on the production of the aforementioned affinity tags fused to GFP (reporter protein) in *E. coli* cells (Fig. 3.1). The gene that encodes for the tagged GFP was designed and subcloned into a commercial expression vector (pET-21c). The cloning step and optimization of overexpression conditions were carried out only for the RWRWRW tagged GFP as proof of concept. Once selected, the nucleotide sequences encoding for the remaining affinity tags fused to GFP were synthesized in the expression vector pET-21c by Geneart, and then the GFP-tagged proteins produced. After

protein expression, the cells were disrupted and fractionated to obtain homogenous crude extracts for further purification through affinity chromatography.

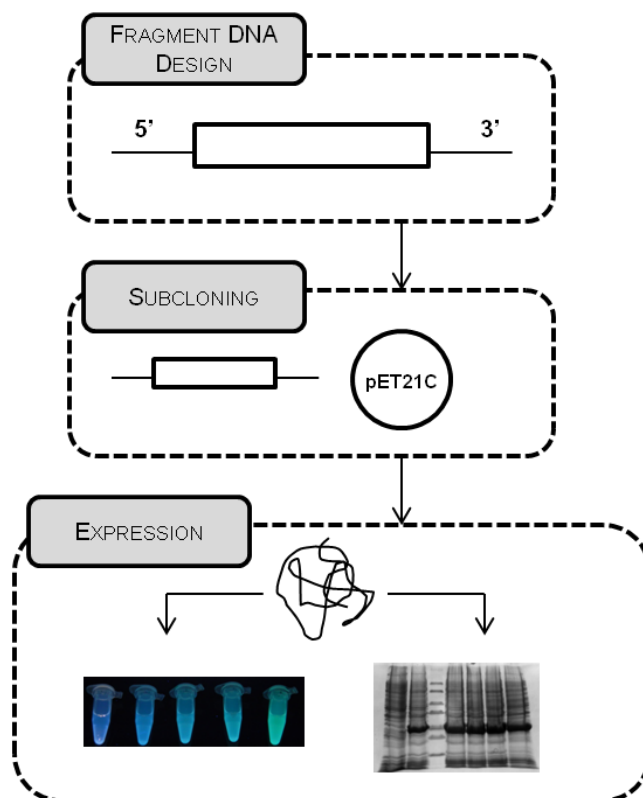


Figure 3.1 Production of affinity tags fused to GFP. The production comprises three main steps: the design of the nucleotide sequence that encodes for the affinity tags (RWRWRW, WFWFWF, NWNWNW, RKRKRK and NNNNN) fused to GFP, the subcloning of this sequence in an expression vector such as pET-21c and then the expression of GFP-tagged by using *E.coli* as host cells.

3.2 FRAGMENT DNA DESIGN

The design of the fragment DNA was carried out by considering the nucleotide sequences encoding for: The reporter protein (GFP) fused to the affinity tag at the N-terminal; The affinity tag and the spacer of three proline residues before enterokinase specific cleavage site; several endonucleases restriction sites (Fig. 3.2).



Figure 3.2 Scheme of the (A) fragment DNA and (B) respective nucleotide sequence that encodes for RWRWRW tagged GFP. The fragment DNA comprises the nucleotide sequence that encodes for the affinity tag and the enterokinase recognition site; also the nucleotide sequence of the reporter protein (GFP) flanked by two polylinkers with multiple restriction sites with a total of 813 base pairs (bp).

Five different GFP-tagged fragment DNA were designed, exhibiting the same nucleotide sequence except the tag nucleotide sequence. The major concern of the design is to avoid the usage of rare codons in *E.coli* [194] to minimize translational errors. The optimized GFP nucleotide sequence used in the design was based on the commercial sequence available in the plasmid pGFPuv (Clontech). The *Nhe* I restriction site was extremely important in the subcloning process because it allows the insertion of the GFP-tagged fragment DNA on the pET-21c expression vector, widely used as an expression vector for the production of recombinant proteins [187, 209]. The incorporation of the enterokinase recognition site was considered for future affinity tag removal after protein purification. The polylinkers are short DNA fragment with recognition sites for specific restriction enzymes and are relevant for the all stages of the subcloning process. The choice of the restriction enzymes for the polylinkers was also careful due to the fact that the restriction enzymes cannot cleave the tag or the GFP nucleotide coding sequence. Therefore, all sequences were analysed with ChromasPro software for restriction mapping. An additional concern for the polylinker P1 was to include nucleotides codifying for amino acids less likely to interfere in the binding between the tag and the immobilised ligand on the purification step. Once designed, the fragment DNA synthesis was carried out on pMA vectors by Geneart™.

3.3 MOLECULAR CLONING OF RWRWRW-GFP ENCODING GENE IN A pET-21c VECTOR

The fragment DNA that encodes for the fusion protein RWRWRW-GFP was subcloned into the expression vector pET-21c, as a proof of concept. The remaining fragment DNA encoding for WFWFWF-GFP, NWNWNW-GFP, RKRKRK-GFP, NNNNNN-GFP and GFP were synthesized and subcloned into the expression vector by GenearthTM. The resulting expression vectors were denominated according with the tag sequence such as pAP001 (RWRWRW-GFP), pAP002 (WFWFWF-GFP), pAP003 (NWNWNW-GFP), pAP004 (RKRKRK-GFP), pAP005 (NNNNNN-GFP) and pAP006 (GFP).

Firstly both plasmids pAP001 and pET-21c were amplified in NZY5 α cells, the plasmid DNA was isolated using a Midiprep kit (Omega) and the final concentration and purity of DNA was determined by standard spectrophotometric analysis. The pAP001 was at 425 ng/ μ l with a purity of 1.73 and the pET-21c at 245 ng/ μ l with a purity of 1.63. Before the subcloning process, the hydrolysis of pAP001 and pET-21c vectors with the restriction enzymes *Nhe* I and *Eco*R I (Fig. 3.3) was required. The fragment DNA that encodes for the tagged protein has 843 bp and the linear plasmid pET-21c presents 5443 bp. Double digestion of pAP001 released the fragment of interest with 748 bp. Afterwards both insert and double digested pET-21c vector were purified by gel electrophoresis (E-gel[®] electrophoresis system).

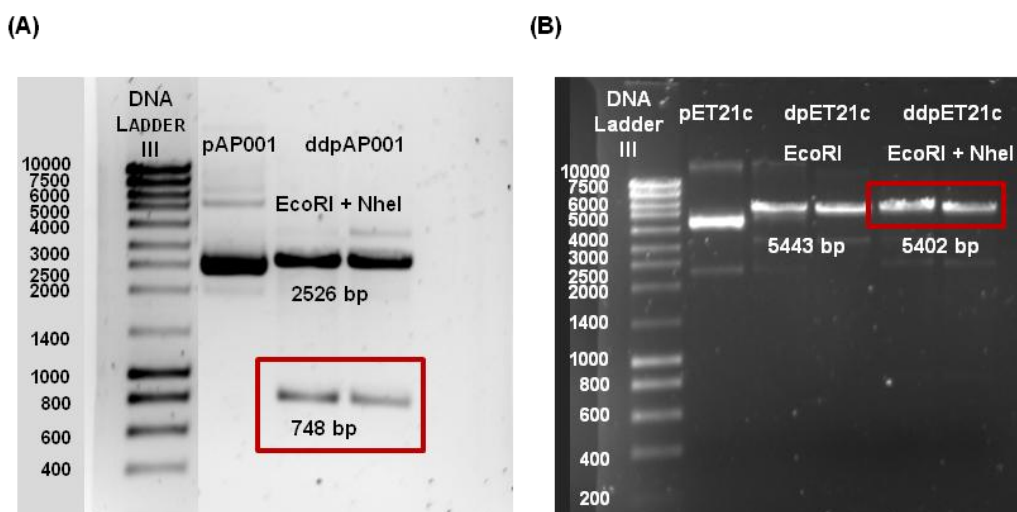


Figure 3.3 Hydrolysis with restriction enzymes *Nhe* I and *Eco*R I of (A) pAP001 plasmid that carries the gene encoding for RWRWW-GFP fusion protein (insert) and (B) pET-21c expression vector. Both plasmids were sequentially digested, first with *Nhe* I followed by hydrolysis with *Eco*R I. A 2 μ l aliquot of each reaction was applied on the agarose gel (0.8%). After the hydrolysis, the pAP001 releases a fragment of 748 bp, corresponding to the insert fragment DNA and the expression vector (ddpET-21c) becomes linear, corresponding to a size of 5402 bp. In the first lane of each gel 5 μ l of DNA Ladder III (NZytech) was applied as a DNA size marker. The agarose gel presents 0.8 % (w/v) agarose).

The quantification of both DNA fragments was carried out according with Fig. 3.4. Each band of the ladder presented specific fluorescence intensity and corresponds to a specific amount of DNA. Considering this, the fluorescence intensity of the bands of interest (insert and double

digested pET-21c) was evaluated with Image J software and the concentration of DNA was estimated by comparison with the DNA mass standards of the ladder.

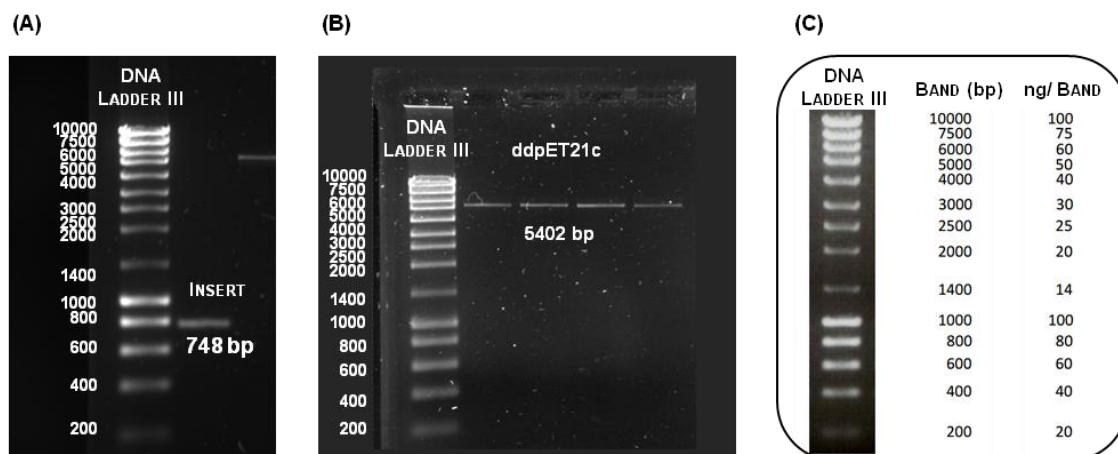


Figure 3.4. DNA quantification of the insert (A) and the double digested expression vector (B) by agarose gel electrophoresis (0.8 % (w/v) agarose) analysis using the (C) NZYDNA Ladder III from Nzytech. After purification, the fluorescence intensity of the fragments was estimated using the Image J software and then compared with the respective band of the ladder, for an estimation of DNA concentration. The volume loaded into the gels of the purified fragment released from pAP001 and the double digested pET-21c (ddpET-21c) was 5 μ l. The final concentration estimated for the fragment released from pAP001 was 1.2 ng/ μ l. Regarding the gel of pET-21c, four lanes present (lanes 1, 2, 3, 4 from left to right) different samples of the purified ddpET-21c and according with the evaluation of the fluorescence intensity of the bands, the estimated concentration of ddpET-21c corresponds to 8 ng/ μ l in case of lane 1 and 2, 9.65 ng/ μ l to lane 3 and 4.02 ng/ μ l to lane 4.

After estimating the amount of DNA, the cloning procedure was performed with the enzyme T4 DNA ligase by using a ratio of insert:vector of 3:1. This enzyme is responsible for the single-strand nicks repair for DNA double helix. The procedure of the subcloning procedure was based on a standard ligation procedure (Nzytech), where the required amount of vector is within 20-50 nanograms. In order to calculate the suitable amount of insert to use in the ligation reaction, eq. 4.1 was used according with the supplier's manual.

$$\text{Insert amount (ng)} = \frac{\text{ng Vector} \times \text{kb Insert}}{\text{kb Vector}} \times \text{Insert : Vector ratio molar} \quad (\text{Eq. 4.1})$$

The product of ligation reaction with T4 DNA ligase was used to transform NZYSTAR competent cells. The transformed cells were incubated overnight in LB/agar plates containing ampicillin. Twelve random white colonies were selected from the LB/agar plate. The choice of the antibiotic was done according with the marker resistance of the expression vector, the amp^r gene. Afterwards, the twelve colonies were grown in LB medium containing ampicillin for plasmid DNA isolation and purification. In order to confirm the presence of positive clones, a hydrolysis with the same restriction enzymes *Nhe* I and *EcoR* I was performed (Fig. 3.5), and those that released a DNA fragment of 748 bp were positive as this would correspond to the GFP tagged

fragment DNA. According with Fig. 3.5, all ligation products released a fragment around 800 bp which corresponds to the insert of interest. Therefore, the plasmid DNA of some colonies, such as 3, 4 and 10, were sequenced by STABvida in order to confirm DNA sequences. Clone 4 was used to transform *E. coli* BL21DE(3) competent cells to begin the GFP tagged expression studies.



Figure 3.5 Evaluation of the hydrolysis of pAP001 of twelve selected colonies obtained from the subcloning procedure with restriction enzymes *Nhe* I and *Eco*R I by agarose gel electrophoresis (0.8 % (w/v) agarose). A volume of 2 μ l of each reaction was applied. The positive clones were identified by the presence of a fragment with 843 bp.

3.4. RWRWRW TAGGED GFP SMALL SCALE EXPRESSION STUDIES

The small scale expression studies of the RWRWRW tagged GFP fragment DNA on pET-21c plasmid were carried out by using *E. coli* BL21(DE3) as host cells. The protein expression on a pET system requires a chromosomal T7 RNA polymerase gene, usually present in *E. coli* BL21(DE3) strain, that is controlled by the T7 *lacUV5* promoter and inducible by IPTG [187, 209-210].

The great advantages of the pET system are the high expression levels and the versatility, as the presence of multicloning sites allows the insertion of different genes [187, 209-210]. Moreover, this system presents a reduced background expression before IPTG induction due to the existence of T7 lysozyme, a natural inhibitor of T7 RNA polymerase [187, 209-210]. In particular, pET-21c already has an AUG start codon and prokaryotic ribosome binding site [187]. The RWRWRW-GFP encoding fragment DNA was cloned in a proper reading frame at the 3'-end of the AUG start codon, which corresponds to a N-terminal methionine in the transcript.

The *E. coli* BL21(DE3) competent cells have several important features besides the presence of a chromosomal copy of the T7 RNA polymerase gene. This host is protease deficient due to the mutations on the gene that encodes *E. coli* outer membrane proteases [187, 209-210], and these cells are compatible with most pET vectors [209].

The small scale expression tests were carried out using the positive clone 4 (pAP001_pET21C), where at two $OD_{600\text{ nm}}$ (0.6-0.8 and 2), different concentrations of IPTG (0.2, 0.5, 1, 1.5 and 2 mM) were used to induce the expression of the target protein. Cell growth was monitored by optical density measurements at 600 nm. The kinetic profile for the production of GFP-fusion proteins was carried out by fluorescence assays ($\lambda_{\text{excitation}} = 485\text{ nm}$ and $\lambda_{\text{emission}} = 535\text{ nm}$) in a microplate format and by SDS-PAGE. The GFP fluorescence assay can be used as a simple, real time quantification method of GFP-fusion proteins produced intracellular without the need to perform cell lysis for further SDS-PAGE analysis. The target proteins can be quantified in a rapid and accurate way. Additionally, this technique is useful to monitor large scale production of recombinant proteins [211].

During the expression of RWRWRW tagged GFP, aliquots were collected every hour between 6 and 22 h of induction. Also, an aliquot was taken at the time of induction of gene expression with IPTG, representing the $t=0$ of the expression profile. All samples were evaluated in order to select the IPTG concentration and induction time leading to higher yields of GFP-tagged protein. As all aliquots correspond to different times of induction with different optical measurements of cell growth media, the amount of protein in each fraction was normalized by the optical density. Regarding, the SDS-PAGE evaluation, the GFP-tagged fusion protein presents a total weight of 31 kDa, 29 kDa corresponds to GFP and the other additional 2 kDa to the affinity tag.

The results obtained for cultures induced at an optical density of 0.6-0.8 are presented in Fig. 3.6 and 3.7, while those achieved for $OD_{600\text{ nm}} \sim 2$ are shown in Fig.3.8 and 3.9. The best conditions for protein expression are those obtained with IPTG concentrations between 0.2 and 1 mM after 22 h of induction (Fig. 3.6). However, for 1 mM IPTG a higher value of GFP-tagged fluorescence was obtained, corresponding to a higher yield of protein produced.

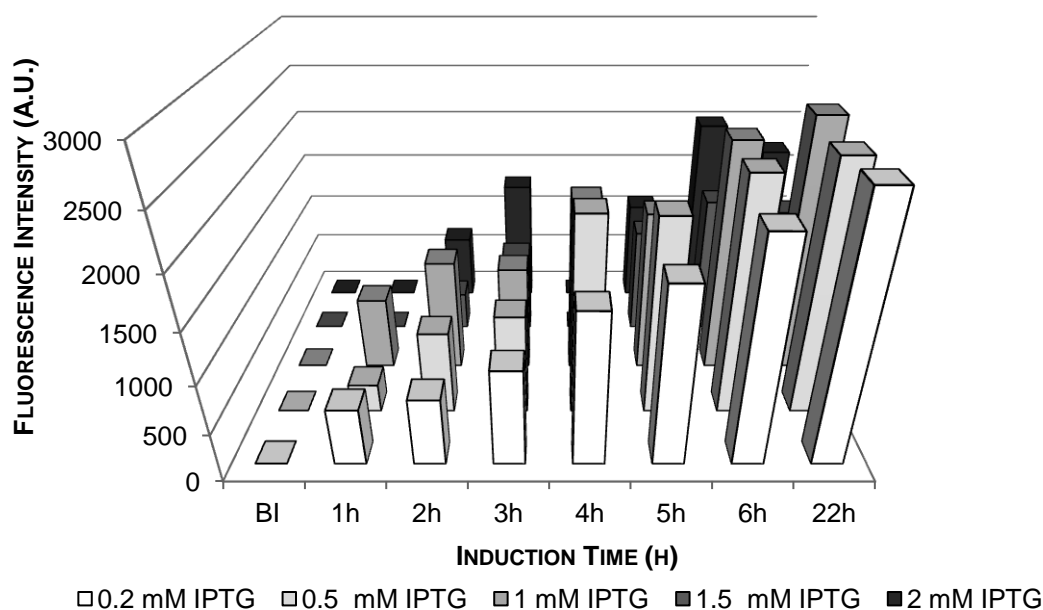


Figure 3.6 Quantification of RWRWRW tagged GFP by GFP fluorescence spectroscopy, produced by induction of gene expression at an optical density 0.6-0.8 in *E. coli* BL21(DE3) cultures. Each condition of IPTG used corresponds to unique assay of the target protein expression. The GFP fluorescence intensity values were normalized according with a constant optical density value (0.8).

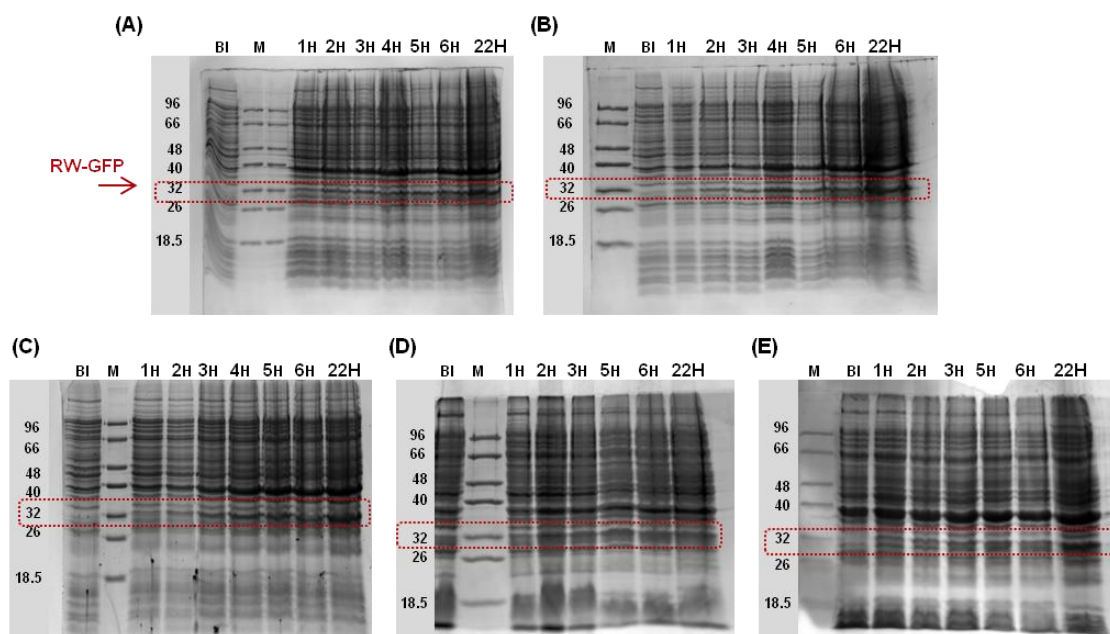


Figure 3.7 SDS-PAGE evaluation of the amount of RWRWRW tagged GFP produced over time, when using different concentrations of IPTG (A) 0.1 mM, (B) 0.5 mM, (C) 1 mM, (D) 1.5 mM and (E) 2 mM at optical density 0.6-0.8. M corresponds to the low molecular weight marker (Nzytech); BI, is the sample collected at t=0 and all other lanes represent the protein profile at different times of induction (1 h, 2 h, 3 h, 4 h, 5 h, and 22 h). The band of RWRWRW- tagged GFP is expected to have an apparent molecular mass of ~31 kDa and without tagging ~29 kDa. The volume of culture applied on the gel was normalized to a final optical density of 1.2. The protein expression was carried out at 37°C with 220 rpm. The gels present 12.5% of acrylamide and were stained with Coomassie Blue R-250.

The analysis of SDS-PAGE confirmed an increase of the amount of RWRWRW tagged GFP fusion protein produced with induction time (Fig. 3.6), as one can observe at the 22 h induction, for all the IPTG concentrations tested. In particular, the conditions that present a more intense GFP-tagged band are Fig. 3.7 (B), (C) and (E), corresponding to IPTG concentrations of 0.5, 1 and 2 mM. This last one is not corroborated by the GFP fluorescence intensity values obtained. However, the SDS-PAGE evaluation is not conclusive on the distinction of the best condition for a higher yield of target protein production. Therefore, the best condition when using lower optical density values was mainly selected by the fluorimetric analysis. The best IPTG concentration for this condition ($OD_{600\text{ nm}}$ 0.6-0.8) was 1 mM.

The results obtained for the production of GFP-tagged when inducing expression at higher optical density values did not lead to a higher yield of target protein. This conclusion is corroborated by the GFP fluorescence intensity values obtained for the different samples that are lower (Fig. 3.8) than the previous ones (Fig. 3.6). Moreover, the bands obtained for GFP-tagged are also not as intense as those obtained in previous results.

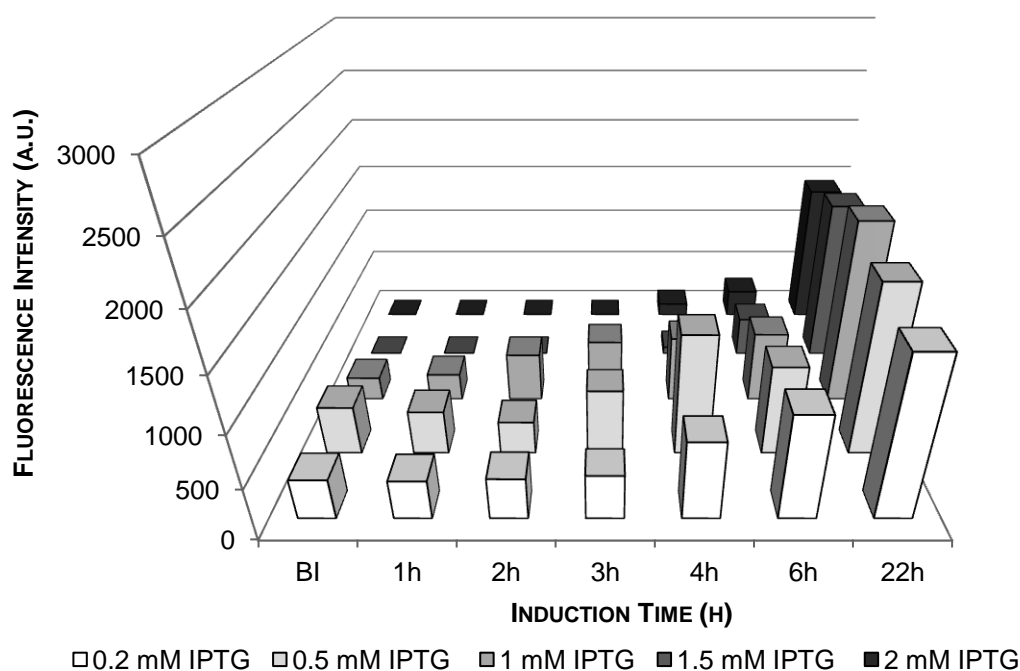


Figure 3.8 Monitoring of GFP Fluorescence intensity when inducing the RWRWRW tagged GFP expression at optical density of 2. The GFP fluorescence intensity values were normalized to a constant optical density value (0.8).

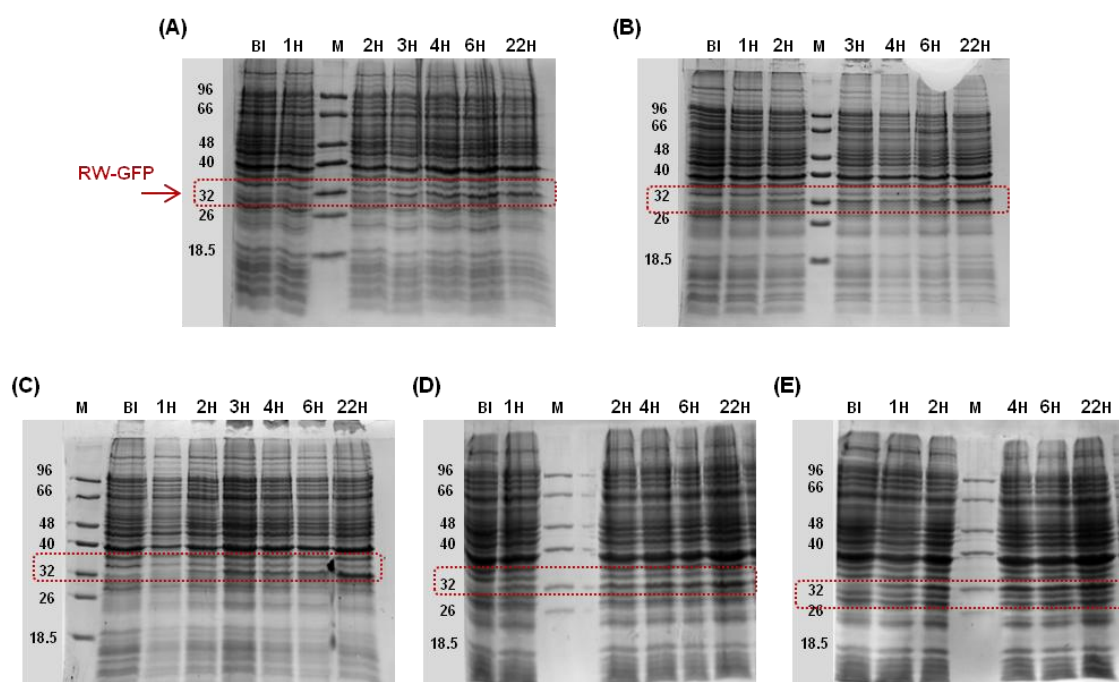


Figure 3.9 SDS-PAGE evaluation of the amount of RWRWRW tagged GFP produced over time, when employing different concentrations of IPTG (A) 0.1 mM, (B) 0.5 mM, (C) 1 mM, (D) 1.5 mM and (E) 2 mM at optical density 2. M corresponds to the low molecular weight marker (Nzytech); BI, is the sample collected at $t=0$ and all other lanes represent the protein profile at different times of induction (1 h, 2 h, 3 h, 4 h, 5 h, and 22 h). The band of RWRWRW- tagged GFP is expected to be ~ 31 kDa and without tagging ~ 29 kDa. The volume used for the sample application was normalized for a final optical density of 1.2. The protein expression was carried out at 37°C with 220 rpm. The gels present 12.5% of acrylamide and were stained with Coomassie Blue R-250.

Overall, the best condition found for the RWRWRW tagged GFP protein production was using an optical density of 0.6-0.8 with a concentration of IPTG of 1 mM during 22 h at 37°C and 210 rpm. This condition was then used to all other systems and the GFP tagged protein production was also evaluated by the techniques described above. Other aspect that should be highlighted from these studies is that the fluorescence techniques are useful reporters for the online detection of the GFP-fusion proteins. These methods allow taking measurements of the intracellular fusion protein production in real time and in a simple and accurate form, when compared with other methods as SDS-PAGE analysis. This can contribute for the optimization of the biochemical engineering processes.

3.5. LARGE SCALE EXPRESSION OF GFP TAGGED PROTEINS

3.5.1 Production of GFP tagged proteins

The large scale production of each affinity tag fused to GFP was performed at the optimal conditions determined on the small scale studies. During the expression, the production of different GFP tagged proteins was monitored by fluorimetric and optical density measurements (Fig. 3.10 and 3.11). SDS-PAGE analysis was also performed in order to visualise the relative

amount of protein in the various fractions. For a direct comparison of the amount produced by different systems, the loaded fractions at various induction times were normalized to a constant specific optical density value. The results obtained are shown in Fig 3.12. There is a direct correlation between the optical density and the GFP fluorescence intensity with increasing induction times (Fig. 3.10). However, for a direct comparison of the production of GFP fused to different tags, a normalization of the fluorescence intensity with the respective optical density measurements was performed and shown in Fig.3.11. This normalization allows a better evaluation of the GFP fluorescence intensity over time for all tags and present the following order in terms of GFP production: NNNNNN tagged GFP > GFP > NWNWNW tagged GFP > WFWFWF tagged GFP > RKRKRK tagged GFP > RWRWRW tagged GFP. The RWRWRW tagged GFP, is produced at a lower yield, the optical density starts decreasing after 6 h of induction, while in the other production systems the optical density tends to increase or stabilize. The other affinity tags induce lower levels of GFP expression when compared with the system containing the non-tagged GFP, except for the NNNNNN affinity tag. Due to the high hydrophilic character of this tag with the highest negative value of logP (§ 4.2), it increases the hydrophilicity of the fusion protein, which may increase the solubility of the fusion protein. The SDS-PAGE analysis also corroborates these observations regarding the GFP production with the different affinity tags fused to GFP.

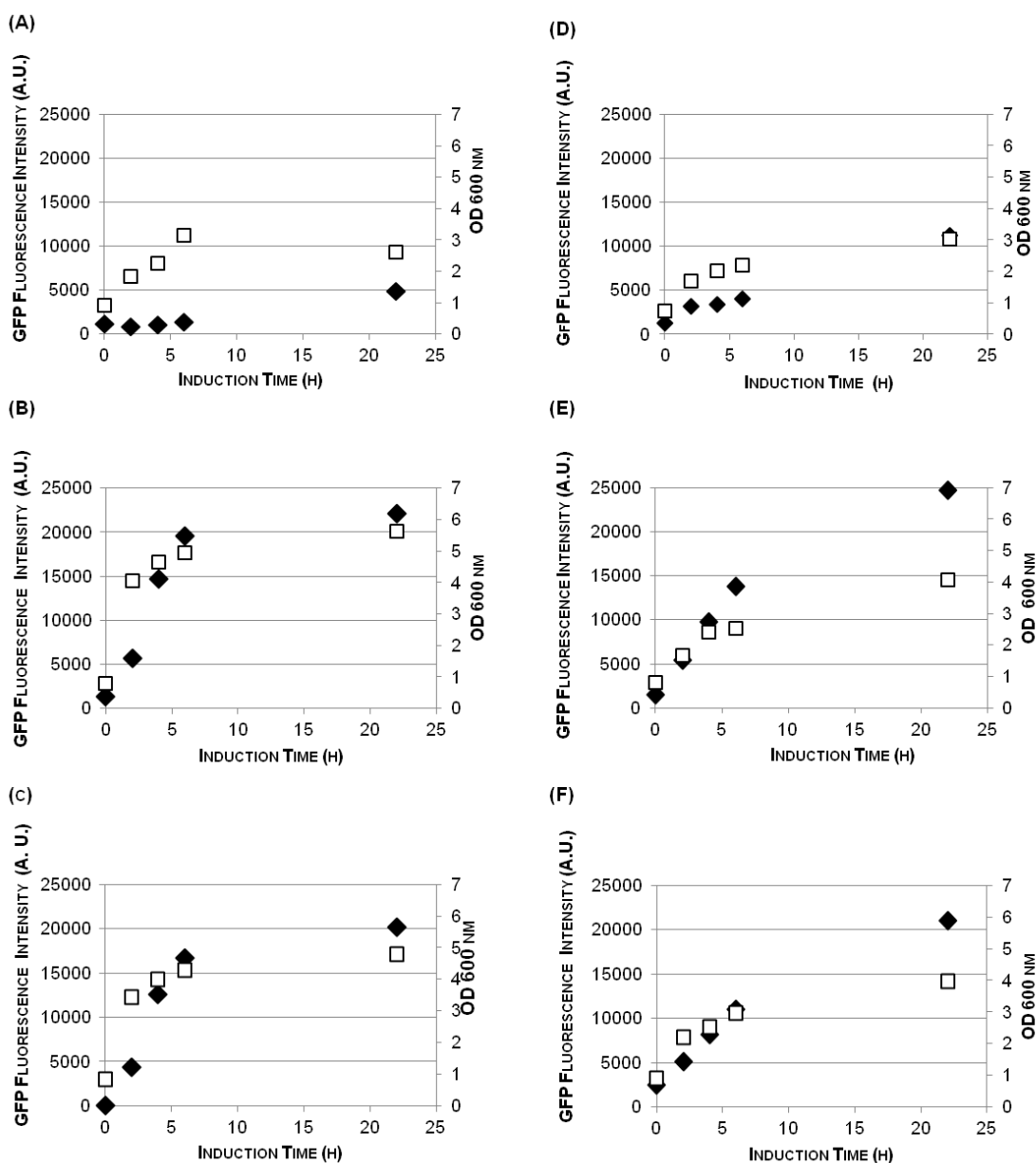


Figure 3.10 Correlation between GFP fluorescence intensity and optical density measurements with induction time during the large-scale expression of (A) RWRWRW tagged GFP, (B) WFWWF tagged GFP, (C) NWNWNW tagged GFP, (D) RKRKRKR tagged GFP, (E) NNNNN tagged GFP and (F) GFP on *E. coli* BL21(DE3) bacterial cells. ◆ GFP fluorescence: □ OD_{600 nm}

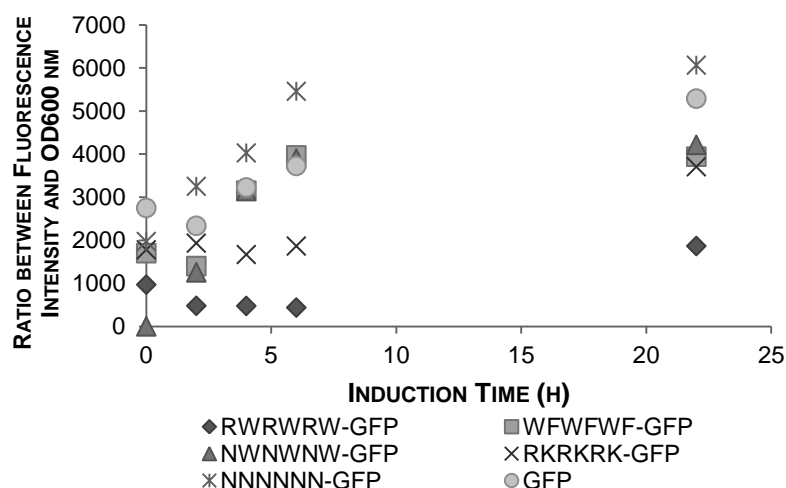


Figure 3.11 Normalization of the GFP fluorescence intensity by the respective optical density measurements. The fluorescence intensity and optical density measurements were performed over time after addition of IPTG (0 h, 2 h, 4 h, 6 h and 22 h) through 96-well fluorimetry and spectrophotometric analysis, respectively.

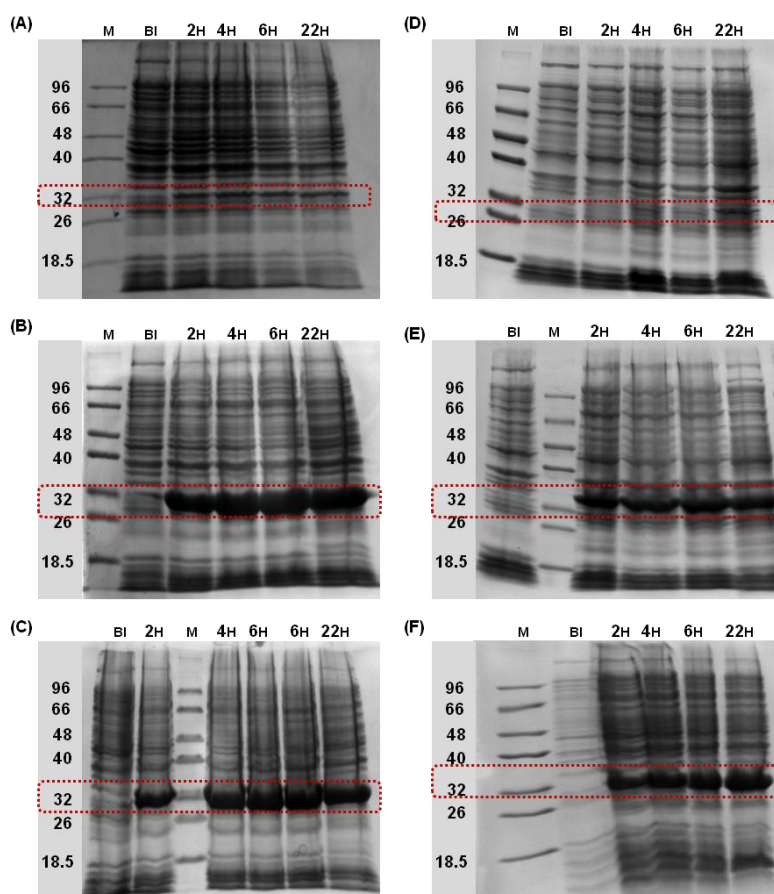
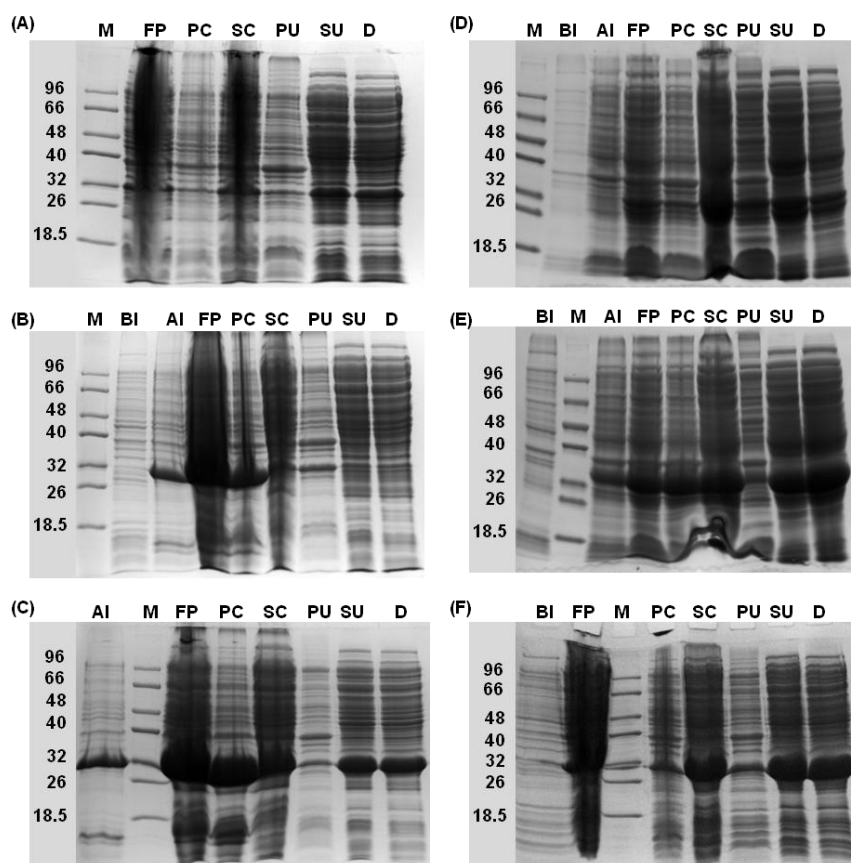


Figure 3.12 SDS-PAGE time course of large scale production of GFP tagged proteins in *E. coli* BL21(DE3). (A) RWRWRW tagged GFP; (B) WFWFWF tagged GFP; (C) NWNWNW tagged GFP; (D) RKRKRKR tagged GFP; (E) NNNNNN tagged GFP; and (F) GFP. M represents the low molecular weight marker (Nzytech); BI, is the sample collected at t=0 and all other lanes represent the protein profile at different times of induction (2h, 4h, 6h and 22h). The band of GFP tagged is expected to be ~31kDa and without tagging ~29 kDa. The volume used for the sample application was normalized for a final optical density of 1.2. The gels present 12.5% of acrylamide and stained with Coomassie Blue R-250.

Once the proteins were produced by the overexpression system, cell harvest and mechanical lysis with French press (4x 800 psi) were required to extract the intracellular proteins. Proteins were then fractionated through several steps of centrifugation to remove cell debris and membrane fraction. The clarified soluble extract was finally dialyzed against the buffer used in the affinity chromatography binding step. In all steps of protein fractionation, an aliquot of each sample was taken for analysis by SDS-PAGE (Fig. 3.10). After the last step of fractionation, all GFP-tagged proteins were quantified by GFP fluorescence and BCA assay to quantify the amount of GFP-tagged produced over the total protein present in soluble crude extracts. The results obtained are shown in Table 3.1.

Analysis of all SDS-PAGE gels (Fig. 3.13) shows an intense band on the supernatant lane, which corresponds to soluble GFP-tagged, except in Fig. 3.13 (B). For the NWNWNW tagged GFP there is a significant amount of protein in the pellet and supernatant after centrifugation. This means that protein has been produced both in soluble and insoluble forms. The amount of soluble protein was enough to proceed with binding assays.



3.13 Cellular fractionation analyzed by SDS-PAGE of (A) RWRWRW tagged GFP, (B) WFWFWF tagged GFP, (C) NWNWNW tagged GFP, (D) RKRKRKR tagged GFP, (E) NNNNNN tagged GFP and (F) GFP. (M) corresponds to the low molecular weight (Nzytech); (BI) the samples collected before induction; and (AI) the sample after 22 h of induction; (FP), after lysis through a French Press; (PC) the resuspended pellet after low-speed centrifugation; (SC) the supernatant after the low-speed centrifugation. This supernatant fraction was then ultracentrifuged, and the pellet (PU) and supernatant (SU) were obtained. The supernatant was dialyzed (D) against the binding buffer required for screening assays in chapter 5). GFP tagged proteins have an apparent molecular mass of ~31 kDa while the non-tagged GFP has ~29 kDa. The volume used for the sample application was normalized for a final optical density of 1.2. The gels were 12.5% in acrylamide and stained with Coomassie Blue R-250.

Table 3.1 Percentage of tagged and non-tagged GFP produced over the amount of total protein produced by *E.coli* cells. The percentage of GFP was determined by GFP fluorescence intensity measurements and total protein through the colorimetric method BCA assay.

Crude Extract	Total Protein (mg/ml)	GFP (mg/ml)	% GFP
RWRWRW tagged GFP	8.55 ± 0.13	0.18	2
WFWFWF tagged GFP	Formation of Inclusion Bodies		
NWNWNW tagged GFP	5.58 ± 0.03	0.48	9
RKRKRK tagged GFP	7.63 ± 0.24	0.33	4
NNNNNN tagged GFP	16.81 ± 0.53	0.91	5
GFP	4.97 ± 0.17	0.66	13

The low amount of RWRWRW-GFP produced can indicate that this particular affinity tag can be a burden for the cell. It is known that cationic peptides with arginine-tryptophan (RW) tandem repeat peptide sequences are recognized as antimicrobial peptides (AMPs) [212], presenting an optimal antimicrobial activity when used with three repeats (RWRWRW, (RW)₃) [213-214]. (RW)₃ repetitions present bactericidal effect against different microorganisms such as *S.aureus*, *S. epidermidis*, *E. coli*, *P. aeruginosa*, and *K. pneumonia* [213]. It is thought that the mechanism of action of these peptides starts by the interaction with the bacterial cell membrane, by establishing electrostatic interaction while the cationic side chains of arginine, whereas the non-polar side chains of the tryptophan residue might induce membrane disruption [212, 215]. However, the exact mode of action of these peptides is not yet fully understood and some models regarding membrane permeabilization and non-permeabilization have been postulated [215]. However, in this particular work, the AMP (RW)₃ is being produced intracellularly and might destabilize the host cell due to its toxicity leading to cell death [216]. AMPs can be produced when fused to a carrier protein in order to mask their toxicity [216-217]. In this work,

GFP acted as the carrier protein and the potential of this system for $(RW)_3$ production was further evaluated (§ 3.5.2).

Regarding the production of WFWFWF tagged GFP, there is a higher amount of protein in the pellet sample after centrifugation, when compared with the amount of protein in the respective supernatant (soluble form) (Fig. 3.13), indicating that this fusion protein was produced as inclusion bodies (IBs) [187], which was confirmed through observation of the pellet under fluorescence microscope (Fig. 3.14). The images shown in Fig. 3.14 reveal bright cylinder spots indicating that portion of the protein is proper folded and functional. The diameter of these spots was estimated with Image J, by analyzing all the spots in the image. The average diameter of these cylinder spots was $0.45 \pm 0.09 \mu\text{m}$, while the range of IBs which are bodies with diameters within the range between $0.2\text{--}0.5 \mu\text{m}$ [166, 190, 218]. In order to proceed with WFWFWF tagged GFP, two different approaches were carried out: the solubilization of IBs followed by protein refolding (§ 3.5.3.1); and the avoidance of IBs formation (§ 3.5.3.2) by using different conditions of protein expression such as temperature and inductor levels.

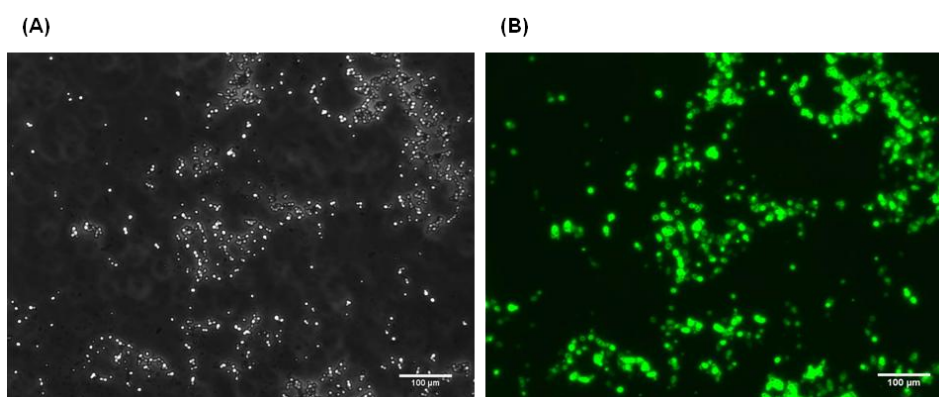


Figure 3.14 Inclusion bodies obtained for the WFWFWF-GFP production system. The images were taken on the fluorescence microscope under (A) bright field filter and (B) fluorescence filter. The inclusion bodies are present in the pellet obtained from the low-speed centrifugation step during WFWFWF tagged GFP cellular fractionation. These images are representative of five images taken of the sample field with 1000x amplification.

All soluble crude extracts of the five affinity tags fused to GFP were then used for the screening with the second generation combinatorial library that contains 64 different affinity ligands immobilized on agarose (Chapter 5).

3.5.2. Antimicrobial properties of the RWRWRW-GFP system

The hexapeptide RWRWRW has been described in the literature as an AMP. In this work, different tests were performed to evaluate the antimicrobial behaviour of $(RW)_3$ when incubated with *E. coli* K12 and also when produced as a fusion protein with GFP in *E. coli* BL21(DE3) cells.

3.5.2.1 Antimicrobial assays with $(RW)_3$

The antimicrobial effect of the peptide RWRWRW was conducted in a sterile 2 ml 96-well block, where different stock peptide concentrations were incubated with *E. coli* K12 in order to evaluate the minimum inhibitory concentration (MIC) values. This parameter is defined as the minimum concentration of an antimicrobial substance that inhibits the growth of a microorganism under specific conditions [219]. The antimicrobial assays were performed by using a HTS method in order to improve the variability of the conditions used and for cost-and-time effectiveness. The monitoring of cell growth under different concentrations of the antimicrobial agent was conducted by optical density measurements at 600 nm every 30 min. The results are shown in Fig. 3.15.

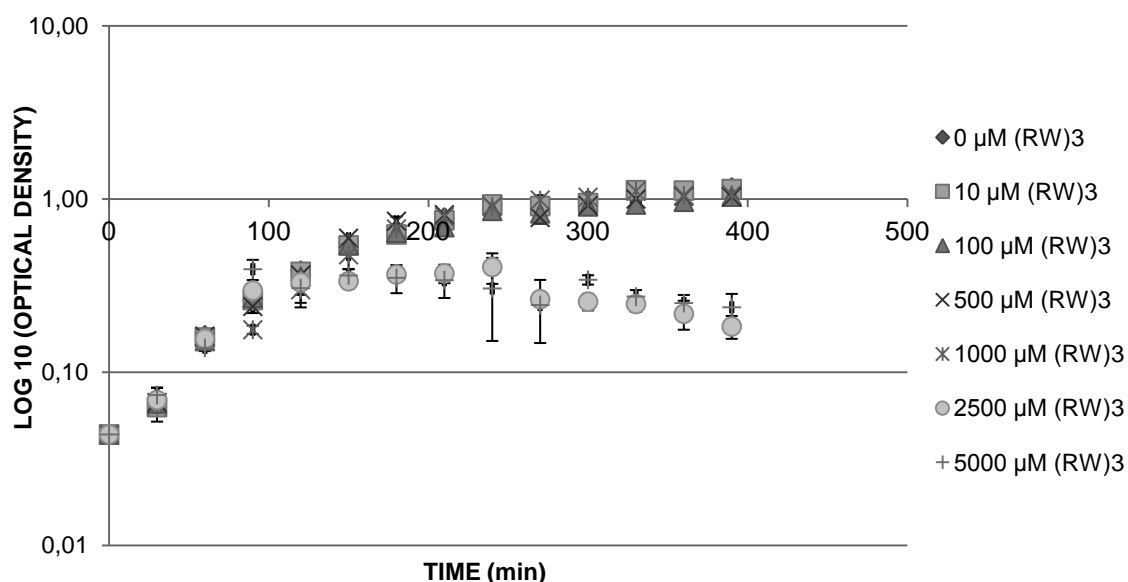


Figure 3.15 Antimicrobial assays between different RWRWRW concentrations and *E. coli* K12. These assays are useful for determining of minimum concentration required for RWRWRW to inhibit the bacterial host growth (MIC values). An overnight grown culture of *E. coli* K12 was diluted for an optical density of 0.045, while the cells were in lag phase. The diluted culture was added to each well of a 96-well block and then incubated at 37 °C with stirring until optical density reached 0.1-0.15. From this point, different peptide concentrations were added to each well and then optical density measurements were taken every 30 minutes in order to monitor cell growth. The error bars corresponds to triplicates of the antimicrobial assays.

When using antimicrobial agents, two types of effects can be observed: bacteriostatic and bactericidal effects. In the former, the antimicrobial agent only inhibits the microorganism growth and this latter cell death is observed [220]. The results suggest that from 2.5 mM of $(RW)_3$ there is a decay of the optical density measurements suggesting the bactericidal effects of this peptide. For concentrations of $(RW)_3$ between 10 and 1000 μM *E. coli* K12 cells tolerate the presence of the antimicrobial agent during growth, as curves are superimposed with the controls.

3.5.2.2. Production of RWRWRW tagged GFP in *E.coli* BL21(DE3) cells

Small scale studies of the expression of RWRWRW tagged GFP were performed with a more detailed monitoring of optical density and fluorescence intensity (Fig. 3.16). GFP has already been described as a possible carrier protein for the expression of AMPs masking their toxicity and preventing their proteolytic degradation [216-217, 221]. Previous works indicate that the carrier protein GFP was not efficient to mask the toxicity of AMPs in bacterial cells, leading to the formation of inclusion bodies [221]. This effect is possibly a host defence mechanism when cells produce antimicrobial peptides.

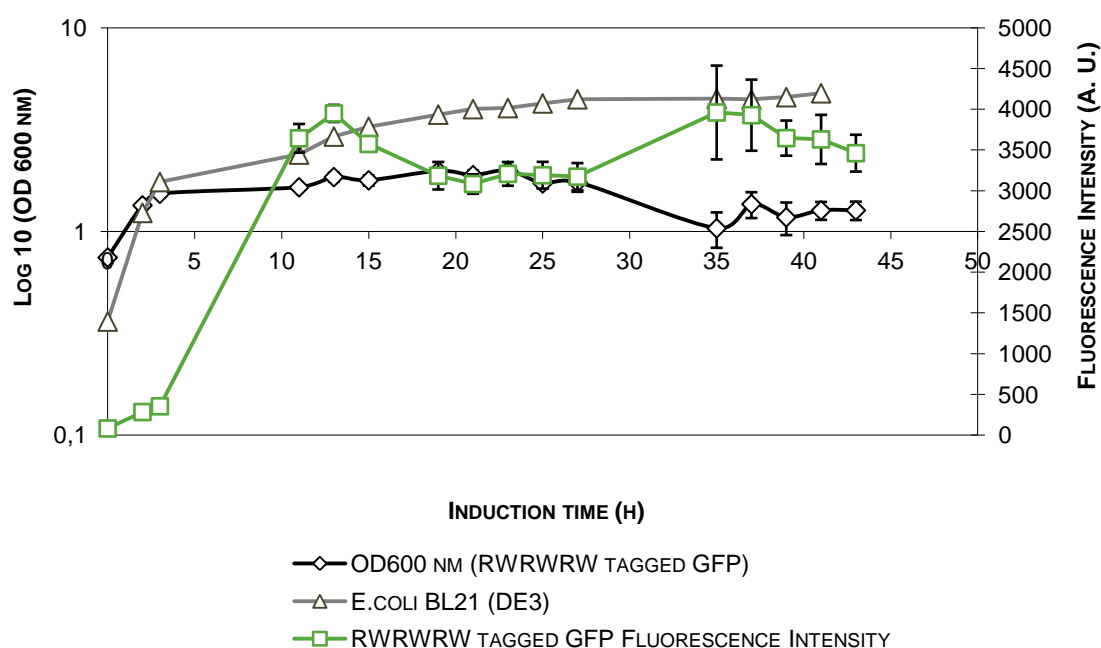


Figure 3.16 Monitoring RWRWRW tagged GFP production on *E. coli* BL21(DE3) cells through GFP fluorescence intensity and optical density measurements over induction time and growth curve for *E. coli* BL21(DE3) cells. The protein expression was induced at 0.6-0.8 with 1 mM IPTG and carried out at 37 °C with 220 rpm as well for *E. coli* BL21(DE3) cell growth. Regarding, the RWRWRW tagged GFP, the samples were collected after the addition of IPTG (0h – 43 h) and then quantified through 96-well fluorimetry and spectrophotometric analysis for the respective monitorization of GFP fluorescence intensity and evaluation of optical density of the cell suspension. The *E. coli* BL21(DE3) cell growth was also monitorized by spectrophotometric analysis. The time 0h corresponds to time after induction for RWRWRW tagged GFP expression and 8h after inoculum for *E. coli* BL21(DE3) cell growth. The error bars correspond to three assays of the expression of RWRWRW tagged GFP.

Overall, the production of the AMP RWRWRW fused to GFP is possible to perform by using bacterial hosts. The use of GFP as a carrier protein seems to be efficient as there is no formation of inclusion bodies. The maximum fluorescence intensity is observed at 15 h and this should be the optimum time to achieve the highest yield in protein production. By analysing Fig. 3.16, it is visible that OD decays approximately 41% between 25 and 35 h after induction, which may be related with the high toxicity provoked by the produced peptide. This is corroborated by *E. coli* BL21(DE3) growth profile on Fig. 3.17, that indicates that at this specific time, the cells

were still in stationary phase but not in death phase. At the same time that the OD values decrease, there was an increase of GFP fluorescence intensity. This can be explained by the release of RWRWRW tagged GFP to culture medium when cell death occurs, and GFP fluorescence was no longer influenced by the bacterial cell wall. In this situation, the purification of RWRWRW tagged GFP could be facilitated as no further steps of cell lysis were required. However, further studies would be necessary to corroborate this hypothesis.

3.5. 3 Production of inclusion bodies with the WFWFWF tagged GFP system

3.5.3.1 Solubilization and Refolding Strategies

The WFWFWF tagged GFP protein was produced as inclusion bodies and attempts were made to solubilise and refold the protein [190]. Protein solubilization was performed with chaotropic agents such as urea, that leads to a random coil structure and loss of secondary structure, therefore exposing protein hydrophobic surfaces [222]. The solubilisation buffer also includes other additives such as dithiothreitol (DTT), ethylenediaminetetraacetic acid (EDTA) and glycerol, where DTT and EDTA prevent unwanted or incorrect disulfide bonds formation and air oxidation during refolding, respectively [190, 222]. Glycerol is also considered an excellent refolding additive [190]. The solubilization was carried out at room temperature according with previous reports [190]. After protein solubilisation, refolding was carried out at 4 °C through subsequent dialysis with small decrements of urea until complete removal of the chaotropic agent. Unfortunately, the fusion protein precipitated and fluorescence was lost during the last dialysis step. This phenomenon was reversible as GFP's fluorophore was restored in the presence of the buffer with the lowest amount of urea (0.5 M urea). Since the refolding attempts were not successful, different solubilisation buffers were employed. These contained the same compounds but with higher salt concentration and slightly higher pH. In addition, the refolding was conducted with different temperatures namely 4 °C and room temperature. According to the literature, the pH used should be at least one unit far from the isoelectric point [190]. Furthermore, salt concentrations used should be a balance between very low concentrations which contribute for salting-in effects and therefore increase the protein solubility, and high concentration that can lead to protein precipitation [165, 190]. Regarding the refolding temperature, there is no optimal temperature to restore native tertiary structure of the protein, however in most cases, the refolding is conducted at room temperature [190, 222]. However, none of these strategies were successful. The last attempt was to perform the refolding on-column, which is further described in Chapter 5 (§ 5.6.1).

3.5.3.2 How to avoid inclusion bodies?

Considering that refolding of WFWFWF tagged GFP was not achieved, different strategies were explored to minimize the formation of inclusion bodies. The high level of protein expression in *E. coli* can lead to the formation of insoluble protein aggregates as inclusion bodies. These aggregates are formed by unfolded or highly misfolded polypeptides [165, 190]. In order to improve correct folding, recombinant proteins should be secreted to the periplasmic space through a peptide leader sequence on the N-terminal, since molecular chaperones and a reducing environment are found in the periplasm [165, 191, 193, 223-224]. Inclusion bodies can be avoided by changing growth conditions such as temperature, inducing agent levels and time, *E. coli* strain [164, 190-191, 223].

Therefore the study of the impact of different expression conditions on the overall production of WFWFWF tagged GFP and on the ratio between soluble GFP fusion protein and IB formation was carried out in two different stages. The first was characterized by the inductor levels influence, where different concentrations of IPTG (0.1, 0.5 and 1 mM) were used for two distinct temperatures (30 and 37 °C) to which the cell cultures were subjected after the addition of IPTG. After selecting the optimal level of IPTG, this condition was employed on the study of temperature influence (27, 30, 33 and 37 °C) during the expression of the target protein, once again after IPTG addition. The production of WFWFWF tagged protein and the ratio between soluble and insoluble protein produced was monitored by optical density and GFP fluorescence intensity measurements as well as by analysis of SDS-PAGE gels.

3.5.3.2.1 Inductor levels Influence

The results showing the impact of IPTG levels on IB formation are shown in Figs. 3.17, 3.18 and 3.19. Fig. 3.17 displays the evolution of GFP fluorescence intensity normalized with the respective optical density measurement over the induction time for different inductor concentrations and temperatures. There was an increase of GFP tagged production with the induction time, and this amount was higher when performing the expression at 30 °C (Fig. 3.17 (A)).

In order to infer on the amount of protein produced in soluble and insoluble forms, the cells were lysed by French press mechanical process and then centrifuged to analyse the supernatant and the pellet [187]. The results obtained are shown in Figs. 3.18 and 3.19. The analysis of the SDS-PAGE gels (Fig. 3.18), confirmed that the induction with IPTG was successful (as no GFP protein was visible before induction), and that the majority of the protein was produced in insoluble form as the GFP protein is clearly present in the pellet lanes (P). The results from Fig. 3.19 indicate that although the amount of tagged GFP produced in insoluble form was higher,

the fluorescence intensity of these samples was lower when compared with that obtained from the soluble form. This can be explained by the fact that IBs are mostly formed by unfolded or highly misfolded polypeptides [166, 190] and the fluorescence intensity of GFP in IBs (pellet) is not representative of the real amount of protein. Quantifications of soluble protein seem more accurate.

Overall, these studies indicate that the lowest temperature (30 °C) is more adequate for protein expression and that the amount of protein produced in both forms is higher for IPTG concentrations of 0.5 mM and 1 mM.

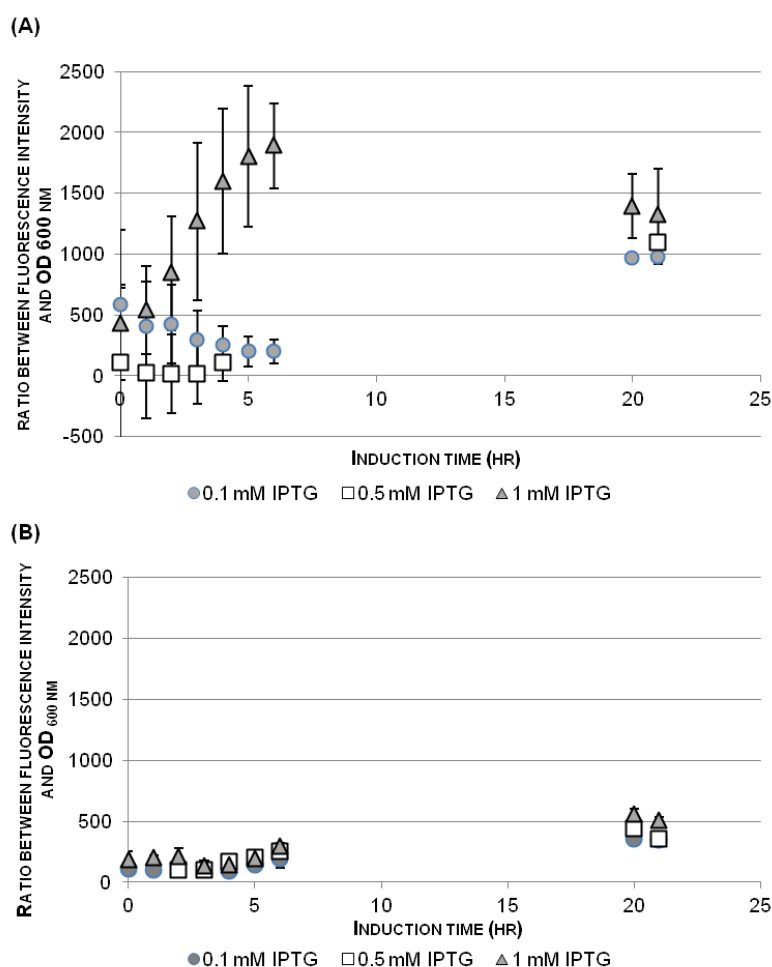


Figure 3.17 Impact of IPTG concentration on the expression of WFWFWF tagged GFP at (A) 30 °C and (B) 37 °C. The WFWFWF tagged GFP expression was induced at OD 0.6-0.8 with three different concentrations of IPTG (0.1, 0.5 and 1 mM), and then carried out at different temperatures. Samples at different induction times were characterised by GFP fluorescence intensity and optical density measurements. The error bars corresponds to triplicates of the WFWFWF tagged GFP expression assays.

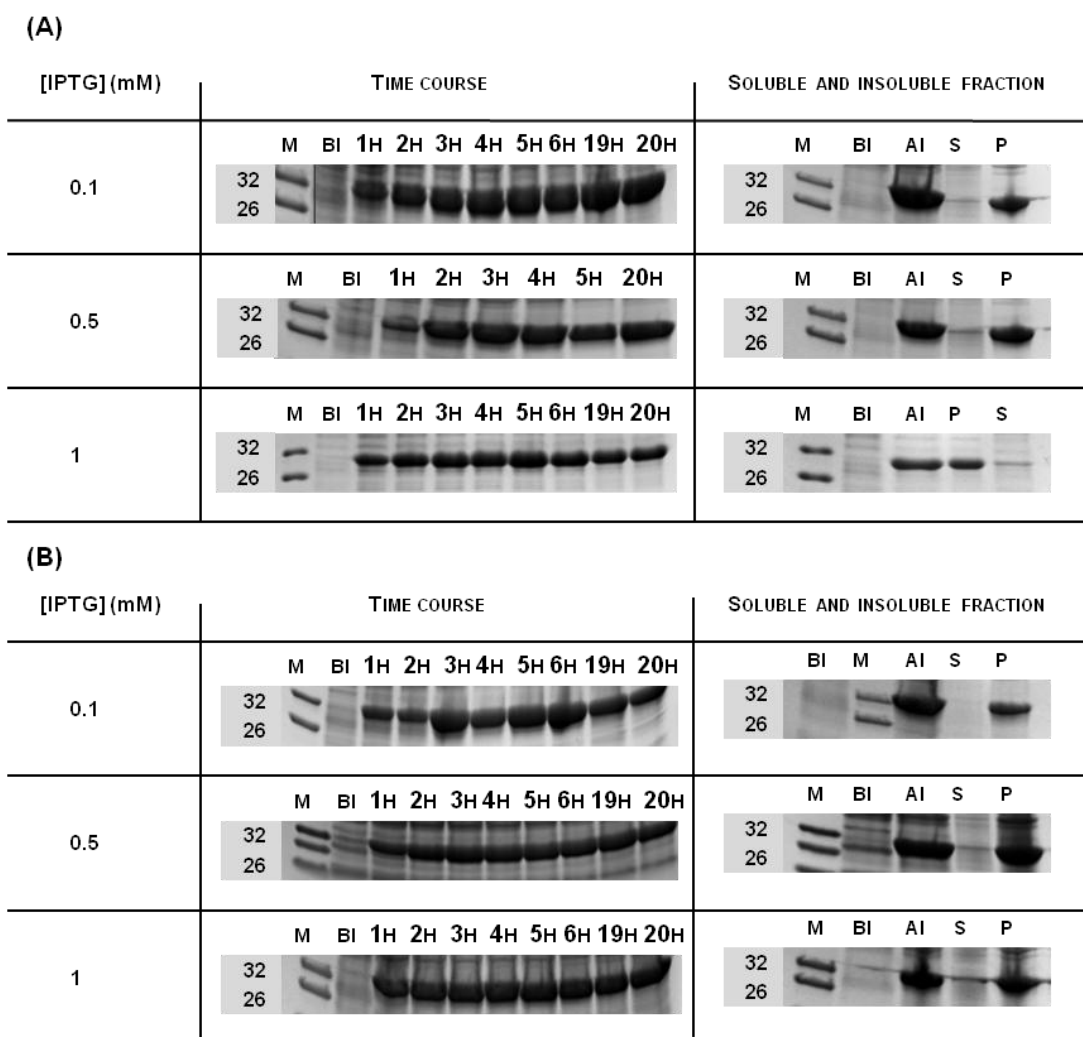


Figure 3.18 SDS-PAGE analysis of the soluble and insoluble WFWWF tagged GFP produced over the induction time for different concentrations of IPTG at (A) 30 °C and (B) 37 °C. The lanes in the “time course” correspond to: the lower molecular weight (Nzytech) (M), samples taken before induction (BI) and at different times of induction (1 h, 2 h, 3 h, 4 h, 5 h, 6 h and 20 h). After the expression of proteins (AI), the cells are lysed with French press, and then centrifuged to evaluate the soluble (S) and insoluble form (P). The band of GFP tagged is expected to be at ~31 kDa. The volume used for the sample application was normalized for a final optical density of 1.2. The gels present 12.5% of acrylamide and stained with Coomassie Blue R-250.

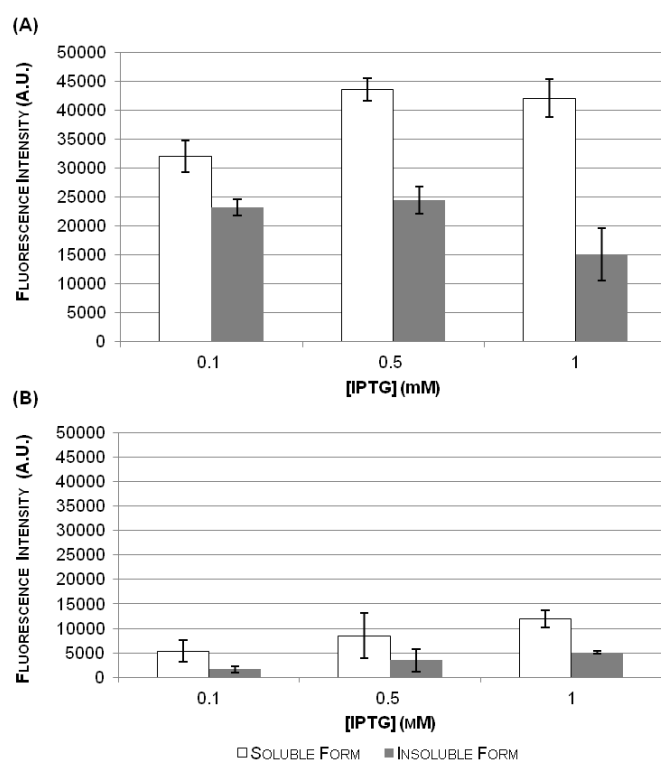


Figure 3.19 GFP fluorescence intensity of soluble and insoluble forms of WFVWFVW tagged GFP obtained after protein expression induced at 0.6-0.8 with different IPTG concentrations at (A) 30 °C and (B) 37 °C. The evaluation of soluble and insoluble form was performed after the expression of WFVWFVW tagged protein, in particular after the steps of cellular fractionation such as cell lysis and low-speed centrifugation. The error bars corresponds to triplicates of the WFVWFVW tagged GFP expression assays.

3.5.3.2.2 Temperature Influence

The influence of temperature during the expression of WFVWFVW tagged induced with 1 mM IPTG was studied by employing four different temperatures 27, 30, 33 and 37 °C. The results obtained from the fluorescence intensity measurements normalized with the respective value of optical density are shown in Fig. 3.20. These results show that with the increase of temperature, there is a higher amount of protein being produced and that 30 °C seems to be the best temperature condition, followed by 27 °C, 33 °C and 37 °C. As previously discussed, the analysis of GFP fluorescence intensity might not be representative of the real amount of protein produced in insoluble form. Therefore, it was necessary to analyse concomitantly the information given from Figs. 3.20-3.22.

When analysing the SDS-PAGE gels (Fig.3.21), it is clear that induction with IPTG was successful for all temperatures (as no GFP protein was visible before induction), and that the majority of the protein was produced in the insoluble form. Finally, data in Fig.3.22 clarify on the relative amounts of soluble and insoluble protein produced. The ratio between the fraction soluble and insoluble corresponds to 1.2, 2.5, 1.8 and 2.3 for the temperatures 27, 30, 33 and

37 °C, respectively. The ratio between the soluble and insoluble form seems to be more pronounced for 30°C. In general, at lower temperatures (27 and 30 °C), the production of soluble protein is similar and more effective than at higher temperatures. The condition T=27 °C seems to be more reproducible and accurate as lower error bars were obtained.

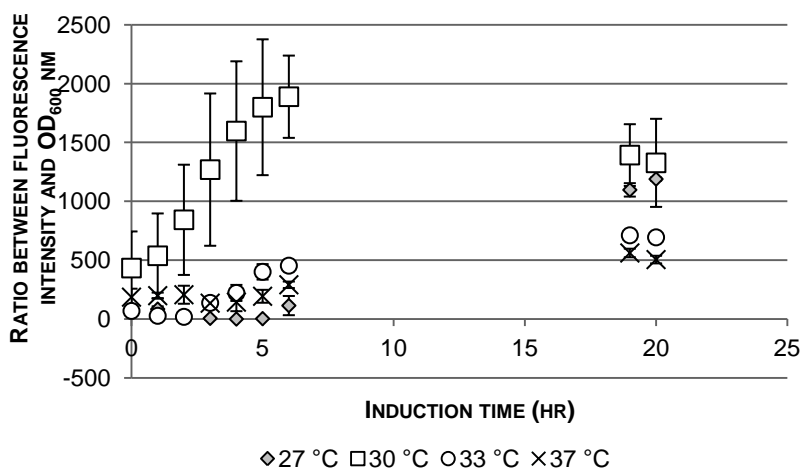


Figure 3.20 Impact of expression temperatures on the overall productivity of WFWFWF tagged GFP. The WFWFWF tagged GFP expression was induced at 0.6-0.8 with 1mM of IPTG and then carried out at different temperatures after the induction. The monitorization was carried out by GFP fluorescence intensity and optical density measurements. For a direct comparison between all samples of different induction times, the GFP fluorescence intensity was normalized with the respective optical density values. This corresponds to ratio between the fluorescence intensity and the optical density at 600 nm. The errors bars correspond to triplicates of WFWFWF tagged GFP expression.

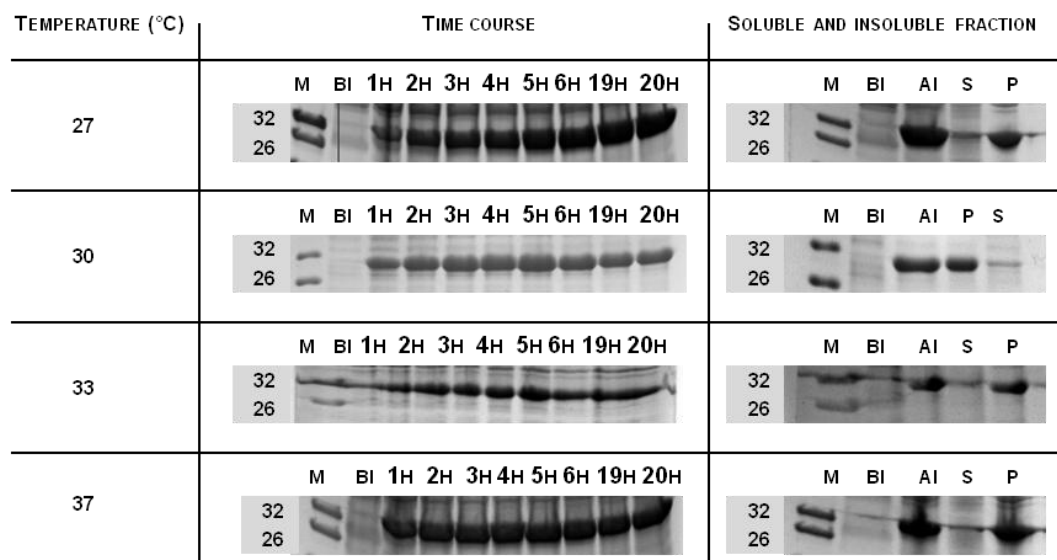


Figure 3.21 SDS-PAGE analysis of the soluble and insoluble WFWFWF tagged GFP produced over the induction time for different temperatures. The lanes in “time course” corresponds to: the lower molecular weight (Nzytech) (M), samples taken before induction (BI) and at different times of induction (1 h , 2 h, 3 h, 4 h, 5 h, 6 h and 20 h) . After the expression of proteins (AI), the cells are lysed with French press, and then centrifuged to evaluate the soluble (S) and insoluble form (P). The band of GFP tagged is expected to be at ~31 kDa . The volume used for the sample application was normalized for a final optical density of 1.2. The gels present 12.5% of acrylamide and stained with Coomassie Blue R-250.

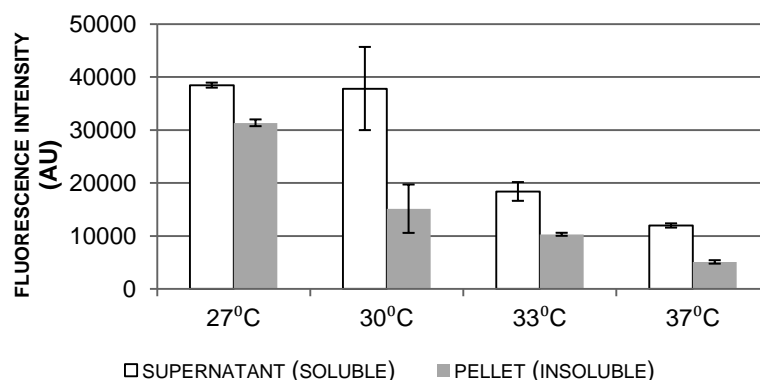


Figure 3.22 GFP fluorescence intensity of soluble and insoluble forms of WFWFWF tagged GFP obtained after protein expression induced at 0.6-0.8 with 1 mM of IPTG at different expression temperatures. The evaluation of soluble and insoluble form was carried out from the first step of centrifugation during protein fractionation after the expression of WFWFWF tagged protein and subsequent cell lysis. The ratio between the fraction soluble and insoluble corresponds to 1.2, 2.5, 1.8 and 2.3 for the temperatures 27, 30, 33 and 37 °C, respectively. The error bars correspond to triplicates of WFWFWF tagged GFP expression assays.

The samples correspondent to the pellets obtained for the various conditions were also visualised by fluorescence microscopy and presented in Fig.3.23. These images indicate that as the expression temperature increases, the formation of aggregates is more prone to occur. For high temperatures, the aggregates presented intense fluorescence indicating the presence of folded tagged GFP. However, these aggregates might interfere with the measurement of GFP fluorescence intensity.

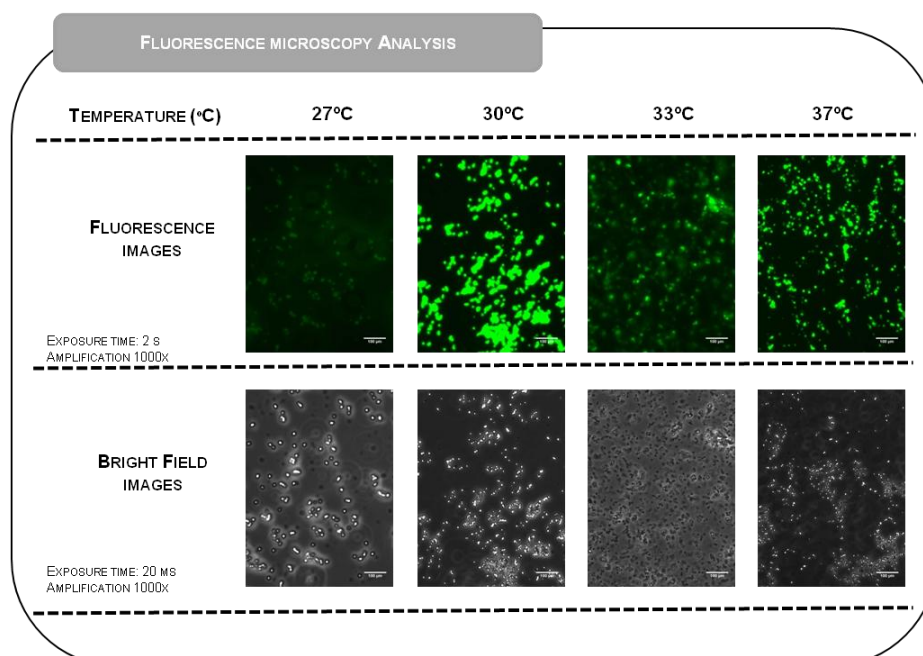


Figure 3.23. Fluorescence microscopy analysis of the insoluble form of WFWFWF tagged GFP produced at different temperatures with 1 mM IPTG. The insoluble form is related with the formation of inclusion bodies. This sample is obtained from the first step of centrifugation during protein fractionation after the expression of WFWFWF tagged protein and subsequent cell lysis. These images are representative of three images taken from the sample field and present a total amplification of 1000x.

Overall it is difficult to avoid the formation of inclusion bodies by controlling the temperature, for the WFWWF tagged GFP protein. However it was possible to increase the amount of soluble protein when employing lower temperatures. The temperature seems to have a higher impact than the inductor levels. Still it was possible to increase the amount of soluble protein produced by a factor of 10 when employing 27 °C and 1 mM IPTG for induction.

3.6. CONCLUDING REMARKS

The development of novel affinity tag started with the successful subcloning of the fragment DNA with the nucleotide sequence encoding the affinity tag fused to GFP in pET-21c expression vector. The proof of concept was performed for RWRWRW tagged GFP, and can be extended to other hosts, expression vectors and target proteins, since the designed fragment DNA included two multicloning sites, for the subcloning to other commercially available expression vectors and also to include nucleotide sequences codifying for other proteins. The production of the RWRWRW tagged GFP was carried out in bacterial cells, and an optimization of expression conditions were performed in order to maximize the yield of protein produced. The production of the five different affinity tags fused to GFP in *E. coli* cells presented different expression profiles (Table 3.2).

For the cases of GFP tagged to RKRKRK, NNNNN and NWNWNW protein expression is similar and a considerable amount of protein has been produced as soluble protein. These fusion proteins were produced under conditions optimized for the RWRWRW tagged GFP, and there is still scope to improve conditions individually for each tag system.

The expression of RWRWRW tagged GFP revealed to be a burden to the host cells and due to its antimicrobial properties can also be lethal after long induction times. Still, it was possible to produce enough soluble protein to carry out further steps of screening Ugi based combinatorial library. However, the antimicrobial properties of this affinity tag might interfere with its main application studied in this work

Table 3.2. Expression profiles summary of each affinity tag fused to GFP.

GFP tagged and non-tagged proteins	Expression Profile					
	Temperature (°C)	IPTG (mM)	GFP (mg/ml)	Total Protein (mg/ml)	Properties	
Hexapeptides	RWRWRW	37	1	0.18	8.55 ± 0.13	The tag presents antimicrobial properties
	WFWFWF	27		0.11	4.96 ± 0.21	Protein produced both as soluble and insoluble form
	NWNWNW	37		0.48	5.58 ± 0.03	Protein produced both as soluble and insoluble form
	RKRKRK	37		0.33	7.63 ± 0.24	Soluble protein
	NNNNNN	37		0.91	16.81 ± 0.53	Protein produced both as soluble and insoluble form
Non-tagged	GFP	37		0.66	4.97 ± 0.17	Soluble protein

The WFWFWF tagged GFP produced at the optimized conditions for RWRWRW tagged GFP, lead to the formation of IBs. As a first approach, the protein in the IBs was solubilised with chaotropic agents and different refolding strategies were attempted. However, no strategy was successful and the matrix assisted refolding was selected for further studies. IBs could be minimised by changing the growth culture temperature. Lower temperatures can contribute for the production of higher amounts of soluble protein. The NWNWNW tagged GFP was also produced as both soluble and insoluble for but the total amount of soluble protein was enough to carry out with further studies. A common feature of the affinity tags that produce GFP protein as IBs is the presence of tryptophan residues on their sequences and their hydrophobic character.

CHAPTER 4

COMBINATORIAL LIBRARY SCREENING WITH CRUDE EXTRACTS CONTAINING GFP-TAGGED AND NON-TAGGED PROTEINS

SUMMARY

A combinatorial library comprised of 64 affinity ligands based on the Ugi reaction was prepared and screened for binding to five GFP tagged proteins and non-tagged GFP, from crude extracts of *Escherichia coli* production. The five GFP tagged proteins used were RWRWRW-GFP, WFWFWF-GFP, NWNWNW-GFP, RKRKRK-GFP and NNNNNN-GFP. The screening was performed using affinity chromatography in a 96-well format, and lead ligands were identified for the GFP -tagged and non-tagged proteins. For the GFP- tagged proteins, the lead systems “tag-receptor” were: A4C8 for WFWFWF-GFP in denaturing conditions, A3C4 for NWNWNW-GFP and A7C1 for RKRKRKRK-GFP. The GFP tagged proteins, RWRWRW-GFP and NNNNNN-GFP were not selected for further studies as no lead ligands were identified. For non-tagged GFP, the lead ligand selected was A4C7.

4.1 INTRODUCTION

The development of synthetic affinity ligands for the purification of a target biologic comprises three main stages: rational design, solid-phase synthesis and high-throughput screening. The rational design involves *in silico* tools to develop a structure that mimics the functionalities of the natural ligand involved in biological recognition, or can be *de novo* designed for increased complementarity with the target protein [1-2]. In order to develop an affinity pair “tag-receptor” for the purification of fusion proteins, two first generation libraries of affinity ligands based on the triazine and Ugi-scaffolds have been previously *de novo* designed and screened on-bead by fluorimetric techniques toward the five hexapeptides RWRWRW, WFWFWF, NWNWNW, RKRKRK and NNNNNN [207-208].

The on-bead screening was initially carried out with the potential targets labelled with the fluorophore FITC (fluorescein isothiocyanate) on the C-terminal, and with two proteins selected as controls, namely GFP and Bovine Serum Albumin (BSA) labelled with FITC. The fluorescence-based techniques employed were Fluorescence microscopy and 96-well Spectrofluorimetry, which presented several limitations, namely the fact that the former is time-consuming, requires the individual analysis of the binding character of each agarose sample under the fluorescence microscope, and is difficult to implement as a HTS method; and the latter is associated with a high incidence of false negative results, and problems with resin pipetting including resin retention in the tip after dispensing, resin settling during the aliquotting procedure, and variability in the aliquotted slurry volume [225]. In addition, these first studies were performed with purified solutions of biomolecules conjugated to a fluorophore. Due to these limitations, this chapter focus on the selection of lead ligands by HTS affinity chromatography employing unlabelled and unpurified samples containing the target biomolecules.

The use of 96-well microtiter plate formats facilitates the screening of a wide range of different adsorbents and the use of different binding and elution buffers, such as buffer composition, pH, concentration and ionic strength. Furthermore, the required amounts of resin and target biomolecule are much less than required for other techniques [225-226]. Besides these advantages, microscale affinity chromatography in a 96- well format is less time consuming, is cost effective and leads to an acceleration of the downstream process development [225-226]. The main differences between 96-well and column formats are related to binding capacity measurements, automation and costs [226]. The binding capacities are different because the 96- well format is operated in batch mode, in which the target biomolecule and the resin are incubated under agitation; thereby generating a suspension of resin beads with the biomolecule feed solution. On column, this measurement is performed in a dynamic mode and is influenced by the flow rate [226]. Furthermore, the residence time and linear velocity are not representative

of industrial applications [227]. Another difference between the methods relates to the automation, as in the case of the 96-well format, it is essential to include a separation step to resolve the resin and the supernatant after incubation. This separation step may be achieved by vacuum filtration, pipette dispensing or by centrifugation [225]. Microscale 96-well formats have already been successfully employed in several applications for the purification of polyclonal antibodies by using a mixed mode cation exchange chromatography [226], in protein purification from cell-free broths [228], in optimization of the resin and elution buffer of protein A for monoclonal antibodies [229] and in the development of chromatographic methods for the purification of α -amylase [230]. Other applications of these methods also include the purification of virus-like particles and the development of peptide tags using metal affinity chromatography [231]. All of these examples are based on the incubation of a sample with the resin and Bergander *et al* [232] were the first to describe the microscale 96-well format for the estimation of the dynamic binding capacity as predictive of the behaviour of a preparative chromatography column [232]. Although the 96-well formats do not substitute for the use of pre-packed columns, they can lead to an indication of the optimal ligand candidate and chromatographic conditions, i.e. they can narrow the search for the conditions for the subsequent on-column studies.

This chapter describes the synthesis and screening of a combinatorial library of affinity ligands based on the Ugi reaction against GFP-tagged and non-tagged protein using microscale affinity chromatography in a 96-well format. A freshly prepared combinatorial library was utilised for the screening with each target protein. The GFP-tagged proteins presented the tag sequence, which included five different tags: RWRWRW, WFWFWF, NWNWNW, RKRKRK and NNNNNN. The non-tagged GFP was the same protein minus the tag. The evaluation of the lead ligands was determined according to Fig. 4.1. A preliminary screening was performed in order to evaluate the putative lead ligands by determining the enrichment factor for each affinity ligand and the respective target biomolecule, while comparing with the theoretical maximum enrichment factor for each system. Putative lead ligands were re-screened using the 96-well format to select the lead ligands for each target biomolecule. In this last step of selection, the purity of each fraction was qualitatively determined by SDS-PAGE analysis and the binding percentages of GFP proteins (tagged and non-tagged) and total protein were assessed for each of the putative lead ligands by measurements of fluorescence intensity and colorimetric techniques.

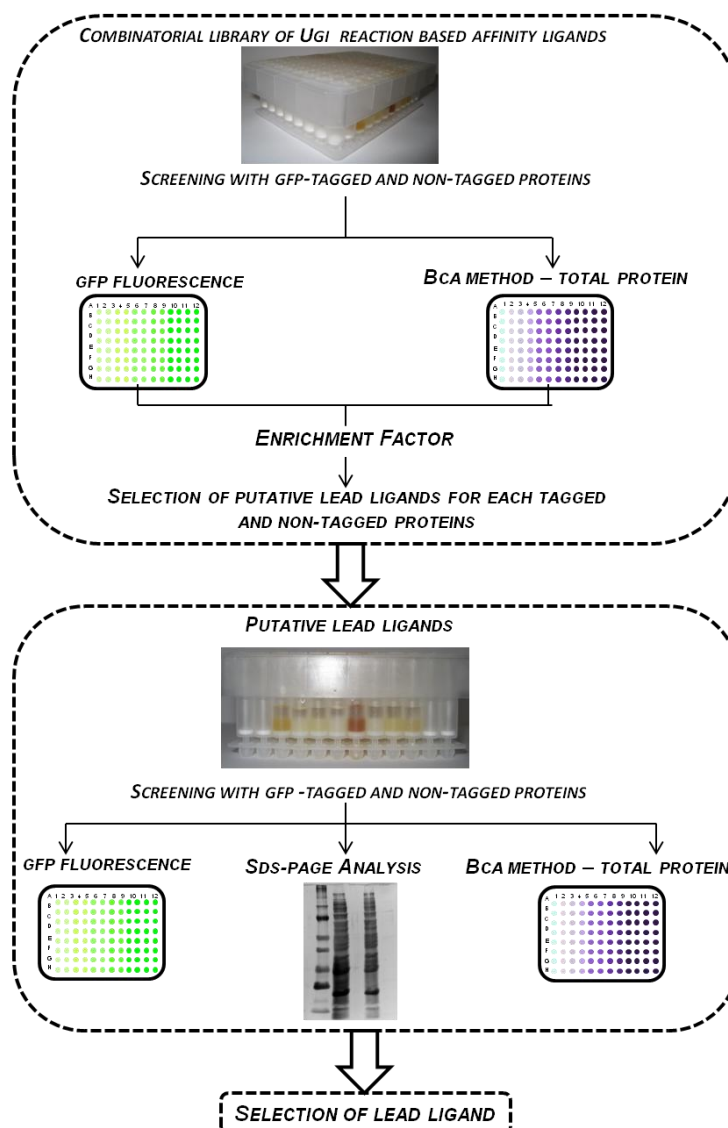


Figure 4.1 Strategy for the selection of lead ligands. The first stage involved the selection of the putative lead ligands through 96-well format screening of a combinatorial library of affinity ligands based on the Ugi reaction and the GFP-tagged proteins. The putative lead ligands selected were those with higher values of enrichment factor, determined by the ratio of GFP fluorescence and the BCA total protein assay. The second stage comprised the refined selection of the lead ligands, by monitoring the % of GFP and total protein bound and SDS-PAGE evaluation.

4.2. SYNTHESIS OF A COMBINATORIAL LIBRARY OF AFFINITY LIGANDS

The synthesis of the libraries of ligands was carried out on a solid phase in a 96-well format. The solid phase presents several advantages over the liquid phase as this latter requires multiple steps of synthesis and isolation procedures. In the solid phase, the purification process can be facilitated by washing off starting materials or other by-products formed by filtration and the amount of starting material required are less when compared with liquid phase. The main disadvantages of the solid phase are related to the instability of the support under the use of certain solvents, as well as the difficulty in characterizing and quantifying the intermediates and

final products synthesized on the solid support due to the poor resolution of the available techniques [233]. In order to overcome the main disadvantages of the solid phase system, the reactions were accomplished by using an excess of starting materials to guarantee the product formation and compatible solvents with the solid support. Moreover, during the solid phase synthesis, quantitative and qualitative tests were performed in order to evaluate the amount of intermediates that were formed until the formation of the final products.

The most conventional matrix used in affinity chromatography for purification purposes is agarose, which is commercially available and well established, as Sepharose CL-6B from GE Healthcare. Agarose is a linear polysaccharide comprised of alternating residues of D-galactose and 3, 6-anhydro-L-galactose and presents a hydrophilic character with a higher amount of primary and secondary alcohols [36]. According to the suppliers, Sepharose CL-6B is a 6% cross-linked agarose with fairly wide particle size distribution of 45 – 165 μm . This material seems to be a good matrix for purification purposes, more specifically in affinity chromatography, as it is a fair base for the coupling of affinity ligands and also has a higher-degree of cross-linking that improves stability under a wide range of conditions such as pH and temperature. The synthesis of the combinatorial library of affinity ligands was based on the Ugi reaction scaffolds (Fig. 4.2, § 2.3.1).

The Ugi reaction is a multicomponent reaction (MCR) that involves an amine (primary or secondary), and oxo compound (aldehyde or ketone), a carboxylic acid and an isocyanide [234-237]. This reaction presents several advantages such as simplicity, time and costs saving, since to create diversity between the combinatorial libraries, only a small amount of starting materials is required [234, 238]. Moreover, the MCRs present selectivity, because from different starting materials, it is possible to achieve only one product. Also MCRs are convergent, because the overall yield of reaction tends to be maximized as the number of step involved decrease, and in this case, as the reactions are one-pot synthesis, this can contribute to maximum convergence [238]. The mechanism of the Ugi reaction is believed to involve as first step the formation of an imine, i.e. a Schiff base. Afterwards, the protonation of the nitrogen atom from Schiff base by the carboxylic acid generates an electrophilic iminium ion [234]. Then, according to what Ugi postulated, the nitrogen atom of the nucleophilic isocyanide attacks the iminium ion and then a Mumm rearrangement occurs to form a stable Ugi product. However, other mechanisms have been suggested [237] and this involves the formation of a hemiaminal, an intermediate compound, originating from the Schiff base and the carboxyl acid group that is further attacked by the isocyanide. The Ugi reaction has already been described in the literature to be a successful strategy for the synthesis of affinity ligands in agarose for immunoglobulin purification [38, 68].

The compounds employed to generate diversity in the first generation libraries were selected to maximise the complementarity with the hexapeptides and are represented in Table 2.1 and 2.2. The libraries was synthesized directly on the solid support by combinatorial methods and further screened micro-scale affinity chromatography. The amines and carboxylic acids employed in the combinatorial library are summarized in Table 2.1 and 2.2.

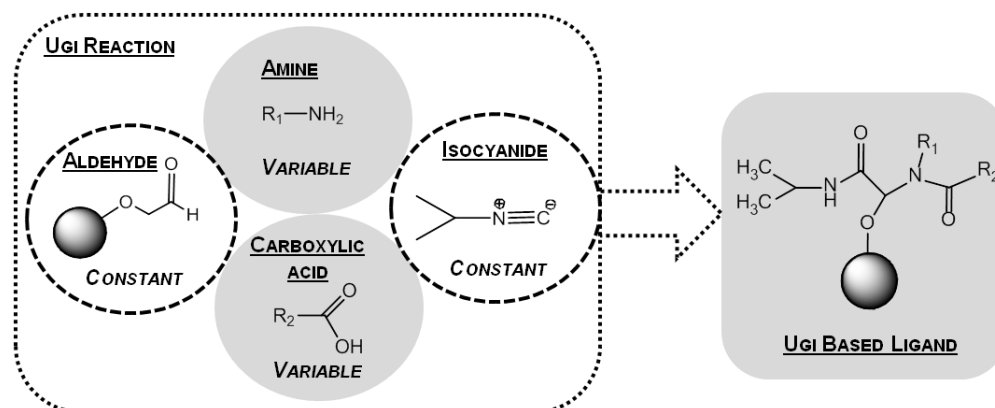


Figure 4.2 Diversity of the compounds used in the synthesis of the Ugi based combinatorial libraries. For the Ugi reaction, the aldehyde and the isocyanide compounds were kept constant and the diversity of the Ugi based ligands was originated from the different amines and carboxylic acids used. The list of amines and carboxylic acids utilised are summarized in Table 2.1 and 2.2

4.3. SCREENING OF THE UGI BASED COMBINATORIAL LIBRARY

Screening of the Ugi based combinatorial library, optimized in terms of cost and variability of compounds structures, was performed through micro-scale affinity chromatography using a 96-well filtration block, where each well contained 250 mg of moist gel functionalized with a different affinity ligand. Crude extracts containing GFP-tagged and non-tagged proteins produced in *E. coli*, rather than pure peptide stock solutions, were utilised in the screening. The results obtained with the crude extracts confirmed simultaneously the affinity and selectivity of the lead ligands. The selectivity was investigated by observing the adsorption of proteins produced by *E. coli*, in the affinity adsorbents in comparison with the adsorption of GFP-tagged or non-tagged.

Molecular recognition is the basis of the interaction between a ligand and its target biomolecule, which is influenced by surface complementarity, non-covalent intermolecular interactions and thermodynamic contributions [9, 239]. Non-covalent intermolecular interactions comprise hydrogen bonds, ionic interactions, van der Waals' forces and hydrophobic interactions. Due to the weak character of these bonds, surface complementarity plays an important role as it allows a higher proximity of the atoms and increases their subsequent interaction [9].

Hydrogen bonds are electrostatic attractions between a donor functional group and an atom or group of atoms able to accept the bond [9, 239]. A proton, covalently bonded to an electronegative donor group carries a partial (δ^+) charge and is thus capable of forming a dipole-dipole interaction with the non-bonded pair of electrons of an electronegative acceptor [240]. Nitrogen and oxygen atoms are those most commonly used in the hydrogen bonding, although sulphur and fluorine are also able to participate [9]. Depending on the functional groups of donor and acceptor, the hydrogen bond can present a different potential [9]. Ionic interactions are weaker and result from the attraction of two oppositely charged atoms (long-range interactions) [9, 239] Van der Waals' interactions are weaker than hydrogen bonding and ionic interactions and result from the attraction between temporary dipoles [9]. Hydrophobic interactions are dependent on surface complementarity to have an acceptable affinity because this type of interaction occurs between non-polar groups such as alkyl chains and aromatic rings [9]

Although these non-covalent intermolecular interactions play an important role in molecular recognition, there are several factors that can influence the strength of the bonds, such as the molecular flexibility, surrounding environment, distance between the acceptor and the donor and the direction of the bond [9, 239].

Crude extracts containing GFP-tagged and non-tagged proteins were incubated with adsorbents in order to promote the binding between the target proteins and the affinity ligands. The temperature was maintained low (4°C) in order to inactivate proteases produced by the host, which are known to be present in the crude samples. After screening, the suspension was centrifuged and all the washes were collected and quantified by GFP fluorescence and the BCA assay. The BCA assay is a sensitive and selective colorimetric assay that allows the quantification of total protein through the measurement of absorbance at 560 nm, due to the product derived from the chelation of two molecules of BCA with one cuprous ion (Cu^+). However, color development in this method is temperature- and time-dependent. Fluorimetric techniques are highly sensitive and require small amounts of ligand and target biomolecule [241]. These two techniques differ in terms of protein detection limits: the fluorescence technique allows quantification of micrograms of protein (0.001-100 $\mu\text{g}/\text{ml}$), whilst in the case of the colourimetric test it is only possible to quantify milligrams of protein (0.2-1 mg/ml). This difference can explain some discrepancies in the determination of the amount of GFP-tagged and non-tagged proteins bound and total protein bound. In order to overcome these issues, in situations where the percentage of total protein bound was zero, the amount of total protein bound was taken as the respective amount of GFP-tagged (or non-tagged) proteins determined by the fluorescence method.

As a first approach, the screening and ligand selection were based on the enrichment factor. The enrichment factor is the ratio between R1 and R2. R1 corresponds to the ratio between

GFP-tagged (or non-tagged) protein bound and total protein bound to the affinity supports; R_2 is the ratio between GFP-tagged (or non-tagged) protein loaded and total protein loaded. The enrichment aimed at assessing the specific binding between the affinity ligand and the target biomolecule. The enrichment factor was determined for each of the affinity ligands and the target biomolecule.

As each GFP production system yields different amounts of total protein and GFP-tagged (or non-tagged) (see Chapter 3), it was necessary to estimate the maximum enrichment factor expected (Table 4.1) in each system. For this, the best case scenario was considered where GFP-tagged (or non-tagged) protein bound totally (i.e. 100% bound), whilst only 5% (w/w) of total protein bound to the affinity support. This last value was based on the minimum non-specific adsorption calculated for plain agarose (Chapter 5 § 5.3).

The selection of the putative lead ligands for each tag system was based on the highest values of enrichment factor. Moreover, promising lead ligands should present only affinity for the respective affinity tag and not for GFP or hosts proteins. Therefore, the screening of the combinatorial library of affinity ligands against GFP can be seen as a negative control and simultaneously will allow the search for the lead ligands displaying selectivity for GFP, which can also be considered as an affinity tag on its own.

Table 4.1 Maximum values of enrichment factor expected for GFP-tagged and non-tagged proteins. The maximum values of enrichment factor were based on the most promising situations possible to obtain in a screening: 100% of GFP protein and 5% protein of total bound.

Crude Extract	% GFP protein in crude extract (w/w)	Maximum Enrichment Factor (-fold)
RWRWRW-GFP	2	15
WFWFWF-GFP (NON DENATURING CONDITIONS)	2	14
NWNWNW-GFP	9	7
RKRKRK-GFP	4	11
NNNNNN-GFP	5	10
GFP	13	6

Therefore, it is important to understand some properties such as isoelectric points (pI) and log P of the peptides that are being used in this work as affinity tags (Fig. 4.2). These parameters were estimated by using the software Marvin Beans, ChemAxon as shown in Table 4.2. However, the parameters were only estimates because there is no information about the conformation of the peptide when fused to the GFP and how these parameters can be influenced upon fusion. The isoelectric point corresponds to the pH where the total net charge of the biomolecule is zero and the log P is the partition coefficient of the compound between octanol and water [242]. The partition coefficient is a hydrophobicity measurement that determines the solvent affinity of specific target biomolecule or bioactive compound [242]. The negative sign of the log P corresponds to a more polar hydrophilic compound whilst a positive relates to a more hydrophobic [242].

The peptides were designed to be used as affinity tags fused to the target protein such that they can recognize the affinity ligand immobilized on the solid support. The C-terminal of the peptides is linked to the linker PPP (three proline residues) and enterokinase recognition site that is present before the target protein to facilitate tag removal after the purification step. Thus,

only the N-terminal and the side chains of the peptides are available to interact with the affinity ligands (Fig. 4.3).

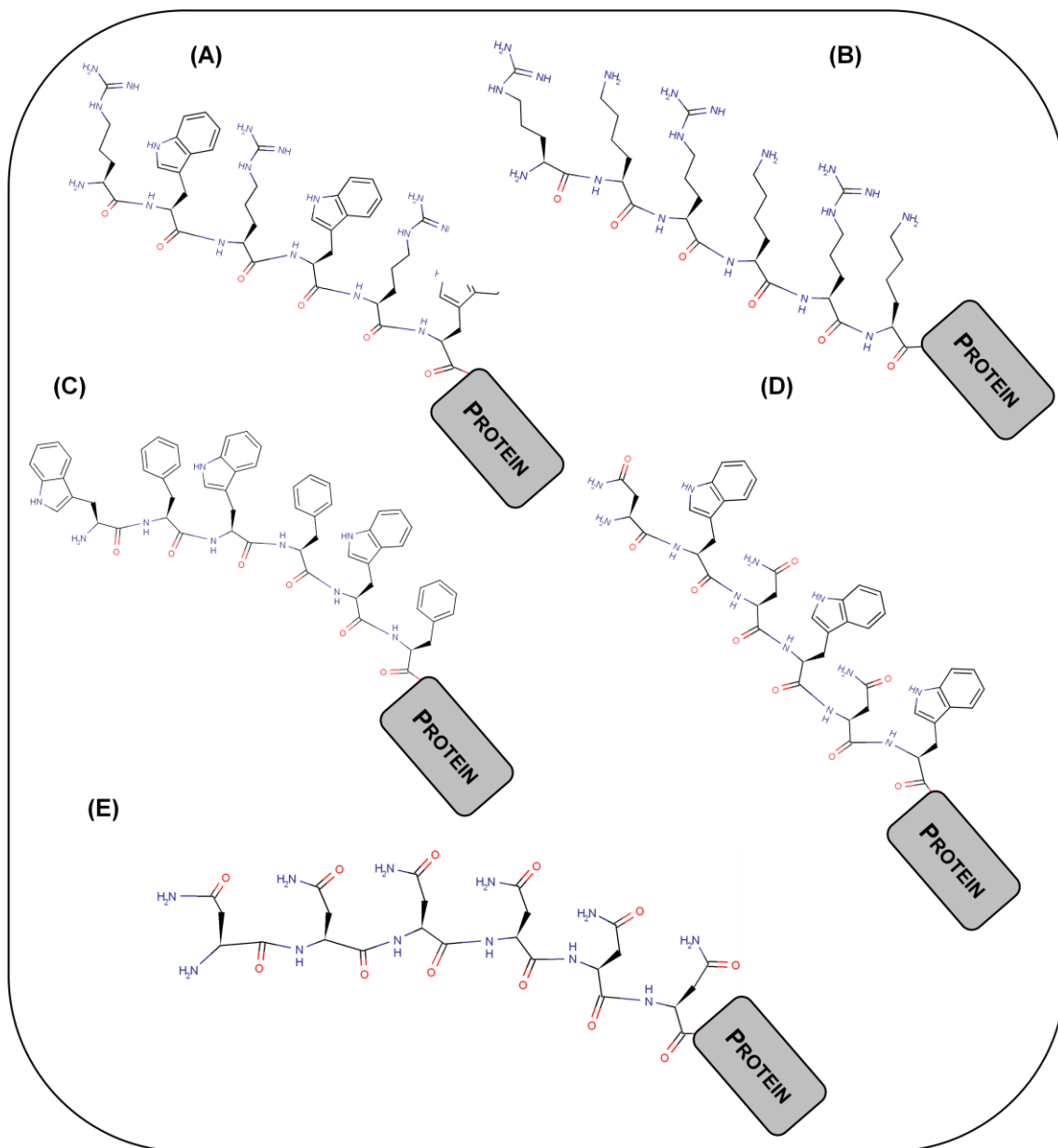


Figure 4.3 Peptide structures used as affinity tags. The peptide structures were designed in Marvin Beans software, ChemAxon. These peptides present a free N-terminal and a C-terminal linked to the recognition site of enterokinase and to GFP protein. The GFP-tagged proteins designed were (A) RWRWRW-GFP, (B) RKRKRKR-GFP, (C) WFWFWFW-GFP, (D) NWNWNW-GFP and (E) NNNNNNN-GFP.

Table 4.2 Properties of the peptides used as affinity tags. The isoelectric point and the logP of the peptides were estimated by using Marvin Beans software, ChemAxon. The estimation of these parameters took into account the fact that only the N-terminal of the peptides is available to interact as the C-terminal is linked to the recognition site of enterokinase and the GFP protein.

Peptide tag	Isoelectric Point without free C-terminal	Isoelectric point with free C-terminal	LogP
RWRWRW	13.23	12.30	-0.72
WFWFWF	10.52	5.18	7.74
NWNWNW	10.05	4.85	-2.94
RKRKRK	13.35	12.35	-6.78
NNNNNN	9.91	4.7	-12.75

The peptide tags RWRWRW and RKRKRK have similar isoelectric point with and without the COOH group available on C-terminal (Table 4.2). The pI estimated for the peptide tags WFWFWF, NWNWNW and NNNNNN by using the C-terminal free is different from those obtained without the free COOH group on C-terminal. For peptides RKRKRK and RWRWRW, the existence of a protonatable COOH group does not significantly alter the pI due to the highly positively charged side chains. For logP no significant changes were observed when estimating the parameter for the peptides with and without free C-terminal because the COOH group on C-terminal might not be significant to alter the hydrophobicity or hydrophilicity of the peptides as the contributions of the side chains of the aminoacids that compose the affinity tags. As shown in Table 5.2, the most negative value, i.e., most hydrophilic character, corresponds to the peptide NNNNNN whilst the most positive value is peptide WFWFWF i.e. the most hydrophobic character. Peptide RKRKRK also presents a negative value, and hence has hydrophilic character, whilst RWRWRW and NWNWNW peptides exhibits a log P value closer to 0, suggesting neutral characteristics.

4.3.1. Screening with RWRWRW tagged GFP crude extract

The affinity tag RWRWRW display a positive charge at pH 7.4 according to Table 4.2. Moreover, this peptide presents a dual hydrophobic and hydrophilic behaviour and due to this, the interactions involved with this peptide might be electrostatic and hydrophobic interactions. The percentage of GFP tagged protein in the crude extract was low (~2% (w/w)) and hence the estimated maximum value for the enrichment factor was 15. The overall results of the library screening are presented in Fig 4.4, where at least 7 affinity ligands were distinguishable: A1C6, A2C4, A2C8, A3C6, A3C7, A3C8, A6C7 and A6C8.

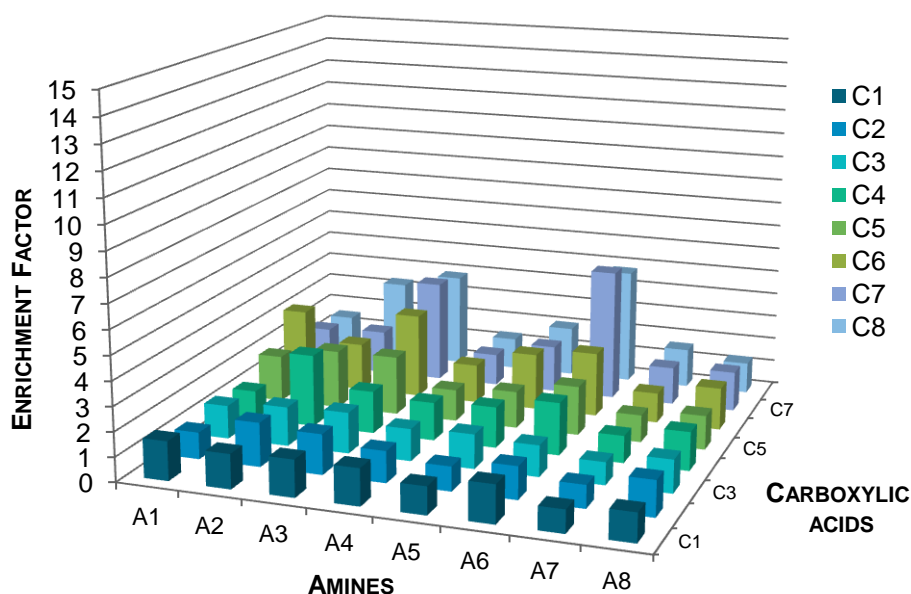


Figure 4.4: Screening of combinatorial library for RWRWRW tagged GFP. Ugi functionalized-Sepharose beads were synthesized in each well of the 96-well filtration blocks (250 mg of resin, c.v.: 250 μ l) and then prepared for chromatographic analysis through regeneration (0.1M NaOH, 30% (v/v) isopropanol alternated with H₂O, 3 c.v. each in a total of 6 c.v.), equilibration (PBS: 10 mM sodium phosphate, 150 mM NaCl, pH 7.4, 12 c.v.) prior to loading (1 c.v. of crude extract containing GFP tagged protein obtained from expression of GFP tagged protein in *E.coli* cells) and incubation (1h at 4°C with manual shaking). The flow-through (1 c.v.) and wash fractions (PBS 12 c.v.) were collected in 96 well black microplates. The fractions were quantified by spectrofluorimetry and BCA assays which determine the amount of GFP and total protein unbound, respectively. Ligand density was 20 μ mol / g moist weight gel. Enrichment factor = R1/R2, where R1= (GFP tagged (or non-tagged) protein bound/ total protein bound) and R2 = (GFP tagged (or non-tagged) protein loaded / total protein loaded).

The highest value of enrichment reached by the affinity ligands A6C7 and A6C5 was ~5. Both ligands possess aromatic rings, which can establish hydrophobic interactions with the tryptophan indole ring from the RWRWRW peptide (Fig. 4.5). It is conceivable that amine A6 is responsible for engaging in hydrogen bonding with the carbonyl group of the peptide backbone or with nitrogen atoms of the arginine side chains.

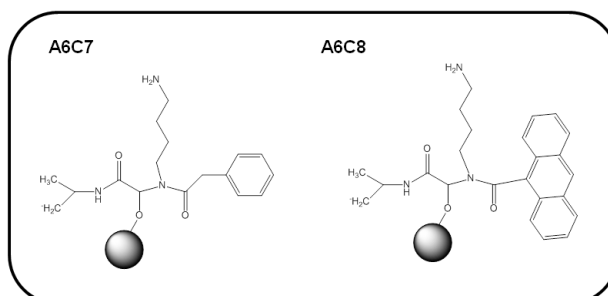


Figure 4.5 Putative lead ligands structures with an enrichment factor ~ 5 for the RWRWRW tagged GFP. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

The other selected ligands A1C6, A2C4, A2C8, A3C6, A3C7 and A3C8 (Fig. 4.6) present an enrichment factor between 3 and 4. Common features of these ligands include amines A2 and A3 which infer the presence of three and two carboxylic acid groups, respectively, combined with aromatic rings (Fig. 4.6). Ligand A1C6 comprises a single carboxylic group and an aromatic ring with a hydroxyl group. It is envisaged that the carboxylic groups will interact with arginine residues and the aromatic moiety will interact with the indole ring of tryptophan.

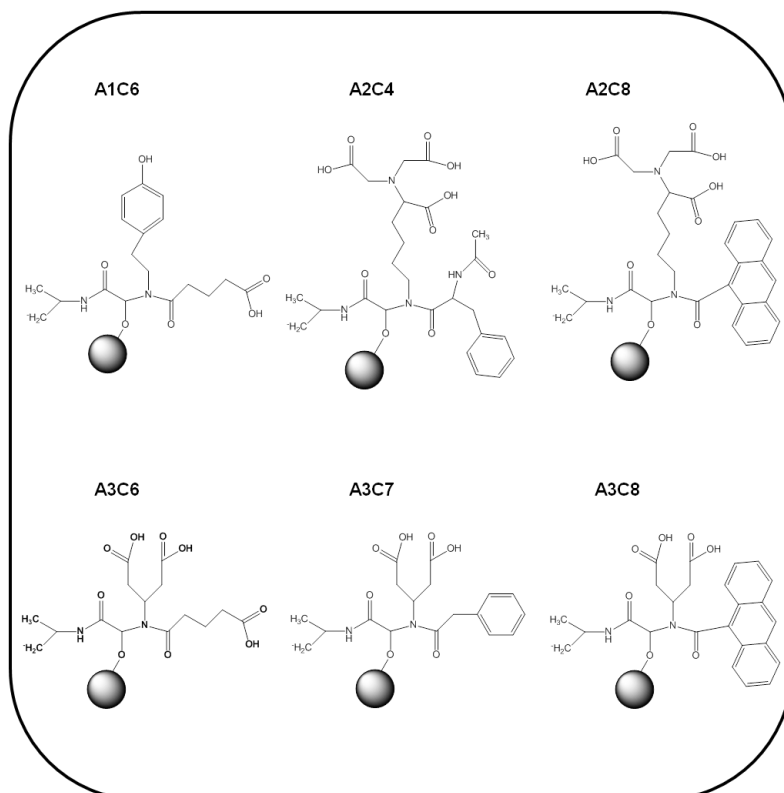


Figure 4.6 Putative lead ligands structures with an enrichment factor ~3- 4 for the RWRWRW tagged GFP. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

Ligands presenting enrichment factors between 2 and 3 were also considered. These included ligands A2C5, A3C5, A5C6 and A6C6 (Fig. 4.7).

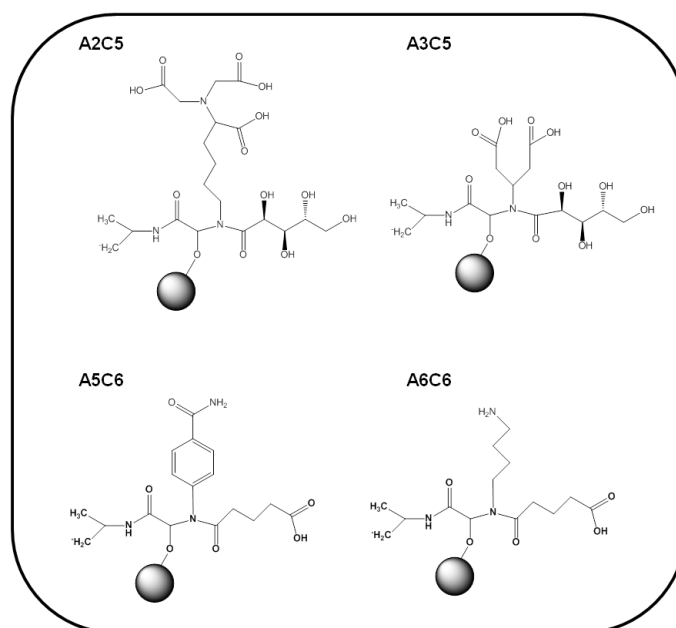


Figure 4.7 Putative lead ligands structures with an enrichment factor ~2- 3 for the RWRWRW tagged GFP. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

The values of isoelectric point (pI) and logP were estimated for each of the putative lead ligands and then correlated with the respective values of enrichment factor (Fig.4.8). Fig. 4.8 (A) suggests that ligands presenting a higher enrichment value (A6C7 and A6C8), also present a similar isoelectric point (12.51 and 12.53, respectively) to RWRWRW, and therefore are also positively charged at pH 7.4. In these cases, electrostatic interactions might not be in the basis of the interaction between the ligands and the affinity tag. For the other selected ligands with enrichment factor below 5, pI values are lower than 7.4, indicating a negative charge of these ligands at the screening conditions employed. It was expected that these ligands interact with the affinity tag *via* electrostatic interactions and therefore, could have higher enrichment factors values. These facts might suggest that the electrostatic interactions are not dominant in the molecular recognition between the putative ligands and the affinity tag. On the other hand, Fig. 4.8 (B) suggested a tendency for the ligands with higher values of enrichment factor, to present more positive values of Log P (a more hydrophobic character).

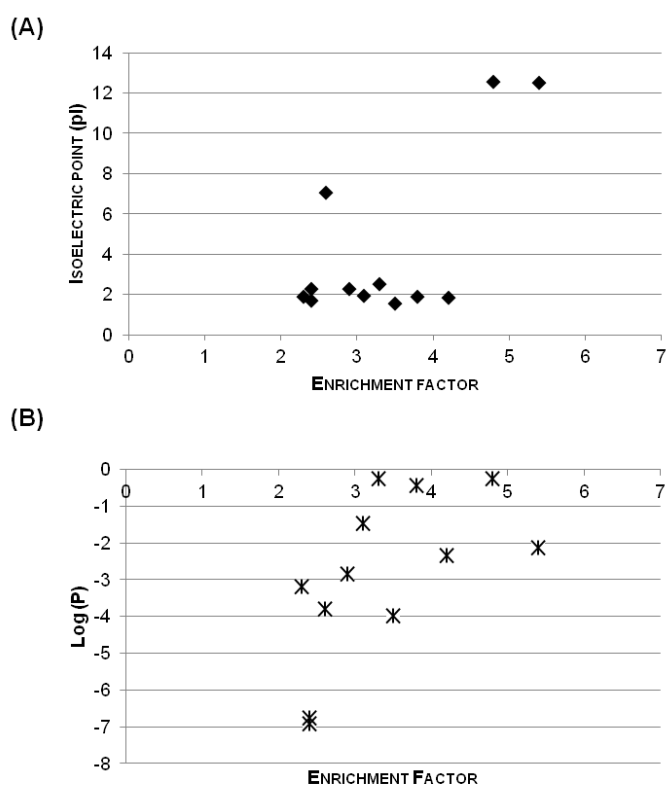


Figure 4.8 Correlation between (A) the isoelectric point and (B) Log P values and enrichment factor value obtained for the screening with the RWRWRW tagged GFP. The values obtained for the isoelectric point and LogP were estimated by the software Marvin Beans, ChemAxon.

4.3.2. Screening with WFWFWF tagged GFP crude extract

The WFWFWF peptide has a strong hydrophobic character and the production of this peptide fused to GFP led to the formation of inclusion bodies. Different culture conditions were selected in order to produce this tagged protein in a soluble form. However, as the formation of inclusion bodies usually allow higher amounts of protein to be expressed, purification under denaturing conditions and subsequent refolding on column has proven as a viable alternative. Thus, in this particular case, the screening was performed both under non-denaturing conditions (i.e. with the soluble GFP-tagged protein) and also in denaturing conditions (i.e. in the presence of 8M urea).

4.3.2.1 Screening under Non denaturing conditions

The WFWFWF peptide possess overall hydrophobic properties but may also engage in hydrogen bonds with the carbonyl group and nitrogen group of the peptide backbone and the indole ring of the tryptophan. As estimated in Table 4.2, the peptide has a low positive charge at pH 7.4.

The results obtained from screening WFWFWF tagged GFP under non-denaturing conditions are presented in Fig. 4.9; the maximum value of enrichment calculated was 14. The enrichment values obtained were promising, as the best affinity ligand presented in Fig 4.9 can enrich around ~10 times, i.e. ~70% of the maximum value of enrichment factor. Affinity ligand A6C1 appears to be very effective as an adsorbent for the WFWFWF tagged GFP. The ligand has a flexible relatively polar structure. Another promising ligand is A1C5, comprising arabic acid as the carboxylic compound and a tyramine as the amine compound (Fig. 4.10), with an enrichment factor around 5. Other putative lead ligands were selected with lower enrichment factors (~ 4-6). These selected promising ligands can be divided in two groups: Those containing amine, A2 (A2C4, A2C5, A2C6, A2C7 and A2C8) (Fig. 4.11), and those containing amine A3 (A3C1, A3C6, A3C7 and A3C8) (Fig. 4.12). Some of these ligands were also identified as putative lead ligands for RWRWRW, suggesting that they might not be selective for the dipeptide repetition on affinity tags, although both tags contain tryptophan.

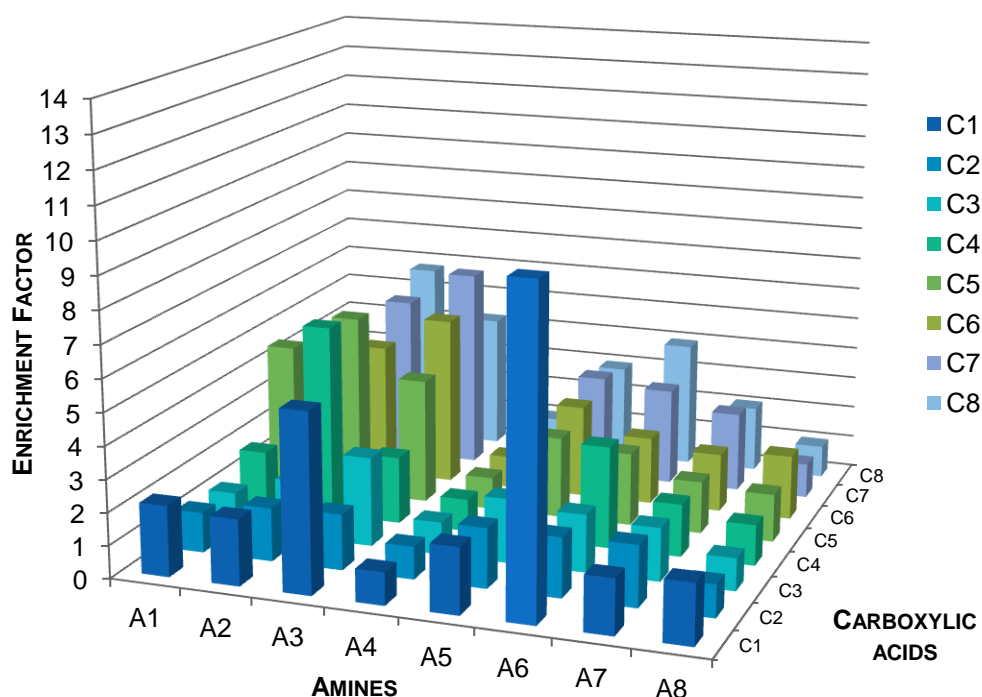


Figure 4.9 Screening of combinatorial library for WFWFWF tagged GFP under non denaturing conditions. Ugi functionalized-Sepharose beads were synthesized in each well of the 96-well filtration blocks (250 mg of resin, c.v.: 250 μ l) and then prepared for chromatographic analysis through regeneration (0.1M NaOH, 30% (v/v) isopropanol alternated with H₂O, 3 c.v. each in a total of 6 c.v.), equilibration (PBS: 10 mM sodium phosphate, 150 mM NaCl, pH 7.4, 12 c.v.) prior to loading(1 c.v. of crude extract containing GFP tagged protein obtained from expression of GFP tagged protein in *E.coli* cells) and incubation (1h at 4°C with manual shaking. The flow-through (1 c.v.) and wash fractions (PBS 12 c.v.) were collected in 96 well black microplates. The fractions were quantified by sprctrofluorimetry and BCA assays which determine the amount of GFP and total protein unbound, respectively. Ligand density was 20 μ mol / g moist weight gel. Enrichment factor = R1/R2, where R1= (GFP tagged (or non-tagged) protein bound/ total protein bound) and R2 = (GFP tagged (or non-tagged) protein loaded / total protein loaded).

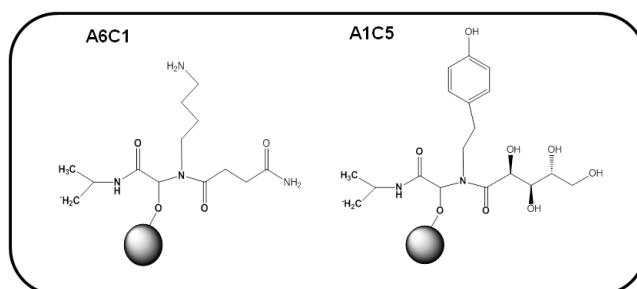


Figure 4.10 Structures of the putative lead ligands **A6C1** and **A1C5** with enrichment factors ~10 and 5, respectively for the **WFVWF** tagged **GFP** under non-denaturing conditions. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

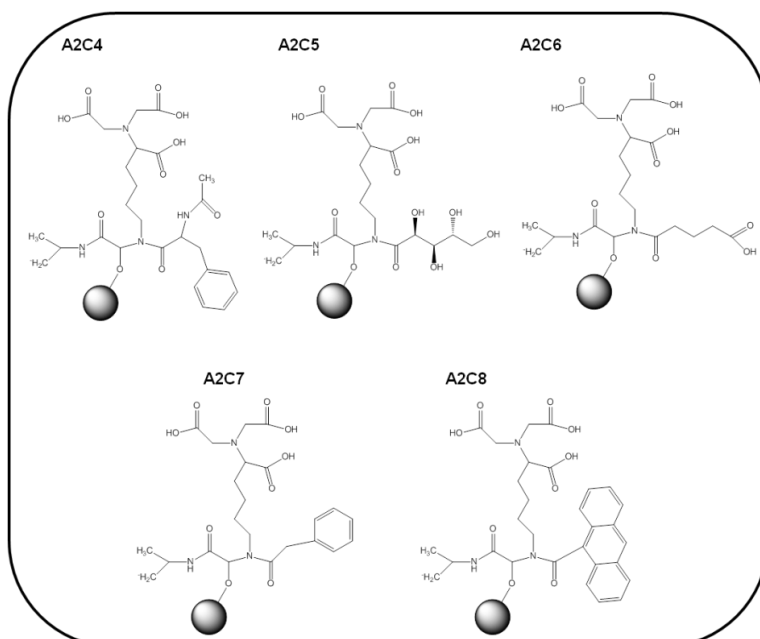


Figure 4.11 Structures of the putative lead ligands with enrichment factors ~2-4 that contain the same amine **A2** for the **WFVWF** tagged **GFP** under non-denaturing conditions. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads.

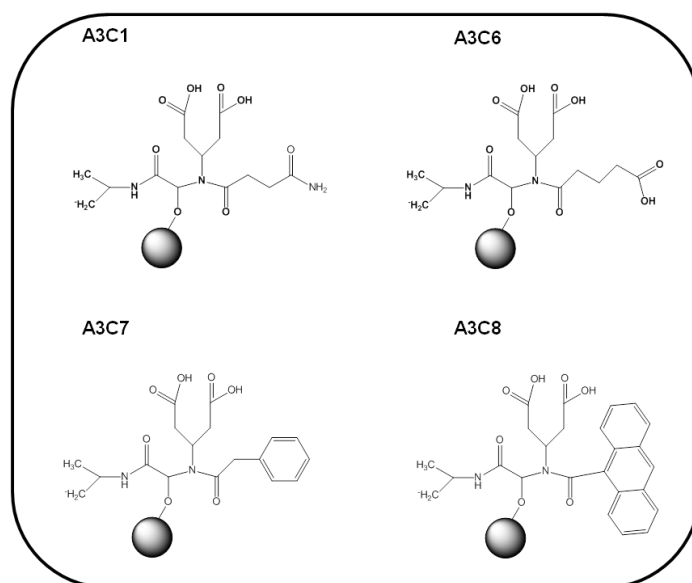


Figure 4.12 Structures of the putative lead ligands with enrichment factors ~2-4 that contain the same amine A3 for the WFWFWF tagged GFP under non-denaturing conditions. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

The properties isoelectric point and LogP were also estimated for the promising lead ligands and further correlated with the respective enrichment factor as shown in Fig.4.13

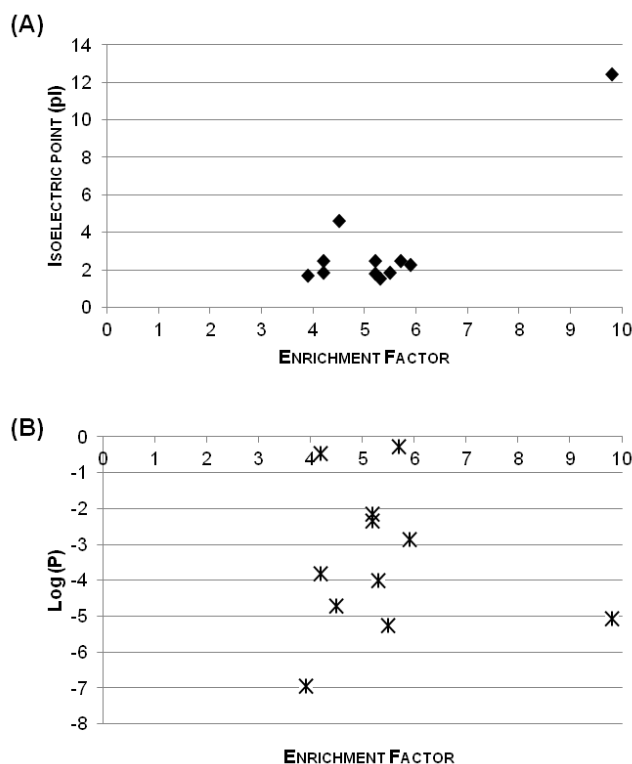


Figure 4.13 Correlation between (A) the isoelectric point and (B) Log P values and enrichment factor value obtained for the screening with the WFWFWF tagged GFP. The values obtained for the isoelectric point and LogP were estimated by the software Marvin Beans, ChemAxon.

By observing Fig. 4.13 (A) and (B), the promising lead ligand A6C1 stands out from the other ligands in terms of enrichment value. At the screening conditions used, this ligand is positively charged (pI equals to 12.43) and presents a hydrophilic character (Log P equals to -5.07); the peptide WFWFWF presents a positive charge and displays a highly hydrophobic character. Therefore, hydrophobic interactions are expected to prevail in the molecular recognition between this ligand and the tag.

The remaining ligands present a similar value of pI (Fig. 4.13(A)) and display a negative charge at pH 7.4. The Log P values are less negative, for ligands A2C8, A2C4 both with enrichment factor ~6 and ligands A3C7, A2C4 with enrichment factor ~ 5. The common feature of these ligands is the presence of at least one aromatic ring on their structures. Therefore both electrostatic and the hydrophobic interactions are likely to contribute for the molecular recognition between the tag and the ligand.

4.3.2.2 Screening under Denaturing Conditions

The screening under denaturing conditions was performed in the presence of 8M urea, which was required to solubilise inclusion bodies [190]. The crude extract was applied to the 96-well filtration block for further purification and refolding on-column. However, in this case, it was only possible to determine the amount of GFP bound through the fluorimetric assay, since the colourimetric BCA assay, can only be used at urea concentrations below 3M, according with supplier instructions. Therefore, it was not possible to determine the enrichment factor. The evaluation of the lead ligands was assessed by determining the amount of bound tagged GFP per gram of resin, as shown in Fig. 4.14. The enrichment and selectivity of the putative lead ligands could be evaluated subsequently by SDS-PAGE analysis.

The WFWFWF tagged GFP protein has a highly hydrophobic character and was expressed as inclusion bodies, which are aggregates of misfolded protein. The GFP chromophore is formed due to a covalent modification of three adjacent amino acids in the GFP sequence [243] and is located in the middle of an α -helix structure in the centre of the β -barrel of the GFP structure [243]. The tertiary structure allows protection of the GFP chromophore from the aqueous environment. However, after solubilization of the GFP-tagged inclusion bodies with high concentrations of chaotropic agents such as urea, GFP loses the secondary structure exposing the hydrophobic surface [222], and a quenching of the GFP fluorescence [243]. The chromophore seems to be resistant, and only when using high temperatures in association with high concentrations of chaotropic agents, it is possible to observe denaturation of the GFP and a complete loss of fluorescence [243].

The WFWFWF tagged GFP still retained some fluorescence in the solubilization buffer containing 8M urea, despite the high concentration of protein corroborated by the SDS-PAGE analysis after expression (§ 3.5.3), where an intense band of GFP-tagged can be seen.

The screening protocol was performed manually and stepwise and consisted of the following steps: resin regeneration, biomolecule loading and refolding during the washing step. Proper folding was not likely to occur since the washes did not contain a slow reduction in the urea concentration, and tagged GFP precipitated. The adsorbents were washed in a total of 12 C.V., where 6 C.V. contained binding buffer with 4M of urea and the remaining washes only binding buffer. However, the principal aim at this stage was to understand which affinity ligands could bind to tagged GFP under these denaturing conditions.

Affinity ligands displaying higher amounts of GFP tagged bound (Fig 4.9) where those containing amines A4 and A8 (Fig. 4.15), in particular, ligands A4C2, A4C5 and A4C8. The hydrophobic character of these ligands increases from A4C5 > A4C2 > A4C8 (Fig. 4.15). Considering the similarities between A4C2 and A4C8 and the fact that the latter displayed a higher binding, ligand A4C8 was considered to proceed with further studies. Ligand A4C5 was interesting in the sense that possess a carboxylic acid group with a very hydrophilic character. On contrary, affinity ligand A8C8 also displaying higher binding per gram of resin presents an hydrophobic character with an indole ring.

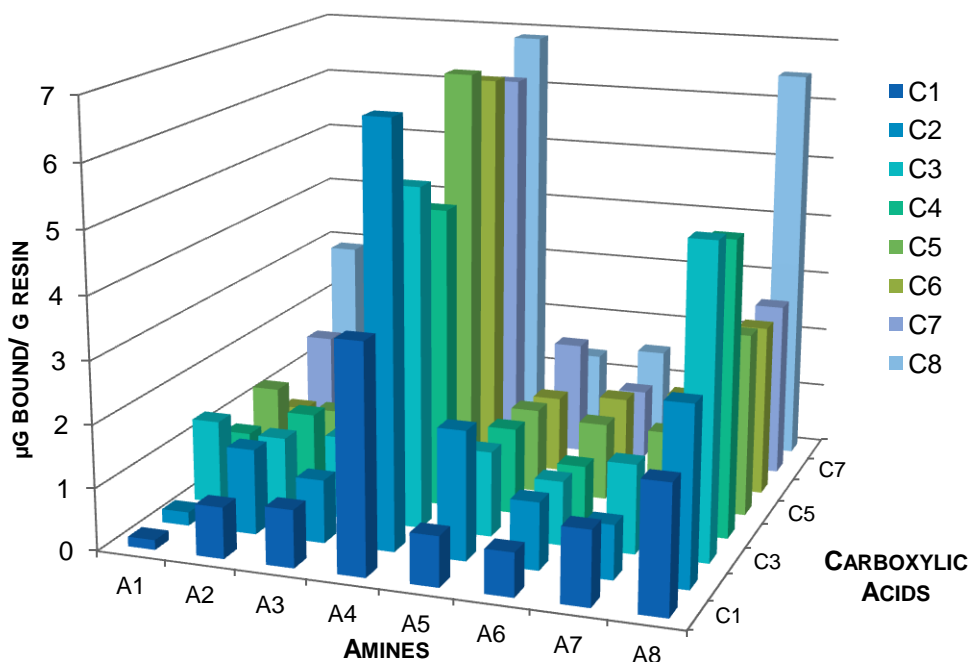


Figure 4.14 Screening of combinatorial library for WFWWF tagged GFP under denaturing conditions. Ugi functionalized-Sepharose beads were synthesized in each well of the 96-well filtration blocks (250 mg of resin, c.v. :250 µl) and then prepared for chromatographic analysis through regeneration (0.1M NaOH, 30% (v/v) isopropanol alternated with H₂O, 3 c.v. each in a total of 6 c.v.), equilibration (PBS: 10 mM sodium phosphate, 150 mM NaCl, pH 7.4 12 c.v.) prior to loading(1 c.v. of solubilised inclusion bodies in 8M urea) and incubation (1h at 4°C). The flow-through (1 c.v.) and washes fractions (PBS 4M urea 6 c.v. and PBS 0M urea 6 c.v.) were collected in 96 well black microplates. The fractions were quantified by spectrofluorimetry and BCA assays which determine the amount of GFP and total protein unbound, respectively. Ligand density was 20 µmol / g moist weight gel). (Protein bound = (Protein loaded)-(sum of the amount protein in flow-through and washes).

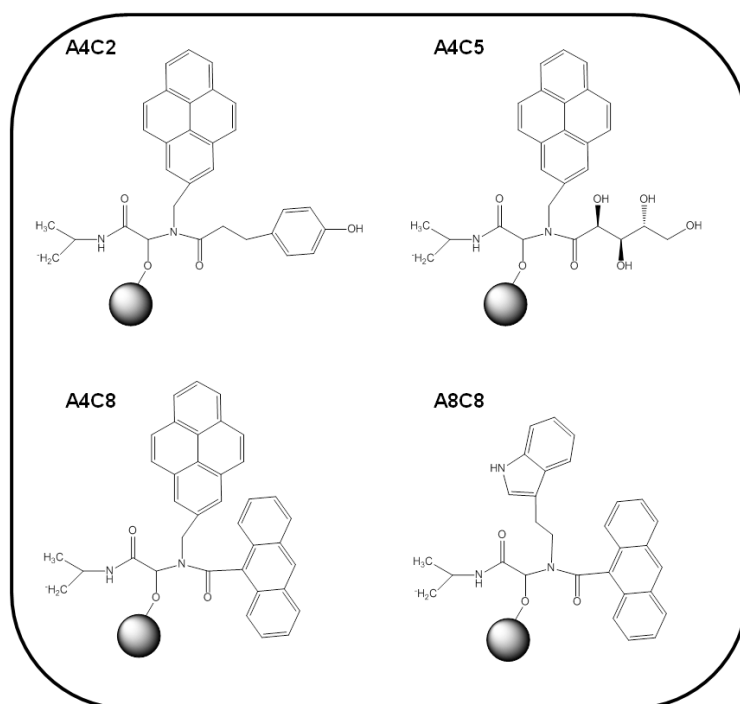


Figure 4.15 Structures of the selected putative lead ligands for the WFWWF tagged GFP under denaturing conditions. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

4.3.3. Screening with NWNWNW tagged GFP crude extract

The maximum enrichment factor possible for the NWNWNW tagged GFP was estimated as 7. Through analysis of the data shown in Fig. 4.16, there were a few affinity ligands with promising properties, in particular, ligand A7C8 which presents an enrichment factor value around 6, i.e. ~ 85% of the maximum. Peptide NWNWNW presents a slightly positive charge at pH 7.4 but has mainly an apolar character.

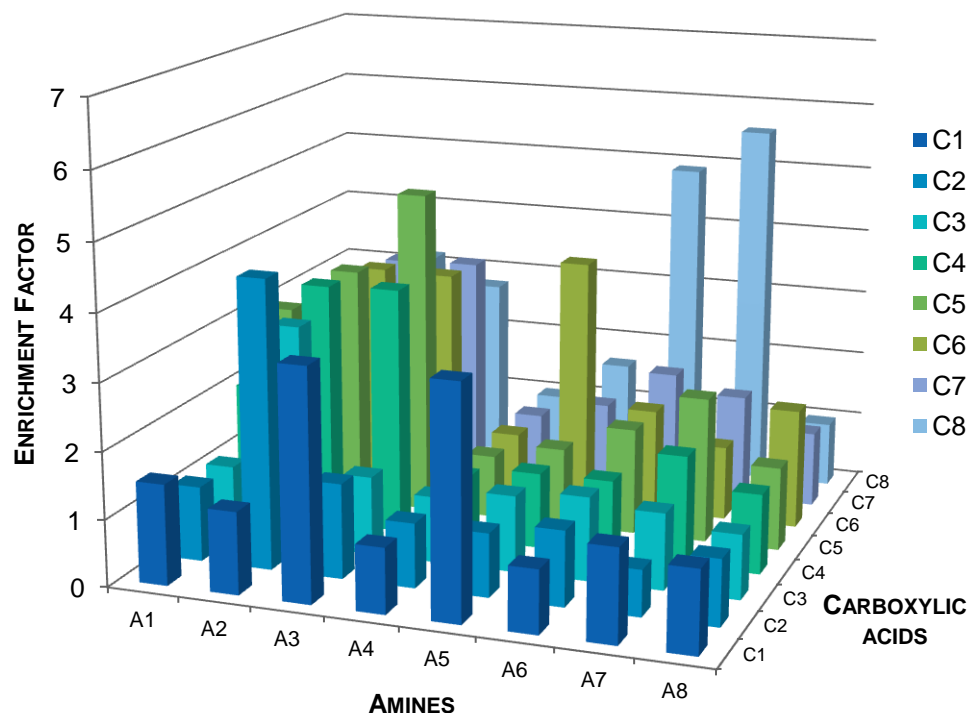


Figure 4.16 Screening of combinatorial library for NWNWNW tagged GFP. Ugi functionalized-Sepharose beads were synthesized in each well of the 96-well filtration blocks (250 mg of resin, c.v.: 250 μ l) and then prepared for chromatographic analysis through regeneration (0.1M NaOH, 30% (v/v) isopropanol alternated with H₂O, 3 c.v. each in a total of 6 c.v.), equilibration (PBS: 10 mM sodium phosphate, 150 mM NaCl, pH 7.4, 12 c.v.) prior to loading (1 c.v. of crude extract containing GFP tagged protein obtained from expression of GFP tagged protein in *E.coli* cells) and incubation (1h at 4°C). The flow-through (1 c.v.) and wash fractions (PBS 12 c.v.) were collected in 96 well black microplates. The fractions were quantified by spectrofluorimetry and BCA assays which determine the amount of GFP and total protein unbound, respectively. Ligand density was 20 μ mol / g moist weight gel. Enrichment factor = R1/R2, where R1= (GFP tagged (or non-tagged) protein bound/ total protein bound) and R2 = (GFP tagged (or non-tagged) protein loaded / total protein loaded).

Ligands A6C8 and A3C5 (Fig.4.17) also displayed enrichment factors around 5. However, these ligands were already selected as putative lead ligands for the RWRWRW tagged GFP, suggesting that they are relatively unselective for the specific peptides. Ligand A2C5 (enrichment factor ~4), has the same structure as A3C5 but comprises one more carboxylic acid group (Figs. 4.17 and 4.18). Other relevant affinity ligands were A5C1 and A5C6; these ligands have enrichment factors of ~ 3 and possess similar structures containing a benzamide group, with the difference that A5C1 has an extra amide and A5C6 has a COOH group (Fig. 4.18). In

this way, both ligands present the same profile of hydrogen bonding with N-H / O-H donors and O-H/ C=O acceptors [240].

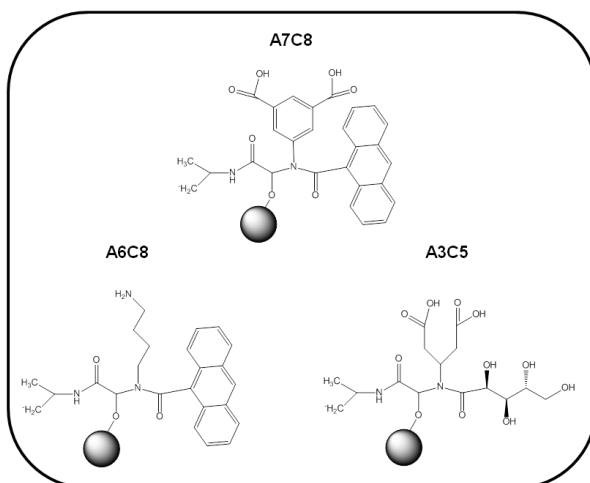


Figure 4.17 Putative lead ligands for NWNWNW tagged GFP, where ligand **A7C8** presents an enrichment factor ~6 and ligands **A6C8** and **A3C5** ~5. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

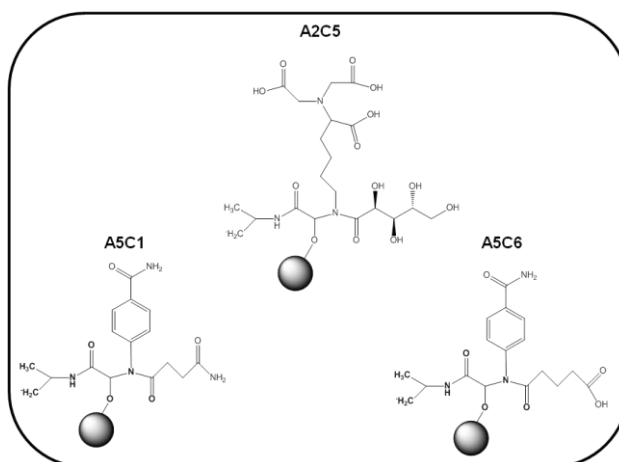


Figure 4.18 Promising lead ligands for NWNWNW tagged GFP, where ligand **A2C5** presents an enrichment factor ~4 and ligands **A5C1** and **A5C6** ~3. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

The other promising lead ligands fall into two categories: Those containing amines A2 and A3. These amines may be strongly engaged in hydrogen bonding with the affinity peptide tags, as already discussed for the RWRWRW peptide (§ 4.2.1).

Ligands with amine A2 are A2C2, A2C3, A2C4 and A2C6 (Fig. 4.19). The first three ligands do not differ very much because all the carboxylic acid compounds contain an aromatic ring or an

indole ring. Thus, not surprisingly, the enrichment factors obtained for these affinity ligands are similar, but the one obtained for A2C2 is slightly higher as it possesses an additional hydroxyl group. Affinity ligand A2C6 also has the same groups but has a more flexible structure. The putative lead ligands with the amine A3 are A3C1 and A3C4 (Fig. 4.20). A3C4 presents a slightly higher enrichment factor because it contains both hydrophobic (aromatic ring) and hydrophilic (COOH groups) functionalities.

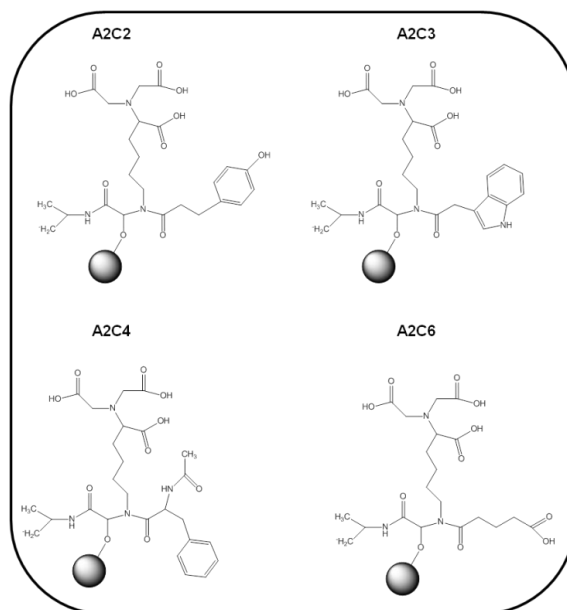


Figure 4.19 Lead ligands for NWNWNW tagged GFP with a common A2 amine and enrichment factor ~2. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

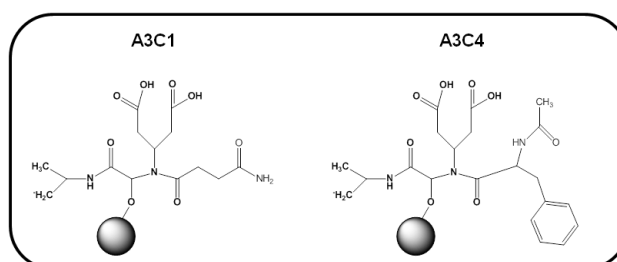


Figure 4.20 Lead ligands for NWNWNW tagged GFP with a common A3 amine and enrichment factor ~2. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

The properties of the selected lead ligands such as isoelectric point and LogP were also estimated (Fig. 4.21). Most ligands present a similar isoelectric point (~2), and the same negative charge at pH 7.4 employed during the screening.

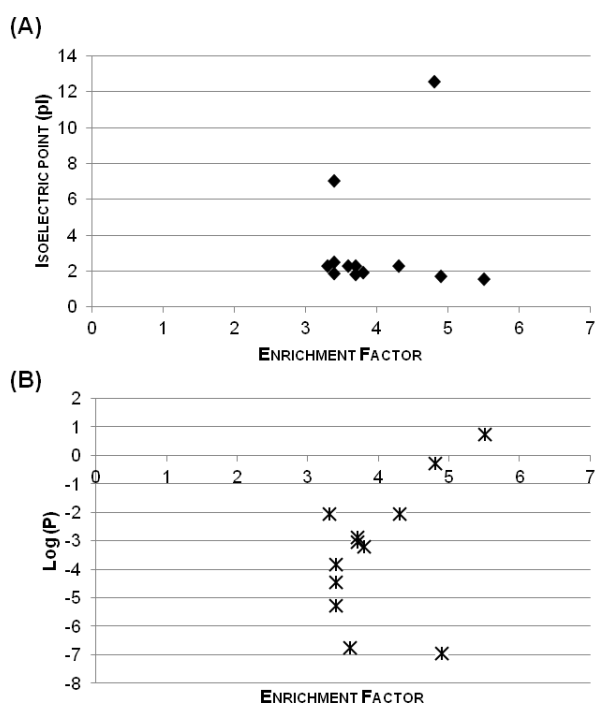


Figure 4.21 - Correlation between (A) the isoelectric point and (B) Log P values and enrichment factor value obtained for the screening with the NWNWNW tagged GFP. The values obtained for the isoelectric point and LogP were estimated by the software Marvin Beans, ChemAxon

The hydrophobic contribution for the interaction between the tag and ligands seem more pronounced. Ligands with more hydrophobic structures (more positive LogP values) also present higher values of enrichment factors. Therefore this can reveals the importance of the hydrophobic interactions establishment between the pair. Overall, the lead ligand with the highest enrichment factor value (~6) was A7C8 with a likely contribution of both electrostatic and hydrophobic interactions between the pair.

4.3.4. Screening with RKRKRK tagged GFP crude extract

The results obtained from the screening of RKRKRK tagged GFP and the combinatorial library of affinity ligands are shown in Fig.4.22. The maximum enrichment value calculated was 11. However, the highest value found was 2, much lower than the theoretical expected; and was observed for ligand displaying amine A7, particularly A7C7 and A7C8 (Figs 4.23 and 4.24). Affinity ligand A7C8 was also selected as a putative lead ligand for the NWNWNW peptide indicating lack of selectivity. Ligands A2C4 and A5C4 (Fig, 4.25) were also selected since they showed values of enrichment slightly below A7 ligands.

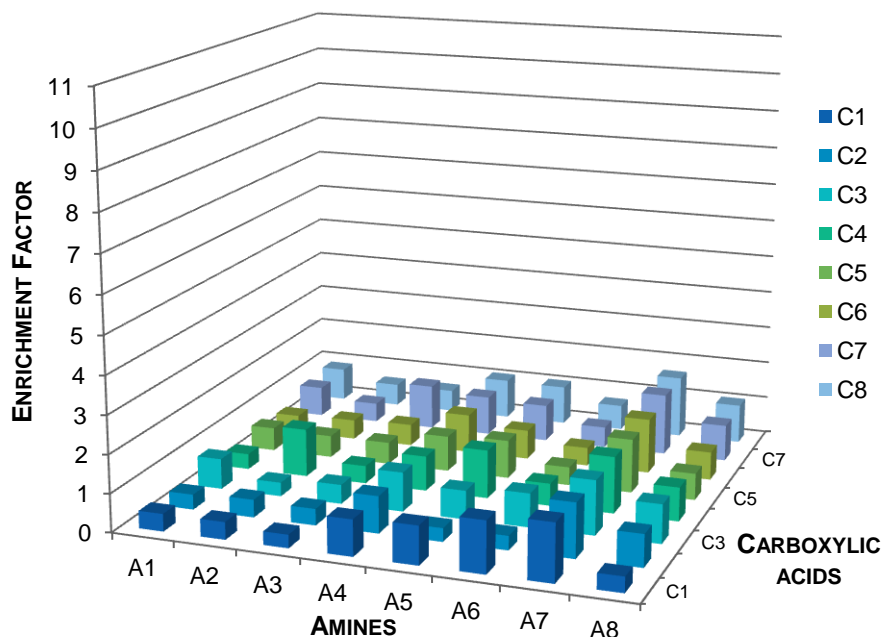


Figure 4.22 Screening of combinatorial library for RKRKRK tagged GFP. Ugi functionalized-Sepharose beads were synthesized in each well of the 96-well filtration blocks (250 mg of resin, c.v.: 250 μ l) and then prepared for chromatographic analysis through regeneration (0.1M NaOH, 30% (v/v) isopropanol alternated with H₂O, 3 c.v. each in a total of 6 c.v.), equilibration (PBS: 10 mM sodium phosphate, 150 mM NaCl, pH 7.4, 12 c.v.) prior to loading (1 c.v. of crude extract containing GFP tagged protein obtained from expression of GFP tagged protein in *E.coli* cells) and incubation (1h at 4°C). The flow-through (1 c.v.) and wash fractions (PBS 12 c.v.) were collected in 96 well black microplates. The fractions were quantified by spectrofluorimetry and BCA assays which determine the amount of GFP and total protein unbound, respectively. Ligand density was 20 μ mol / g moist weight gel. Enrichment factor = R1/R2, where R1= (GFP tagged (or non-tagged) protein bound/ total protein bound) and R2 = (GFP tagged (or non-tagged) protein loaded / total protein loaded).

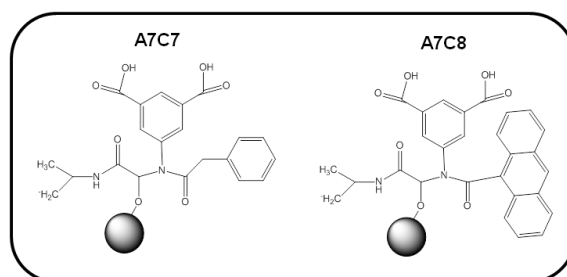


Figure 4.23 Lead ligands A7C7 and A7C8 for RKRKRK tagged GFP. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

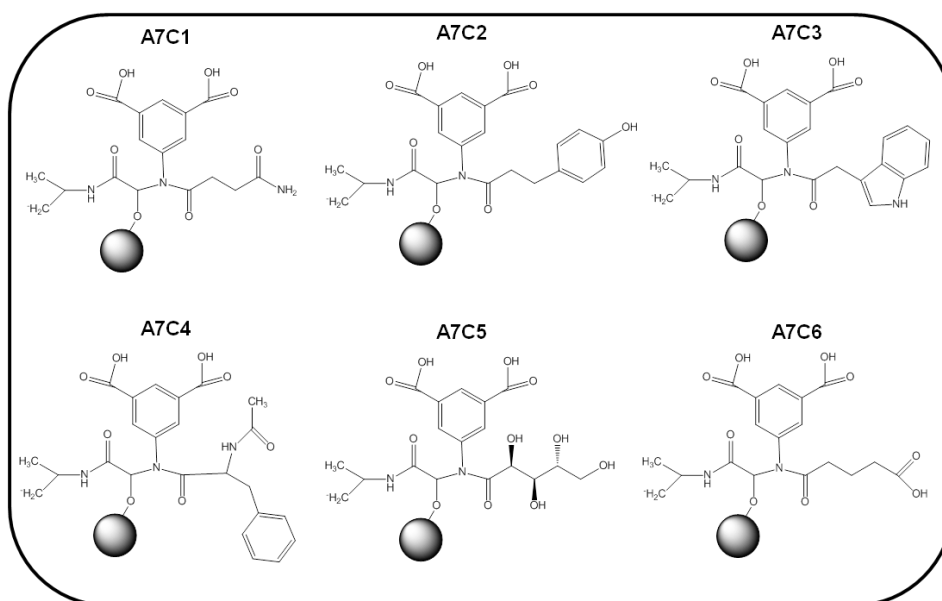


Figure 4.24 Lead ligands with A7 amine, with enrichment factor ~2. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

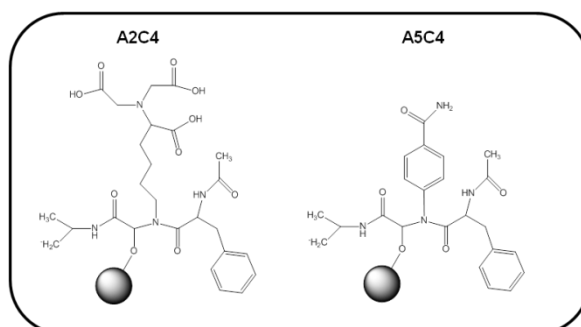


Figure 4.25 Ligands A2C4 and A5C4, selected as putative lead ligands for the RKRKRK tagged GFP. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

When considering the properties of the ligands, they all present a negative charge at the screening conditions employed as the pI is below 7.4 (Fig. 4.26(A)). The peptide RKRKRK has a positive charge at pH 7.4. Therefore, the electrostatic interaction could be in the basis of the molecular recognition between the affinity pairs. According to the Fig 4.26 (B), the hydrophobicity of the ligands varies between them, however all the ligands display at least one aromatic ring. These aromatic rings should contribute for enhance the rigidity of the ligands and therefore contribute for a better interaction with the tag, as this tag presents a flexible character. Due to ligands structures, also hydrogen bonding can reinforce the interaction between the affinity pairs.

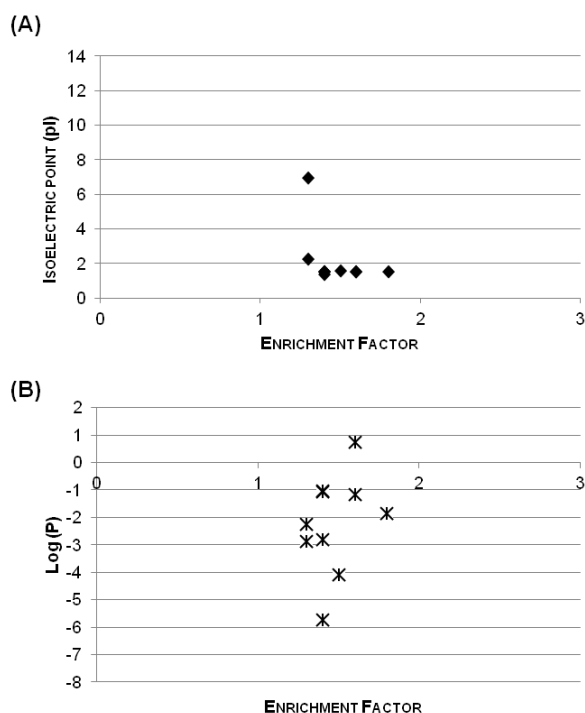


Figure 4.26 Correlation between (A) the isoelectric point and (B) Log P values and enrichment factor value obtained for the screening with the RKRKRK tagged GFP. The values obtained for the isoelectric point and LogP were estimated by the software Marvin Beans, ChemAxon

4.3.5. Screening with NNNNNN tagged GFP crude extract

The results obtained from the screening between the combinatorial library and NNNNNN tagged GFP (Fig. 4.27) showed extremely low enrichment factors. The maximum factor that the affinity ligands could enrich tagged GFP was 10, but in fact the maximum value reached by any ligand was 1, i.e. only 10% of the total possible enrichment. This peptide has a negative LogP, i.e. is hydrophilic. Under these circumstances, no putative lead ligands were selected for further studies and work on this affinity tag was discontinued.

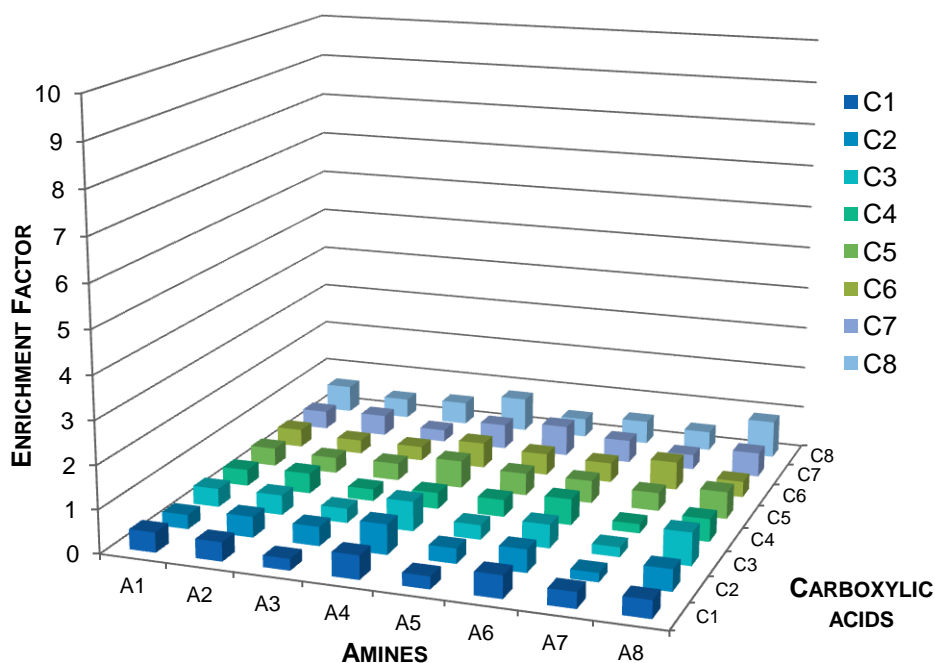


Figure 4.27 Screening of combinatorial library for NNNNNN tagged GFP. Ugi functionalized-Sepharose beads were synthesized in each well of the 96-well filtration blocks (250 mg of resin, c.v.: 250 μ l) and then prepared for chromatographic analysis through regeneration (0.1M NaOH, 30% (v/v) isopropanol alternated with H₂O, 3 c.v. each in a total of 6 c.v.), equilibration (PBS: 10 mM sodium phosphate, 150 mM NaCl, pH 7.4, 12 c.v.) prior to loading (1 c.v. of crude extract containing GFP tagged protein obtained from expression of GFP tagged protein in *E.coli* cells) and incubation (1h at 4°C). The flow-through (1 c.v.) and wash fractions (PBS 12 c.v.) were collected in 96 well black microplates. The fractions were quantified by spectrofluorimetry and BCA assays which determine the amount of GFP and total protein unbound, respectively. Ligand density was 20 μ mol / g moist weight gel. Enrichment factor = R1/R2, where R1 = (GFP tagged (or non-tagged) protein bound / total protein bound) and R2 = (GFP tagged (or non-tagged) protein loaded / total protein loaded).

4.3.6. Screening with non-tagged GFP extract

The library of affinity ligands was screened for binding to non-tagged GFP with two purposes: firstly is served as a negative control since the affinity ligand must recognize only the tag and not the protein; secondly, GFP itself can be regarded as an affinity tag and therefore there is interest in looking for specific receptors. At present, there is no specific and selective procedure for the purification of GFP fusion proteins in the literature apart from immunoaffinity separations. GFP has an isoelectric point \sim 5, indicating at pH 7.4 the protein should have an overall negative charge. The maximum value of the enrichment factor for GFP was determined as 6 (Fig 4.28) and the most promising affinity ligands such as A6C1 (Fig. 4.29) present an enrichment factor \sim 3.

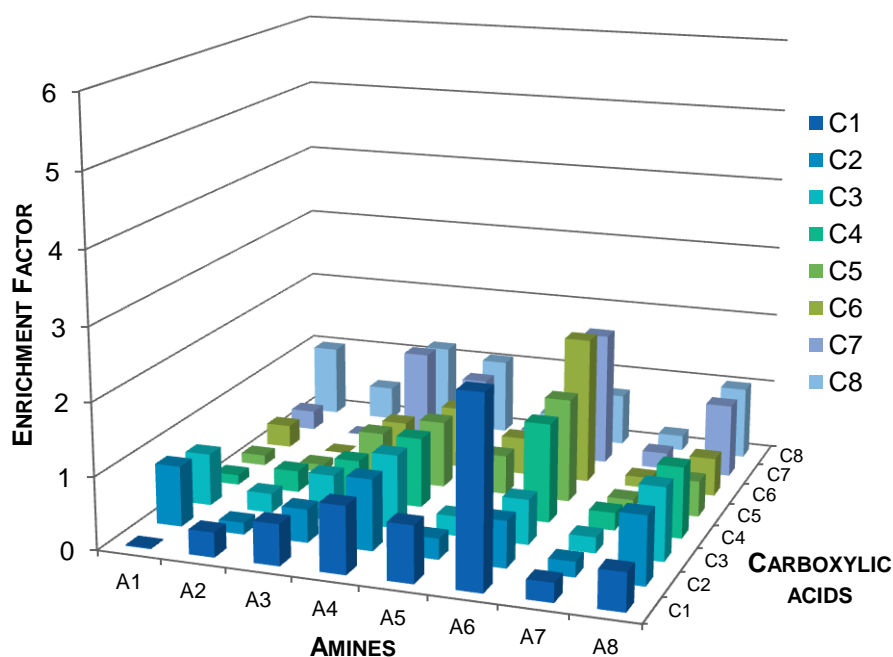


Figure 4.28 Screening of combinatorial library for GFP. Ugi functionalized-Sepharose beads were synthesized in each well of the 96-well filtration blocks (250 mg of resin, c.v.: 250 μ l) and then prepared for chromatographic analysis through regeneration (0.1M NaOH, 30% (v/v) isopropanol alternated with H₂O, 3 c.v. each in a total of 6 c.v.), equilibration (PBS: 10 mM sodium phosphate, 150 mM NaCl, pH 7.4, 12 c.v.) prior to loading (1 c.v. of crude extract containing GFP tagged protein obtained from expression of GFP tagged protein in *E.coli* cells) and incubation (1h at 4°C). The flow-through (1 c.v.) and wash fractions (PBS 12 c.v.) were collected in 96 well black microplates. The fractions were quantified by spectrofluorimetry and BCA assays which determine the amount of GFP and total protein unbound, respectively. Ligand density was 20 μ mol / g moist weight gel. Enrichment factor = R1/R2, where R1 = (GFP tagged (or non-tagged) protein bound / total protein bound) and R2 = (GFP tagged (or non-tagged) protein loaded / total protein loaded).

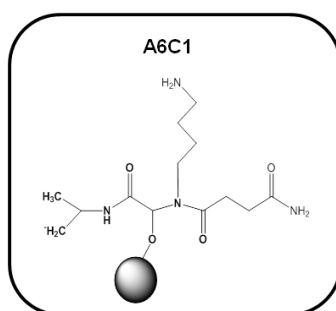


Figure 4.29 Lead ligand for GFP with the highest enrichment factor ~3. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

Ligand A6C1 was also selected as a putative lead ligand for the WFWFWF peptide tag under non-denaturing conditions and therefore it may be interacting not with the peptide tag but with GFP, although it was only detectable for the WFWFWF-GFP system. Ligands A6C4, A6C5, A6C6 and A6C7 (Fig. 4.30) were also presented good enrichment factors towards GFP, where ligands A6C6 and A6C7 were already selected for the RWRWRW tagged GFP system (§ 4.2.1).

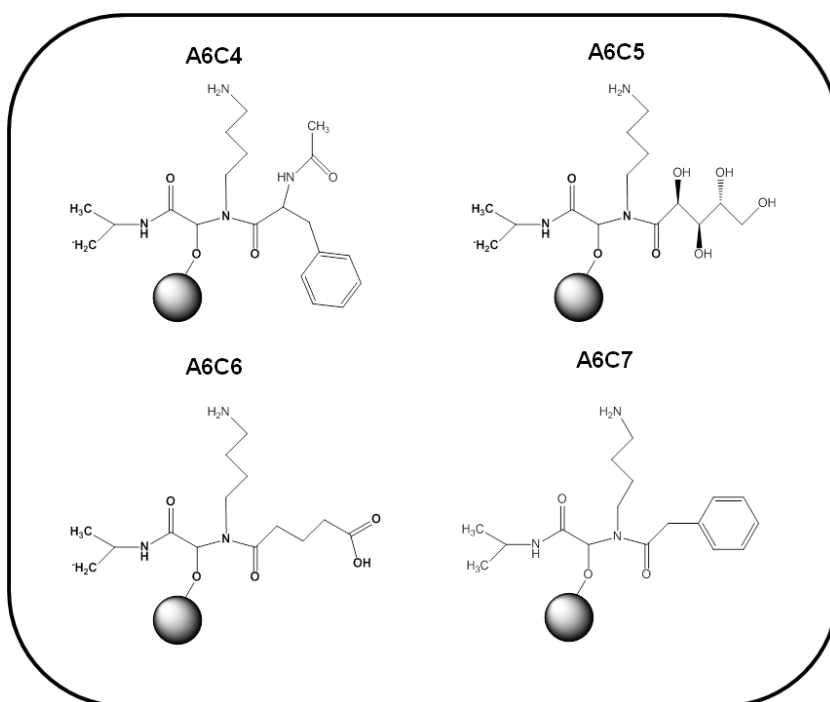


Figure 4.30 Ligands selected also as putative lead ligands with enrichment factor ~ 1.4 -2 for GFP. The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

On the other hand, ligands A6C4 and A6C5 were selected solely for non-tagged GFP protein. As these ligands were not rationally designed and this protein presents several hydrophobic patches [244], it is possible that affinity ligands comprising of hydrophobic moieties will bind to GFP. Affinity ligands A3C7, A3C8, A4C7 and A4C8 (Fig. 4.31) were also included for further studies although the enrichment factor was not high. Regarding A3C7 and A3C8, they also showed affinity for RWRWRW and WFWFWF (under non-denaturing conditions) peptides while A4C8 revealed affinity for WFWFWF tagged GFP under denaturing conditions.

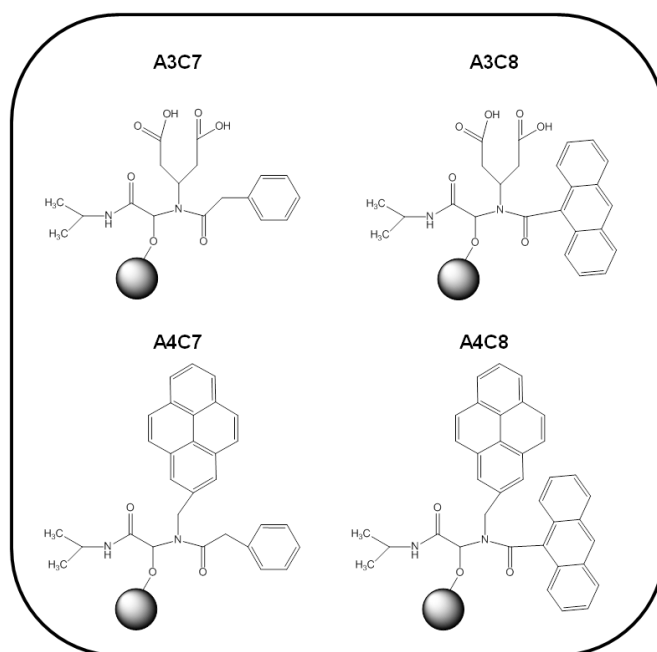


Figure 4.31 Lead ligands for GFP with an enrichment factor ~ 1 . The structures were designed in ChemDraw, CambridgeSoft, being the aldehyde group from the Ugi based ligand covalently linked to the solid support, agarose beads (●).

The isoelectric point and the LogP parameters were estimated for each of the lead ligands selected for GFP and were correlated with the respective enrichment factors (Fig. 4.31). In general, ligands displaying a higher isoelectric point also present higher values of enrichment (Fig. 4.32 (A)). At the screening conditions, the GFP displays an overall negative charge and the lead ligands, with pI higher than 7.4, present a positive charge.

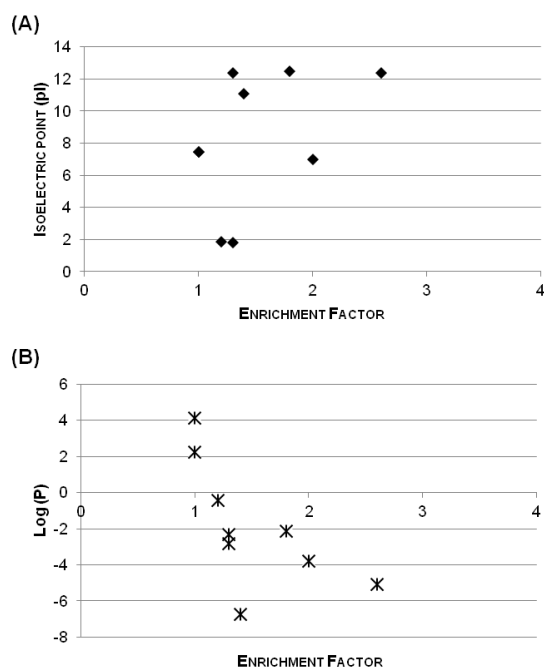


Figure 4.32 Correlation between (A) the isoelectric point and (B) Log P values and enrichment factor value obtained for the screening with non-tagged GFP. The values obtained for the isoelectric point and LogP were estimated by the software Marvin Beans, ChemAxon

Furthermore Fig.4.32 (B) indicates that the ligands with for high values of enrichment possess a more hydrophilic character. These facts might indicate that the electrostatic interactions might prevail over the hydrophobic interactions between the lead ligands and the GFP.

4.4 SELECTION AND RE-SCREENING OF THE PUTATIVE LEAD LIGANDS

Following the initial evaluation described in section 4.2, the putative lead ligands for each tagged- and non-tagged GFP were selected for further studies (Table 4.3).

Table 4.3 Putative lead ligands selected for each of GFP-tagged and non-tagged proteins.

GFP TAGGED AND NON-TAGGED PROTEINS	PUTATIVE LEAD LIGANDS
RWRWRW-GFP	A1C6; A2C4; A2C5; A2C8; A3C5; A3C6; A3C7; A3C8; A5C6; A6C6; A6C7; A6C8
WFWFWF-GFP (NON DENATURING CONDITIONS)	A1C5; A2C4; A2C5; A2C6; A2C7; A2C8; A3C1; A3C6; A3C7; A3C8; A6C1
WFWFWF-GFP (DENATURING CONDITIONS)	A4C5; A4C8; A8C8
NWNWNW-GFP	A2C2; A2C4; A2C5; A2C6; A3C1; A3C4; A3C5; A5C1; A5C6; A6C8; A7C8
RKRKRK-GFP	A2C4; A5C4; A7C1; A7C2; A7C3; A7C4; A7C5; A7C6; A7C7; A7C8
GFP	A3C7; A3C8; A4C7; A4C8; A6C1; A6C4; A6C5; A6C6; A6C7

Selected putative lead ligands were re-synthesized in duplicate in 96-well filtration blocks and the combinatorial libraries re-screened as previously (§4.2). Two different elution conditions were tested: 0.1M Glycine-HCl pH 3.0 and 0.1M Glycine-NaOH pH 11.0, except for the system WFWFWF-GFP in denaturing conditions, where the elution at pH 11 was substituted by 400 mM NaCl. The loading, flow-through and elution fractions were quantified for total protein and GFP concentration and were also analysed by SDS-PAGE (Fig. 4.33). The gels were ran under denaturing conditions and revealed by silver staining, which allows the detection of nanograms of protein ($\sim 0.1 \text{ ng/mm}^2$) according to the supplier. GFP alone has a molecular weight of 29kDa and the GFP-tagged protein $\sim 31\text{kDa}$. The additional 2kDa correspond to the affinity tags containing the hexapeptide sequences plus the cleavage sequence. Fig. 4.33 shows three typical examples of systems where GFP-tagged (or non-tagged), where more or less retained by the affinity adsorbents.

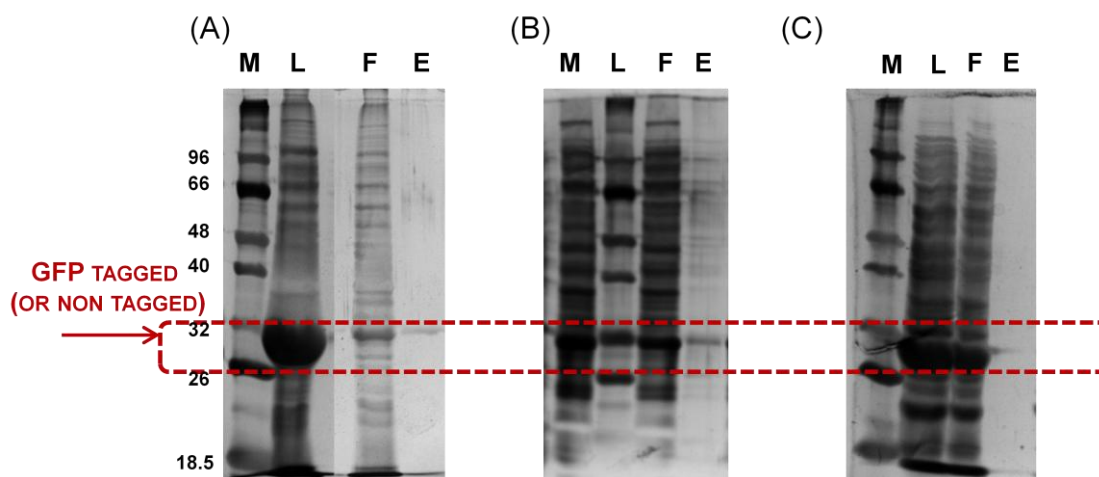


Figure 4.33 Estimation of GFP proteins retention onto the affinity adsorbents by SDS-PAGE analysis. M is the molecular weight marker, L the loading, F, the flow-through and E, the elution fraction. This image shows typical images of different GFP protein adsorption profiles, (A) corresponds to a high binding, (B) medium and (C) low binding for the affinity resins.

4.4.1. Re-Screening with RWRWRW tagged GFP crude extract

The results obtained from the screening of the RWRWRW tagged GFP are shown in Figs. 4.34 and 4.35. Putative lead ligands presented similar GFP binding percentages (40-50%) and enrichment factors values being these last slightly lower than those obtained in the first screening (Fig. 4.4). The difference between enrichment factors obtained from the two screening assays was more pronounced for ligands A3C6, A3C7, A3C8, A6C7 and A6C8. At a first glance in Fig 4.34 (A), Ligand A2C4 seemed to be the most promising ligand, presenting a percentage of RWRWRW tagged GFP around 45% and the total protein $< 15\%$, however this difference was not translated in a higher value of enrichment factor (Fig. 4.34 (B)).

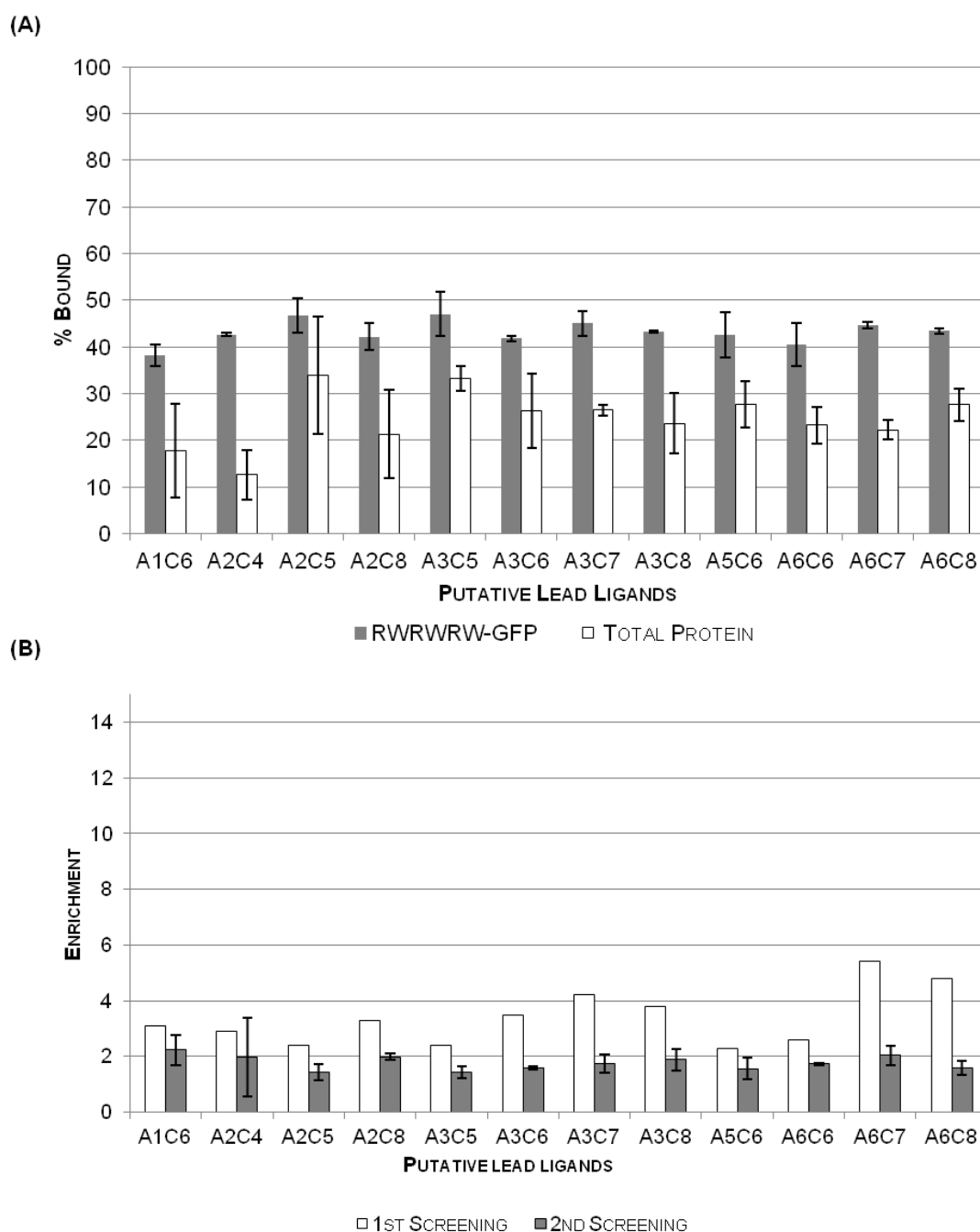


Figure 4.34 (A) Binding of GFP and total protein for the putative lead ligands and RWRWRW tagged GFP and (B) comparison of enrichment factor values obtained from the two screening assays. The screening was carried out as in Fig. 4.4. The error bars were obtained from duplicates (n=2).

From the analysis of the gels shown in Fig. 4.35, there was little evidence of an affinity ligand with GFP retention similar to the one presented in Fig. 4.33 (A), except for affinity ligands A5C6 or A6C6. However, no GFP protein is observed in the elution lanes. Ligand A2C4 presented a medium GFP retention according to Fig. 4.33 (B), and also a small band of RWRWRW tagged GFP under basic elution conditions. However, the amount of tagged GFP was not possible to

quantify. Ligands A6C7 and A6C8 presented a small band for tagged GFP in the elution; however, there was also evidence of contamination with host proteins. Overall, there was little evidence of a promising lead ligand able to retain GFP-tagged from host proteins with a relevant yield of recovery. Furthermore, this tag displayed antimicrobial properties and thus might not be ideal for recombinant protein purification purposes. No further studies were conducted on this tag.

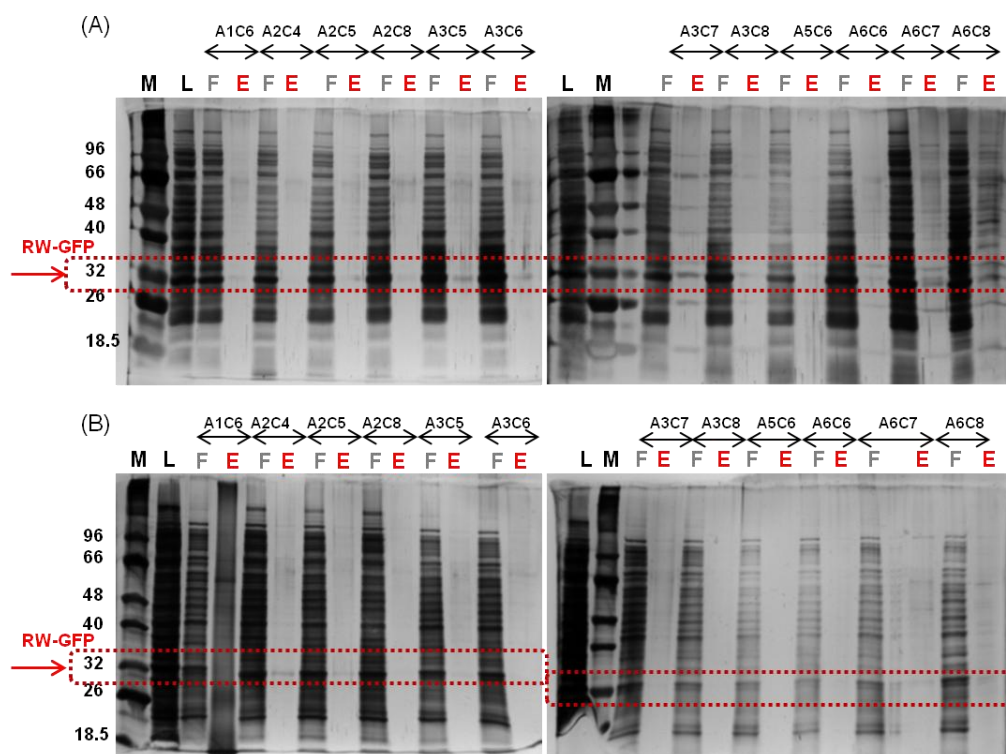


Figure 4.35 SDS-PAGE evaluation of the screening of putative ligands for RWRWRW-GFP. The Loading (L), flow-through (F) and elution fractions (E) obtained from the 96-well screening were loaded into SDS-PAGE gels (12.5% acrylamide), ran 1h at 120V and then silver stained. Two elution approaches were tested (A) 0.1M glycine-HCl pH 3 and (B) 0.1M glycine-NaOH pH 11. (M=molecular weight marker).

4.4.2. Re-Screening with WFWFWF tagged GFP crude extract

4.4.2.1 Non-Denaturing Conditions

From the analysis of Fig 4.36(A), all ligands showed a similar GFP binding profile. The enrichment factor values were maintained between the two screening assays, but with higher errors associated in this last screening. The first screening with the WFWFWF tagged GFP indicated ligand A6C1 as the most promising one, and although this ligand was not expected as a lead ligand due to its structure, the results obtained for A6C1 were maintained regarding the

enrichment factor value (10) and the GFP and total binding percentage was ~50% and 10%, respectively.

Ligands A3C8 and A3C4 also presented similar binding percentages as A6C1, however with lower enrichment factor values. Analysis of the elution profiles at pH 3 and 11 by SDS-PAGE gels (Fig. 4.37) was taken into account for further lead ligand selection. Under the specified elution conditions, the amount of tagged-GFP eluted was $\leq 1\%$ at pH 11, particularly for ligands A6C1 and A1C5. Under acidic elution conditions, there is no evidence of WFWFWF tagged GFP being eluted.

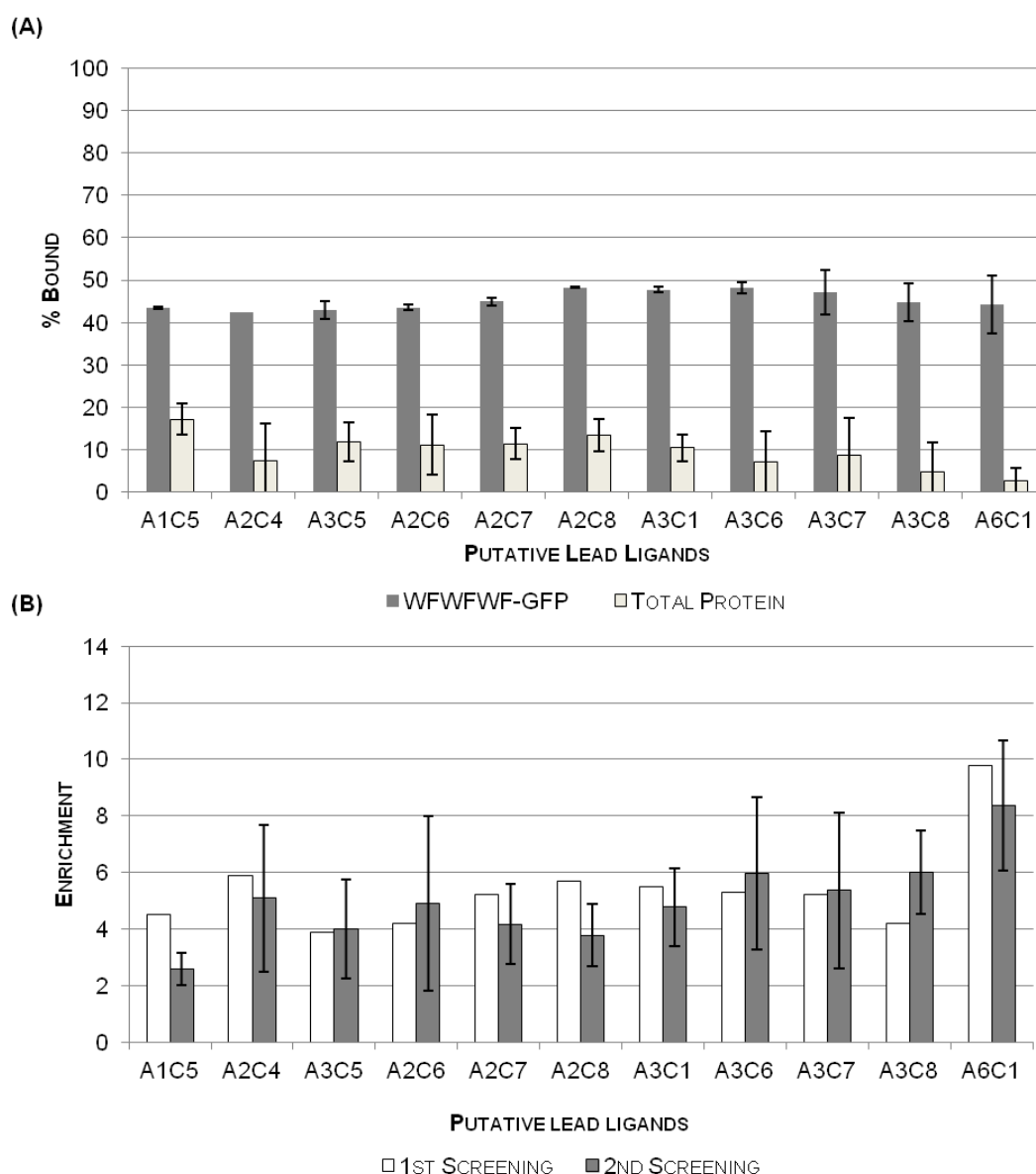


Figure 4.36 (A) Binding of GFP and total protein for the putative lead ligands and WFWFWF tagged GFP under non-denaturing conditions and (B) comparison of enrichment factor values obtained from the two screening assays. The screening was carried out as in Fig. 4.9 (n=2).

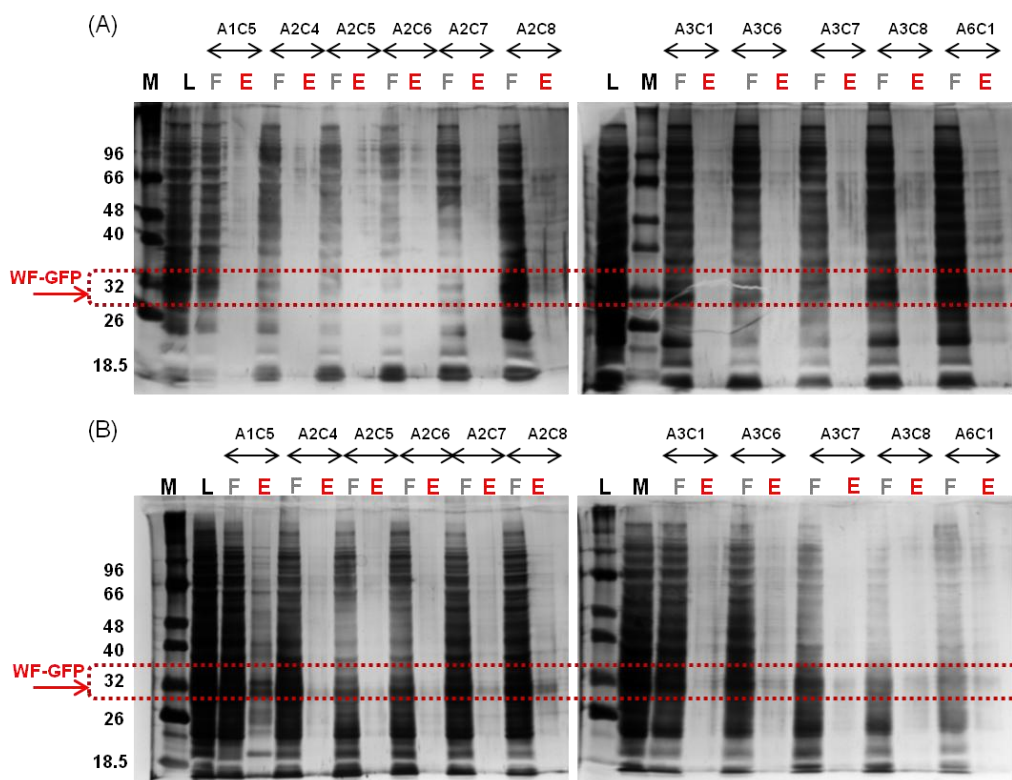


Figure 4.37 SDS-PAGE evaluation of the screening of putative ligands for WFWFWF tagged GFP. The Loading (L), flow-through (F) and elution fractions (E) obtained from the 96-well screening were loaded into SDS-PAGE gels (12.5% acrylamide), ran 1h at 120V and then silver stained. Two elution approaches were tested (A) 0.1M glycine-HCl pH 3 and (B) 0.1M glycine-NaOH pH 11. (M=molecular weight marker).

The SDS-PAGE gels for this GFP-tagged protein were not very promising as the flow-through lanes of all the affinity ligands were similar to the profile in the loading lane. Ligands A1C5 and A2C8 showed a band of the WFWFWF tagged GFP under basic elution conditions, and ligand A6C1 showed a band of GFP-tagged at acidic elution conditions (Fig. 4.37 (B)), although there were also a significant amount of contaminants being eluted. Thus, SDS-PAGE analysis indicated that selected ligands were not particularly selective for the WFWFWF tagged GFP.

4.4.2.2 Denaturing Conditions

Screening of the WFWFWF tagged GFP for the putative lead ligands under denaturing conditions in 8M urea was more promising for ligands A4C5, A4C8 and A8C8 which showed the best binding characteristics, confirmed by SDS-PAGE gels (Fig. 4.38 and 4.39). These affinity ligands presented a flow-through lane very similar to that shown in Fig. 4.33 (A), suggesting that these ligands are preferentially binding to the GFP-tagged over the total host proteins present in the crude extract.

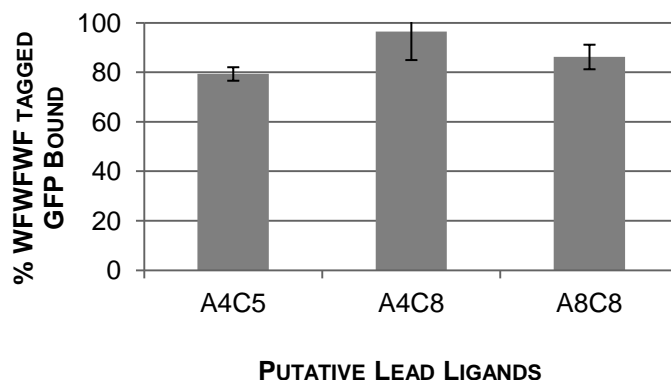


Figure 4.38 Binding of GFP for the putative lead ligands and WFWFWF tagged GFP under denaturing conditions. The screening was carried out as in Fig. 4.14 with n=2.

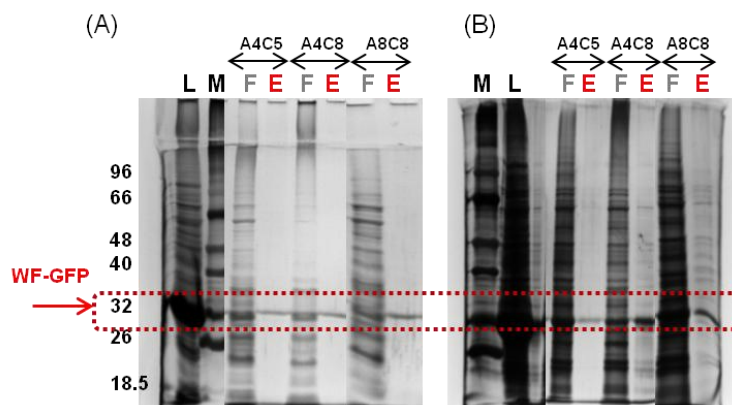


Figure 4.39 SDS-PAGE evaluation of the screening of putative ligands for WFWFWF tagged GFP under denaturing conditions. The Loading (L), flow-through (F) and elution fractions (E) obtained from the 96-well screening were loaded into SDS-PAGE gels (12.5% acrylamide), ran 1h at 120V and then silver stained. Two elution approaches were tested (A) 0.1M glycine-HCl pH 3 and (B) 0.1M glycine-NaOH pH 11. (M=molecular weight marker).

Furthermore, the elution band for the GFP-tagged from ligand A4C8 in the presence of elution buffer containing 400mM of NaCl was significant, although these GFP-tagged proteins did not possess fluorescence, suggesting the elution of GFP-tagged in a denatured form. Ligand A4C8 was selected as the lead candidate for further optimization. This selection was based on the higher percentage of GFP bound, i.e. the low percentage of GFP-tagged band present in the flow-through when compared with the loading sample. However, as these ligands bound selectively to GFP-tagged, the strategy suggested is to perform the refolding of GFP-tagged on-column prior to the elution step, an approach already explored by other authors [49, 138]. These studies will be performed in an automatic system such that the proper negative gradient of urea and slow positive increments of salt can be achieved in order to enhance the likelihood of

proper folding (Chapter 5 § 5.6.1). It is expected that hydrophobic interactions are predominant for this affinity pair “tag-receptor” according to Fig4.46. Moreover, the overall charge of the affinity ligand is neutral and the tag seems to be slightly positive pH 7.4 (Fig. 4.46 and Table 4.2) meaning that might not be occurring ionic interactions between the affinity system.

4.4.3 Re-Screening with NWNWNW tagged GFP crude extract

The screening of lead ligands for NWNWNW tagged GFP showed that the binding percentage were similar for all selected putative lead ligands, corresponding to ~40-50%, for GFP-tagged and ~20-30% for total protein (Fig. 4.41 (A)). In terms of enrichment factor (Fig. 4.40 (B)), the values obtained in the re-screening were lower compared with those obtained previously. Therefore, the evaluation of the GFP-tagged retention and respective recovery through SDS-PAGE analysis (Fig. 4.41) might contribute for a better selection of the lead ligands.

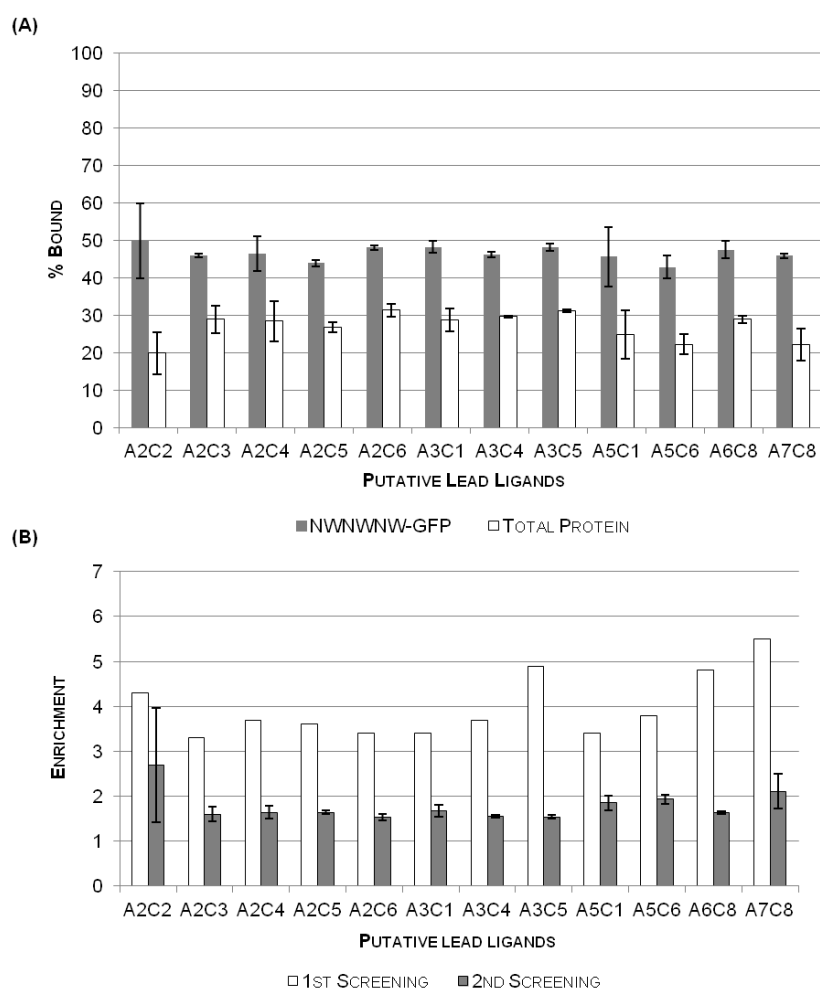


Figure 4.40 (A) Binding of GFP and total protein for the putative lead ligands and NWNWN tagged GFP and (B) comparison of enrichment factor values obtained from the screening assays. The screening was carried out as in Fig. 4.16. with n=2.

Regarding the elution at pH 3, the ligands A2C5, A2C6, A3C1, A3C4 A5C6, A6C8 and A7C8 seemed to display a GFP-tagged band on the elution lane (Fig 4.41 (A)). However, no protein was detected by GFP fluorescence. Basic elution conditions were also not effective on the GFP-tagged recovery and eluted only 0.1%, 0.03% and 0.01% for ligands A3C1, A3C4 and A6C8, respectively. However, purity of the eluted fractions seemed greater at pH 11 (Fig. 4.41(B)). From the SDS-PAGE analysis, there is a wide range of ligands binding preferentially to tagged-GFP.

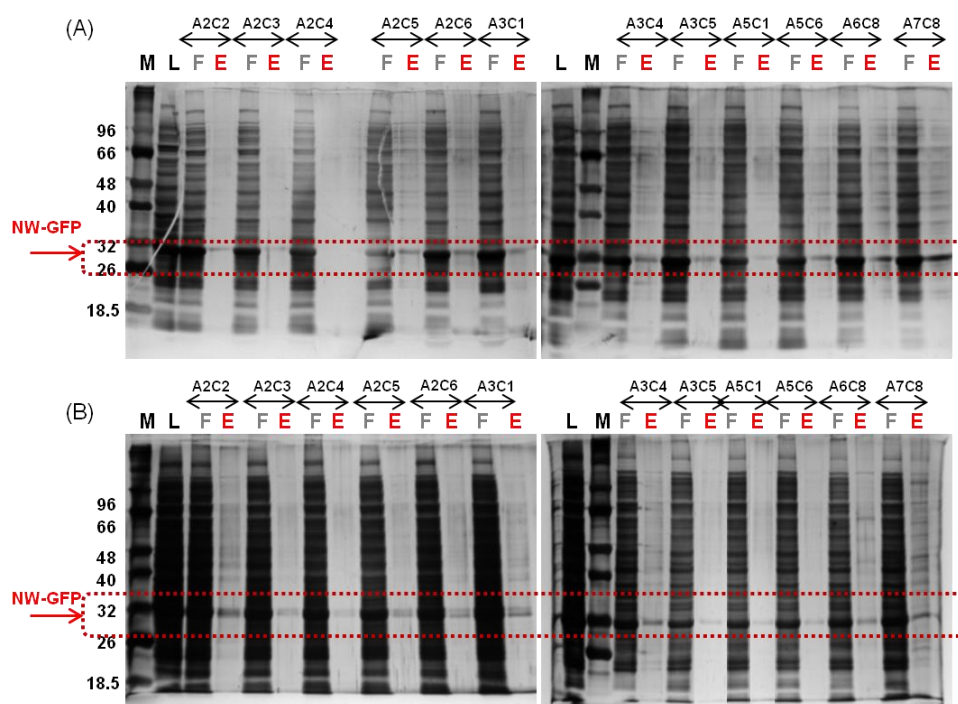


Figure 4.41 SDS-PAGE evaluation of the combinatorial ligand screening of the putative ligands and NWNWNW tagged GFP. The Loading (L), flow-through (F) and elution fractions (E) obtained from the 96 well format screening were loaded into SDS-PAGE gels (12.5% acrylamide), run 1h at 120V and then staining with silver staining. The screening was carried as described before but with different elution approaches (A) 0.1M glycine-HCl pH 3 and (B) 0.1M glycine-NaOH pH 11. (M=molecular weight marker)

Ligands A2C2 and A3C4 present similar structures and a more promising profile in terms of tagged-GFP adsorption/elution at both pH values. However, at pH 11 the elution profile seemed to show a purer product. Affinity ligands A6C8 and A7C8 also showed a band of GFP tagged protein on the elution profile. For A7C8 purity in the elution fraction was low at either pH values. Affinity ligands A2C3 and A2C4 did not show any eluted GFP-tagged. Ligands A3C5, A2C5 and A2C6 present a highly flexible structure, which might explain why these were selected as putative lead ligands for other GFP-tagged systems as RWRWRW-GFP and WFWFWF-GFP. Overall, the best candidates as lead ligands were the affinity ligands A3C4 and A6C8; however, the elution profile for A3C4 seemed to yield a purer product and the eluted amount quantified by GFP fluorescence was higher. At the pH of binding (pH 7.4), the affinity tag presents a slightly

positive charge and the lead ligand is negatively charged. The predominant interactions occurred between the two entities are expected to be hydrogen bonding and hydrophobic interactions

4.4.4 Re-Screening with RKRKRK tagged GFP crude extract

The potential lead ligands for RKRKRK tagged GFP were retested for binding and eluting target protein. The percentage of GFP-tagged bound and the enrichment factor values were similar between the two screening tests, revealing reproducibility between the screening assays performed for the lead ligand selection. Ligands with amine A7 showed a higher percentage of GFP-tagged bound than other ligands such as A2C4 and A5C4 (Fig. 4.42(A)).

The ligands A2C4 and A5C4, allowed the elution of GFP-tagged but only at pH 11 (Table 4.4) and with a yield of ~10% of bound protein. It is worth noticing that at pH 3 (Fig 4.43 (A)), the elution profiles showed two intense bands. In general samples eluted at pH 11 seemed to present a higher degree of purity. Ligands A2C4 and A5C4 presented the lowest percentage of GFP-tagged bound and a medium protein adsorption.

In terms of purity, the elution profiles at pH 3 showed a low purity whilst at pH 11, no tagged-GFP elution was observed. The most promising results were obtained with the ligands possessing amine A7, such as A7C1, A7C6, A7C7 and A7C8. Affinity ligands A7C2 and A7C3 displayed a lower GFP band on the flow-through but the elution profile presented a higher amount of contaminants. Thus the best candidates for lead ligands were A7C1, A7C6 and A7C7, although ligand A7C7 might be too hydrophobic. Affinity ligands A7C1 and A7C6 were similar in terms of binding and elution. Ligand A7C8, displayed a pure eluted fraction of GFP-tagged under basic elution conditions but it was not selective for this affinity tag. Therefore, the lead ligand selected was A7C1. Under the binding conditions, the affinity tag is highly charged due to the presence of the amino acids lysine and arginine. The respective lead ligand presents a negative charge, due to the deprotonation of the existent carboxylic acids (Fig 4.46), and the predominant binding interactions are electrostatic also according to Fig. 4.26 A). However, as the affinity tag is flexible due to the long chains of arginine and lysine aminoacids, it is hypothesized that the carboxylic acids of the ligand, linked to an aromatic ring, contribute for increased rigidity in the interaction with the affinity tag.

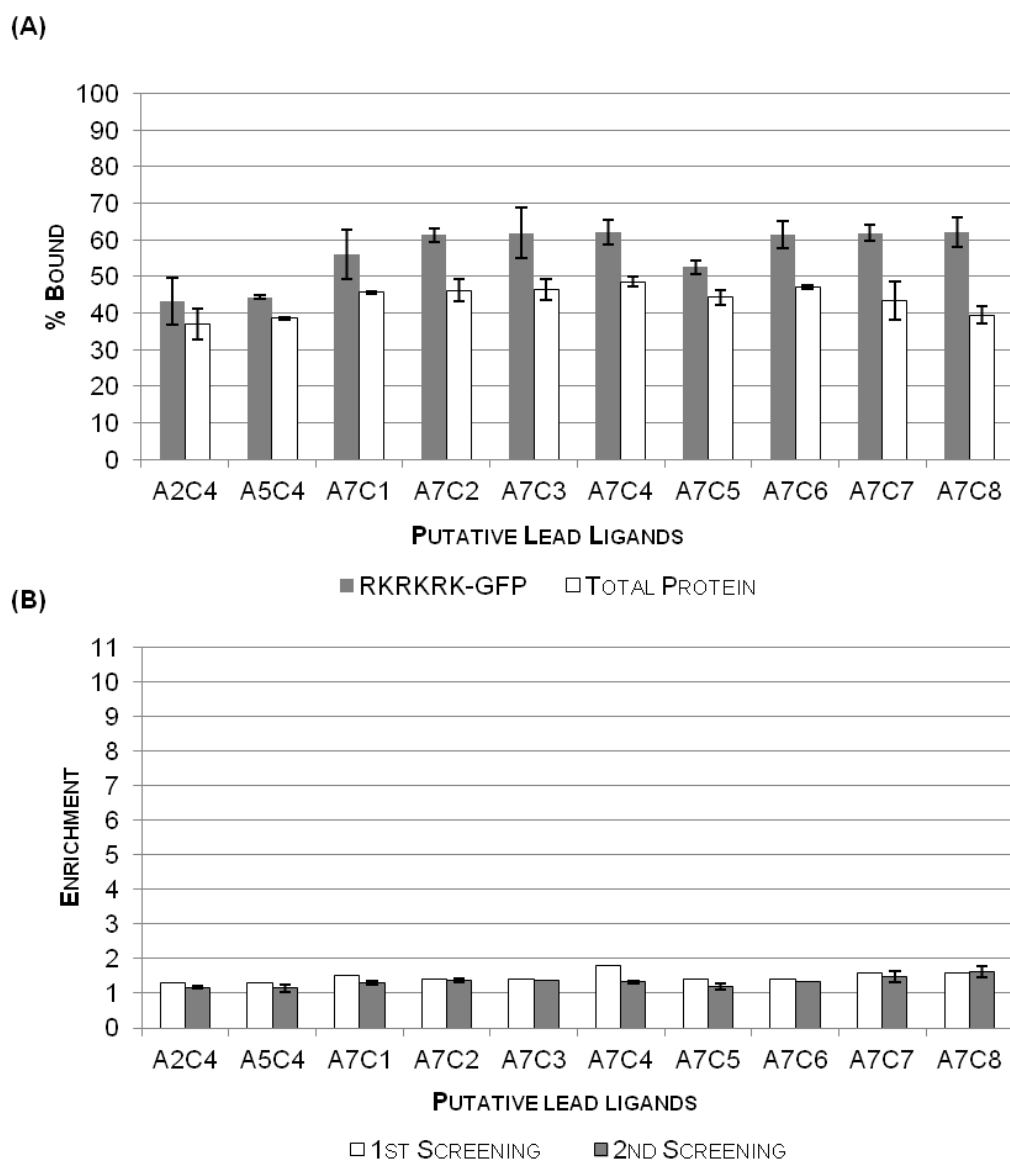


Figure 4.42 (A) Binding of GFP and total protein for the putative lead ligands and RKRKRK tagged GFP and (B) comparison of enrichment factor values obtained from the screening assays. The screening was carried out as in Fig. 4.22 with n = 2.

Table 4.4 Elution of RKRKRK tagged GFP at pH 11. After protein binding onto the affinity resins, 5 c.v. of the elution buffer (0.1M glycine-NaOH pH 11 was added in each one of the wells of the 96-well blocks containing the promising lead ligands. All elution fractions were then quantified by the GFP fluorescence in order to obtain the amount of GFP-tagged eluted.

LEAD LIGANDS	RKRKRK TAGGED GFP
	ELUTED (%)
A7C1	3.5
A7C2	6.1
A7C3	6.8
A7C4	5.9
A7C5	0.6
A7C6	9.5
A7C8	8.5
A7C8	8.5

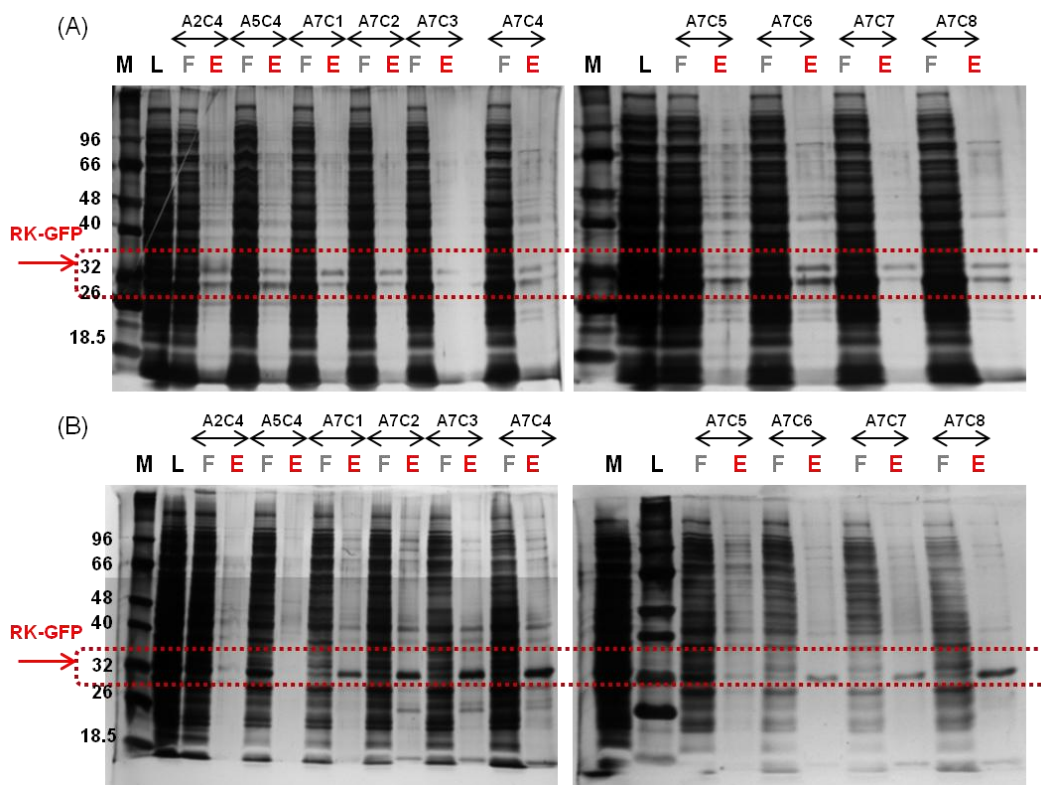


Figure 4.43 SDS-PAGE evaluation of the screening of putative ligands for RKRKRK tagged GFP. The Loading (L), flow-through (F) and elution fractions (E) obtained from the 96-well screening were loaded into SDS-PAGE gels (12.5% acrylamide), ran 1h at 120V and then silver stained. Two elution approaches were tested (A) 0.1M glycine-HCl pH 3 and (B) 0.1M glycine-NaOH pH 11. (M=molecular weight marker).

4.4.5 Re-Screening with non-tagged GFP crude extract

The results obtained for the re-screening of the promising lead ligands and the non-tagged GFP (Fig 4.44) revealed consistency and reproducibility with those obtained in the first screening in terms of GFP enrichment, and also indicated that ligands with the highest GFP binding percentage showed the lowest enrichment factors values.

GFP displays several patches of hydrophobic amino acids on its surface and can thus establish hydrophobic interactions with the affinity ligands. Ligands with higher hydrophobic character bound ~100% GFP, even if they did not present the highest enrichment factor values (Fig. 4.44(B)). Alkaline pH appears to be the elution condition that allows a higher proportion of GFP to be recovered, in particular for A4C7 with ~5% of eluted protein. The remaining ligands presented a protein recovery below than 3%. Affinity ligands A4C7 and A4C8 displayed a small band of GFP in the flow-through and a highly pure band on elution (Fig. 4.45 (B)).

Based on the flow-through of the two candidates as lead ligands, ligand A4C7 seemed to be more promising as this lane possesses a higher amount of unbound proteins than A4C8 and thus A4C7 was considered as lead ligand for GFP. Although the affinity ligand A4C7 was not very selective for binding to GFP under the tested conditions, presented potential for the selective recovery of the protein GFP (§ 5.5.3) At pH 7.4, the GFP protein is negatively charged according with the isoelectric point of GFP around 5 [206], whereas the lead ligand presents a neutral charge (Fig 4.46). The aromatic and benzene ring from the ligand can behave as hydrogen bonding acceptor and also are available to perform hydrophobic interactions. The carbonyl groups and the NH groups from the ligand backbone can act as acceptor and donor, respectively, of hydrogen bonding [206, 245].

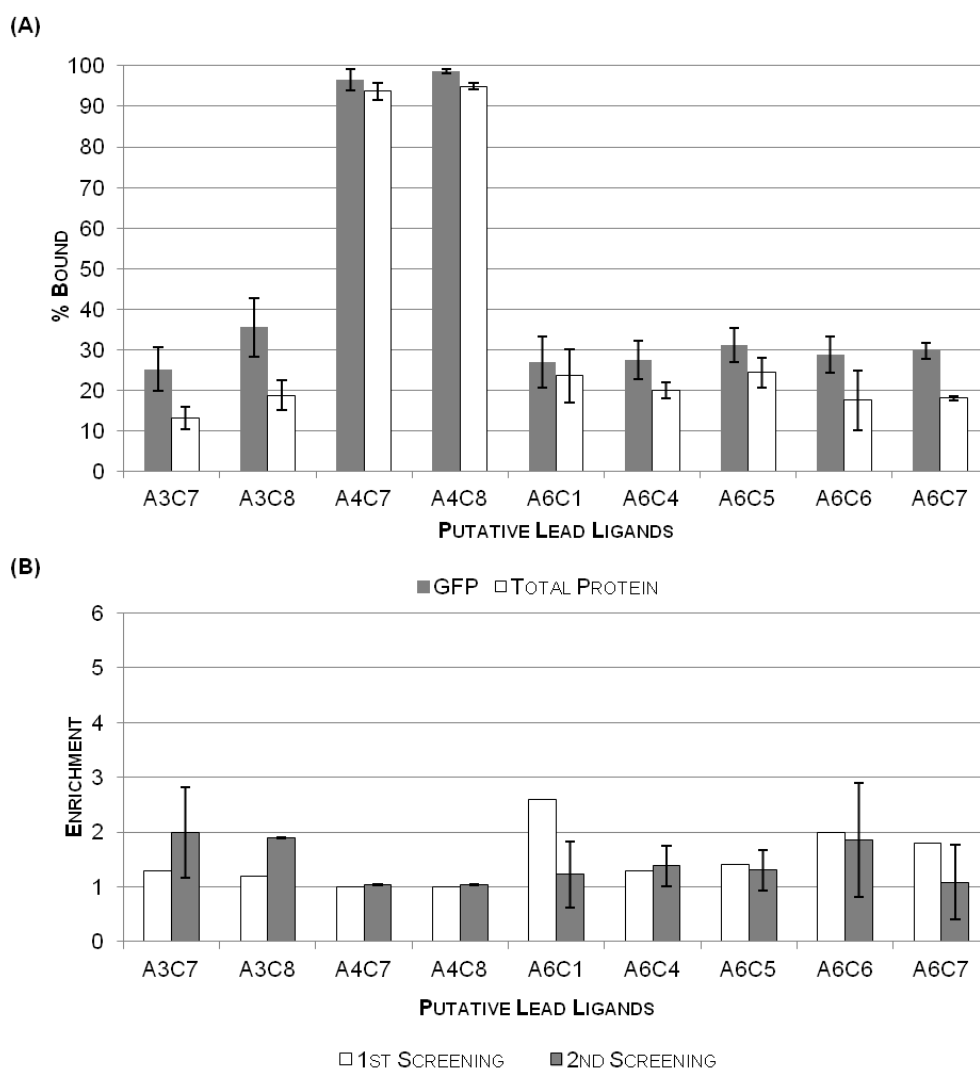


Figure 4.44 (A) Binding of GFP and total protein for the putative lead ligands and GFP and (B) comparison of enrichment factor values obtained from the screening assays. The screening was carried out as in Fig. 4.28 with n=2.

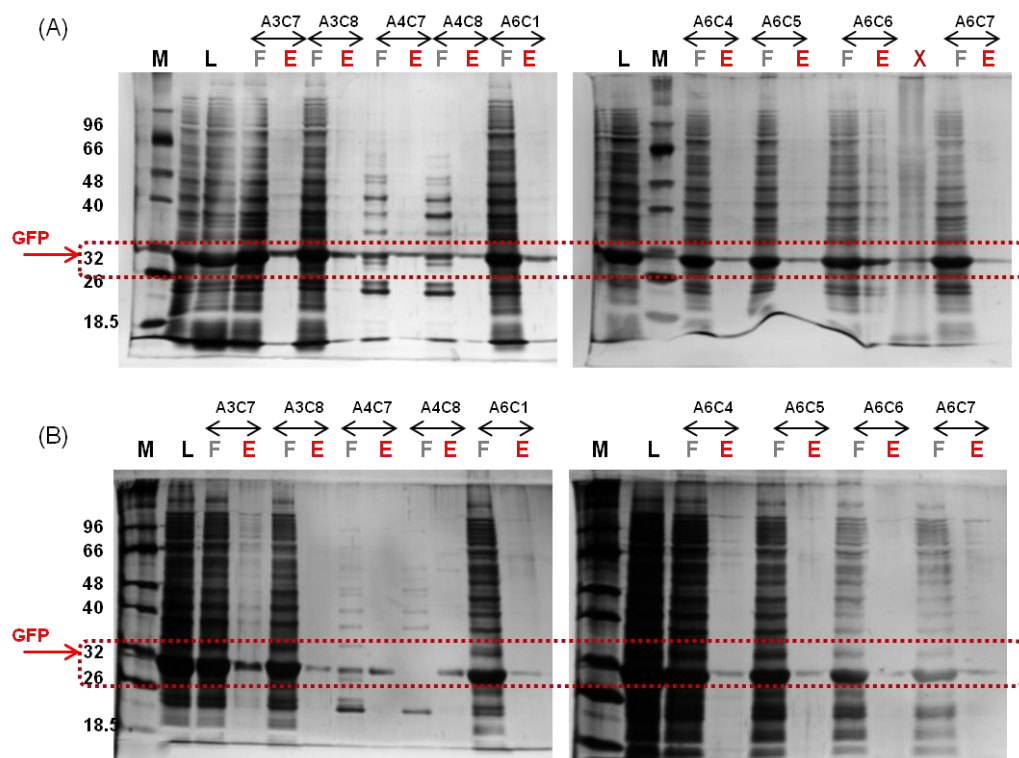


Figure 4.45 SDS-PAGE evaluation of the screening of putative ligands for non-tagged GFP. The Loading (L), flow-through (F) and elution fractions (E) obtained from the 96-well screening were loaded into SDS-PAGE gels (12.5% acrylamide), ran 1h at 120V and then silver stained. Two elution approaches were tested (A) 0.1M glycine-HCl pH 3 and (B) 0.1M glycine-NaOH pH 11. (M=molecular weight marker).

4.5. THE LEAD AFFINITY PAIRS “TAG-RECEPTOR”

The synthesis and screening of an Ugi-library of affinity ligands with cell culture supernatants extracts of *E. coli* producing tagged and non-tagged GFP proteins yielded three potential affinity pairs “tag-receptor”, namely, ligand A4C8 for the tag WFWFWF under denaturant conditions, ligand A3C4 for the tag NWNWNW and ligand A7C1 for the tag RKRKRK. In addition, the synthetic ligand A4C7 presented ability for the selective recovery of non-tagged GFP (Fig. 4.46).

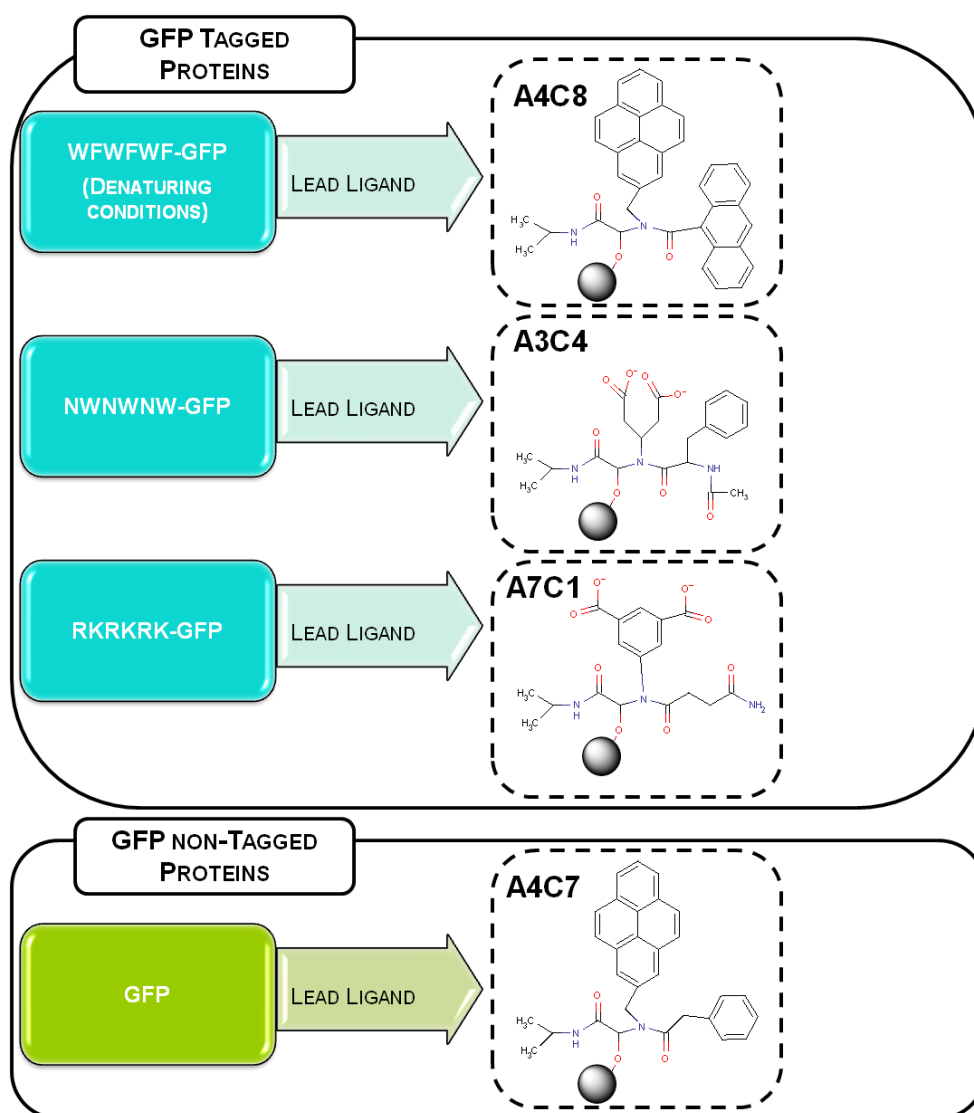


Figure 4.46 Lead ligands selected for GFP tagged proteins and non-tagged GFP. The structure of the lead ligands for the respective are represented at pH 7.4. The ligand was designed and then the overall charge was determined by using the software Marvin Beans, ChemAxon.

A comparison of the amount of GFP bound by each one of the lead systems is given in Fig. 4.47. It is worth noticing that the binding conditions for each affinity system were not optimized in this work, as it would have involved dialysis of the *E.coli* crude extracts with different buffers, a time consuming and costly approach.

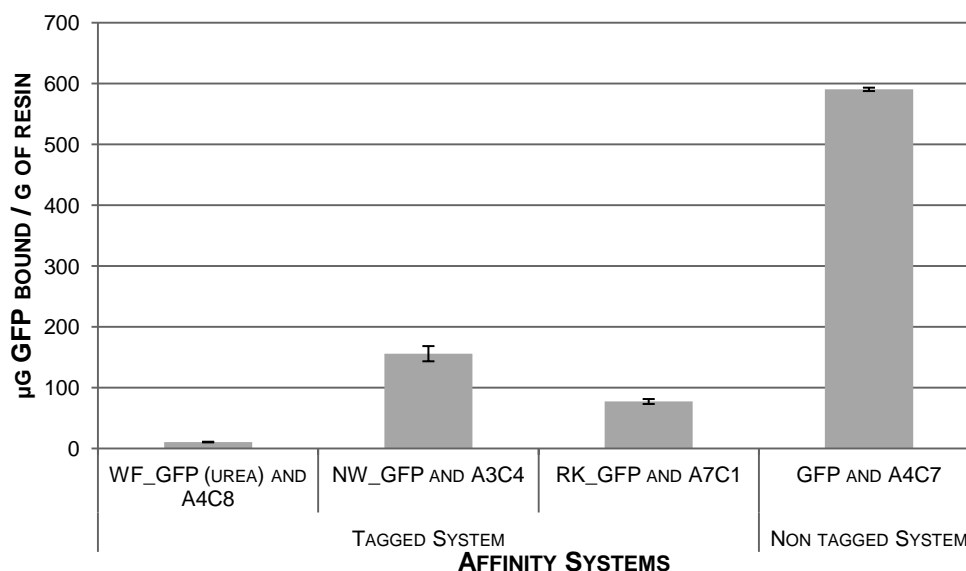


Figure 4.47 Binding evaluation (μg bound/ g resin) of each one of the GFP tagged proteins and non-tagged GFP and the respective affinity ligand. The binding assay was carried out on 96 well format, where each one of the well presented the Ugi functionalized-Sepharose beads (250 mg of resin). The binding evaluation assay was conducted in several stages as: regeneration (0.1M NaOH, 30% isopropanol (v/v) alternated with H_2O , 3 c.v each in a total of 6 c.v.), equilibration with binding buffer (10 mM sodium phosphate, 150 mM NaCl, pH 7.4 12 c.v), loading(1 c.v of the respective crude extract was loaded: WF-GFP in presence of 8M urea 12 μg loaded/g of resin, NW-GFP 265 μg loaded/g of resin, RK-GFP: 150 μg loaded/g of resin and for GFP 593 μg loaded/g of resin) and then incubation (1h at 4°C). The flow-through (1 c.v) and washes fractions with binding buffer or 4M urea 6 c.v followed by 6 c.v. binding buffer for A4C8 were collected in 96 well black microplates. The fractions were quantified by spectrofluorimetry. Ligand density = 20 μmol / g moist weight gel). (Protein bound = (Protein loaded)-(sum of the amount protein in flow-through and washes). The errors bars presented in the graph correspond to the standard deviation of 36 samples except for WF-GFP (urea) that corresponds to duplicates.

4.6 CONCLUDING REMARKS

The screening of the combinatorial libraries based on the Ugi reaction led to the selection of three ligands, each of which displayed affinity for a tag fused to GFP, and a further lead ligand with affinity for GFP as shown in the Fig 4.46. The amine component of the Ugi based ligand seems to have a pivotal role in the molecular recognition of the tags. The 96-well screening format proved to be a reproducible, easy to perform and time saving protocol.

All three systems present different characteristics: The WFWFWF tagged GFP/A4C8 pair was selected in denaturing conditions, i.e. in 8M urea. This tag might be useful for protein purification where the proteins are produced as inclusion bodies, and is presents potential for on-column refolding prior to elution. The second affinity system pair corresponds to tag

NWNWNW/A3C4 and the third system to RKRKRK/A7C1. Both systems were selected by screening under physiological conditions. However, in the case of NWNWNW, this tag presents a more neutral character compared to the tag RRKRKRK, which is positively charged. In addition, an affinity ligand (A4C7) was selected for GFP, which might be useful in future for the purification of GFP fusion proteins. However, these preliminary results required optimization in terms of full characterization of the binding thermodynamics, elution conditions optimization and scale up in an automated system

CHAPTER 5

OPTIMIZATION AND CHARACTERIZATION OF THE LEAD AFFINITY PAIRS “TAG-RECEPTOR”

SUMMARY

This chapter focus on the detailed study and optimization of the lead affinity pairs “tag-receptor”. For both pairs tag-receptor NWNWNW-GFP/A3C4 and RKRKRKR-GFP/A7C1, the purification was carried out as a standard affinity chromatography process, being the former more adequate for recombinant protein purification. Ligand A3C4 presented a static binding capacity of 0.34 ± 0.59 mg of GFP-tagged per mg of resin displaying an affinity constant $7.22 \times 10^4 \text{ M}^{-1}$. Ligand A7C1 presented slightly higher static binding capacity (0.47 ± 0.57 mg of RKRKRK tagged GFP per mg of resin) and displayed an affinity constant $7.45 \times 10^5 \text{ M}^{-1}$. The affinity pair WFWFWF-GFP/A4C8 was selected under denaturing conditions (e.g. presence of urea), with a subsequent matrix-assisted refolding stage. An affinity ligand toward GFP was also assessed with a static binding capacity of 1.02 ± 6.85 mg of GFP per mg of resin displaying an affinity constant of $2.38 \times 10^5 \text{ M}^{-1}$. This was extended for the purification of GFP fusion proteins produced on *Lactococcus lactis*. Regarding the elution conditions optimization, the best elution conditions were competitive elution with arginine (500 mM) under mild conditions (PBS pH 7.4 150 mM NaCl) and under alkaline conditions such as 0.1M glycine-NaOH pH 11 with 150 mM NaCl for the affinity pair RKRKRK tagged GFP/A7C1. In case of non-tagged GFP/A4C7, the best elution conditions were also observed under alkaline conditions such as 0.1M glycine-NaOH pH 9 with and without 50% ethylene glycol. All the attempts used to recover NWNWNW tagged GFP were unaccomplished due to the extreme low amounts of protein recovery (< 3%) and low reproducibility of the system. In terms of the scale-up by using an automated liquid preparative chromatography such as ÄKTA avant 25, the systems RKRKRK tagged GFP/A7C1 and non-tagged GFP/A4C7 revealed to be promising affinity pairs to be further used on purification of recombinant proteins.

5.1 INTRODUCTION

A successful purification process based on affinity chromatography requires a highly specific capture step of the target protein and its full recovery with high purity [35, 246]. The capture is an adsorption step that relies on the specific binding between the target protein and the complementary affinity ligand, with a minimization of the non-specific interactions. The highly specific binding can be a result of the establishment of interactions such as electrostatic, hydrophobic, hydrogen bonding and van der Waals' forces [9]. However, the binding efficiency also depends on the kinetics of the interaction and this can be influenced by concentration of the target protein, the amount of the immobilized ligand and the flow rate [34-35]. The binding conditions are critical for the establishment of a favourable affinity environment and this means that optimal components as molarity, pH and ionic strength should be taken into account [34-36]. After the binding step, the elimination of the undesirable contaminants and unbound species must be assured by a washing step [36]. The following elution step consists on the disruption of the ligand-target protein complex [34-36, 246-247]. The choice of the eluents needs to be a compromise between the total recovery of the bound target protein and the purity and maintenance of the biological activity and integrity of the target protein [34-36, 246-247].

The optimization of binding and elution conditions can be performed on a 96-well micro-scale format because of the great advantages of using wide range of adsorbents and different operating conditions [225-226, 248]. The 96-well microscale procedure is a batch process and because of this case the target biomolecule is incubated with the respective affinity matrix. This allows a better mixing between the two entities, nevertheless the flow stops after the application [36]. In this way, after the selection of the best binding conditions, these conditions need to be applied in columns where the flow rate is taken into account. In case of using small packed columns without an automated system driven by gravity, the flow rate will depend on the type of matrix and the physical dimensions of the column [36]. In this way, the parameters that characterize the maximum binding between the ligand and the target in small scale experiments, needed to be adjust on the scaling up. Moreover, the differences on the flow rate from small to scale up can also influence the yield of purified protein [36].

The main aim of this chapter was the optimization and characterization of the lead affinity systems (Fig. 5.1), which involved (i) the improvement of the elution profiles, performed in a rational and individual manner according with the different properties of each affinity system; (ii) determination of the affinity constants and the static binding capacity of each affinity system and (iii) studies with an automated preparative liquid chromatography system AKTA™ avant 25. The latter involved higher amounts of adsorbent bead volume, a better evaluation of the parameters (e.g. flow rate) and a detailed analysis of each one of the steps involved in the affinity chromatography process.

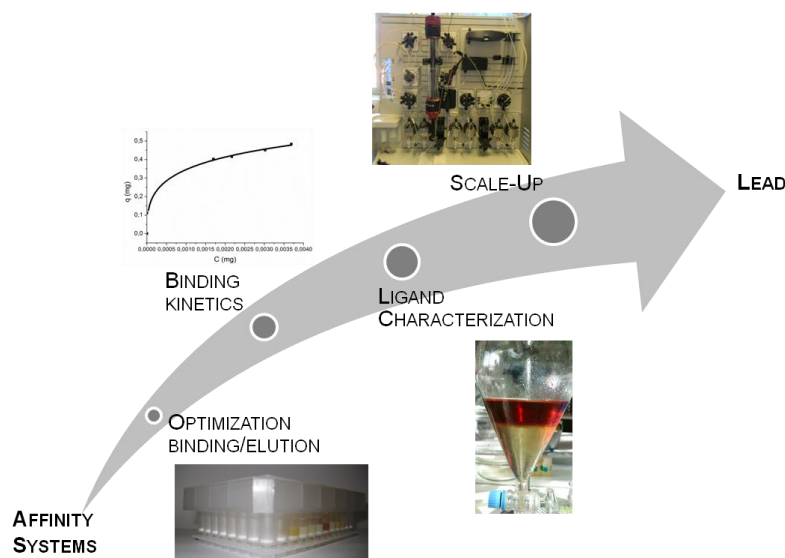


Figure 5.1-Different stages required for the optimization of the lead affinity systems.

5.2. COMPARISON OF THE AFFINITY SYSTEMS PERFORMANCE ON 96-WELL FORMAT AND ON-COLUMN

The binding evaluation for each lead ligand was initially carried out in 96-well filtration blocks (0.8 cm of well diameter and 3 cm well height) and in a batch mode. The same amount of adsorbent (0.25 mg) used in each well containing the lead ligand was packed on small columns (0.8 cm of diameter and 6.5 cm of height) and the assays performed under gravity. Overall, results obtained from 96-well format and on-column are comparable (Fig. 5.2).

The main difference of the two methods was related with the incubation of the crude extract with the immobilized ligands, the ligand packing and the flow rate employed. In the 96-well format, the protein samples were incubated 1h at 4°C with the adsorbents prior washing, the ligand was not packed and all the washes were performed through a centrifugation step. Regarding the studies on column, the affinity adsorbents were packed on the columns, the crude sample was not incubated with the resin, passing through the resin under a gravitational flow (0.75 ± 0.09 ml/min), originating a lower residence time of the sample within the resin. Although ligand density, adsorbent amount, and loading sample were maintained constant, it was expected that the binding percentage determined by the 96-well format would be higher when compared with those obtained on column, as the interaction between ligand and protein was promoted in the former (Fig 5.2). In terms of percentage of GFP bound (Fig 5.2A), there was not a significant decrease of binding, which further confirms the selectivity of the systems. As shown in Fig 5.2 B, the total protein bound on column was significantly lower (less than 10%), except for the non-

tagged system. This indicates that the on-column binding can contribute for the maintenance of the selectivity of the ligands but with reduced non-specific interactions. Regarding the non-tagged system, ligand A4C7 presented a high percentage of binding for the GFP protein, and also for the total protein, as already observed in Chapter 4. The selectivity of this system is therefore based on the choice of the correct elution buffer. The reduction of non-specific interactions obtained from the on-column screening with tagged affinity systems, also contributed for the increase of the enrichment factor (Fig 5.2(C), reaching the maximum values calculated for each GFP-tagged (Table 4.2). In case of non-tagged systems, the enrichment maintained the same value between the screening formats, revealing once again the lack of specificity of the binding. However, as already described, the elution step can be optimized to improve selective recovery of GFP.

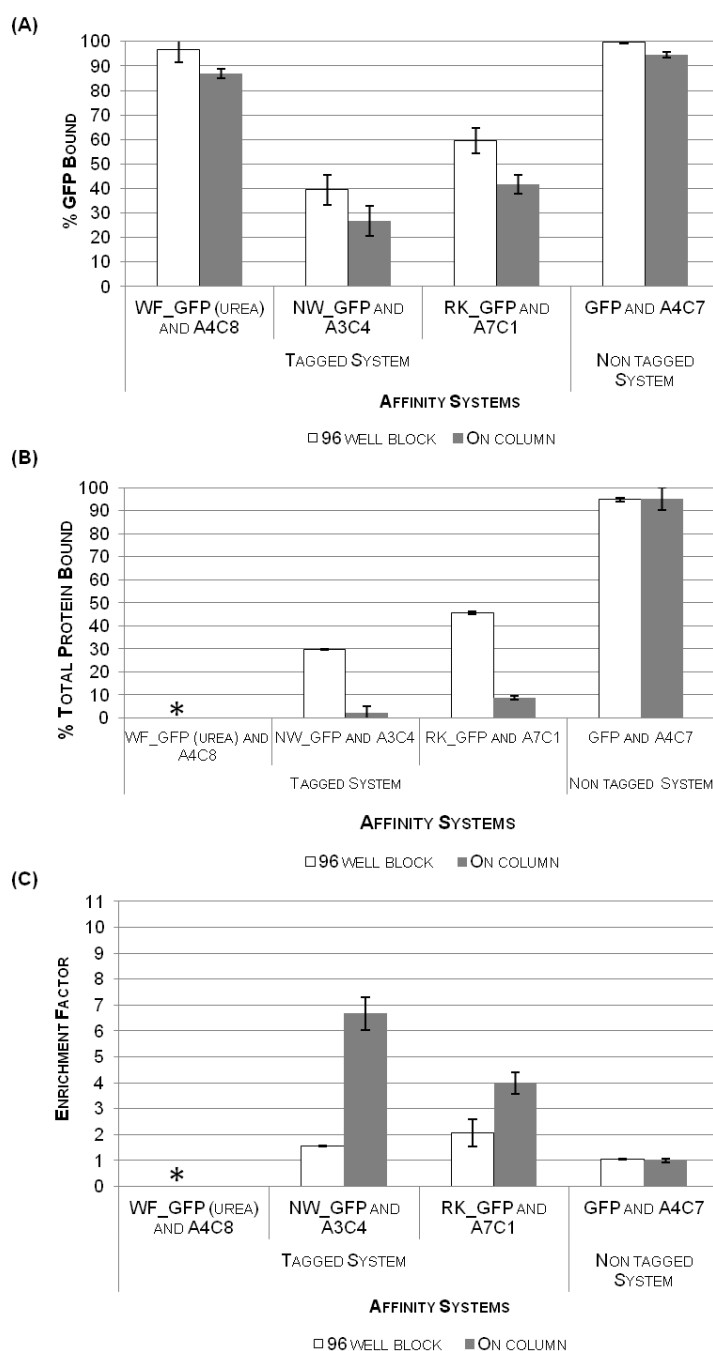


Figure 5.2 Comparison of the percentage of binding of (A) GFP protein and (B) total protein and (C) enrichment factor for the 96-well format and on-column assays for the lead ligands. * Total protein bound and enrichment were not determined due to the high concentrations of chaotropic agents employed which interfered with the BCA assay. The screening assays were carried out in duplicates (n=2) % Bound = (protein bound x 100)/ (protein loaded), where the protein bound was given by protein bound = (Protein loaded)-(sum of the amount protein in flow-through and washes). The amounts of GFP-tagged and non-tagged loaded on 96-well block were 3 ± 0.09 , 40 ± 0.10 , 28 ± 0.76 and 134 ± 4.17 μ g of WFWFWF (8M urea), NWNWNW, RKRKRK tagged GFP and GFP. The respective amounts of total protein were, except for WFWFWF system, 0.71 ± 0.03 , 0.91 ± 0.04 and 1.24 ± 0.01 mg. Regarding the amounts of GFP tagged and non-tagged loaded on-column was 10 ± 0.01 , 45 ± 0.45 , 33 ± 0.40 and 135 ± 3.6 of WFWFWF (8M urea), NWNWNW, and RKRKRK tagged GFP and GFP. Also the respective amounts of total protein were 0.87 ± 0.01 , 0.98 ± 0.01 and 1.08 ± 0.01 mg. In all the cases, the amounts loaded were approximately the same and therefore, the results are comparable in terms of binding percentage.

The comparison of the evaluation of the binding on a 96 well format and on-column was also extended to the controls. The controls refer to unmodified agarose (Fig 5.3). Screening on the 96-well format presents higher values of GFP bound (Fig 5.3A), which can be explained by the higher incubation times employed in the batch system enhancing non-specific interactions. In terms of total protein bound, the percentage of binding is similar and presents a maximum of binding of 15%, except for the hydrophilic tagged system RKRKRK-GFP. This can be explained non-specific interactions and ionic interactions that can be occurring between the contaminants and the GFP tagged in crude samples and the unmodified agarose.

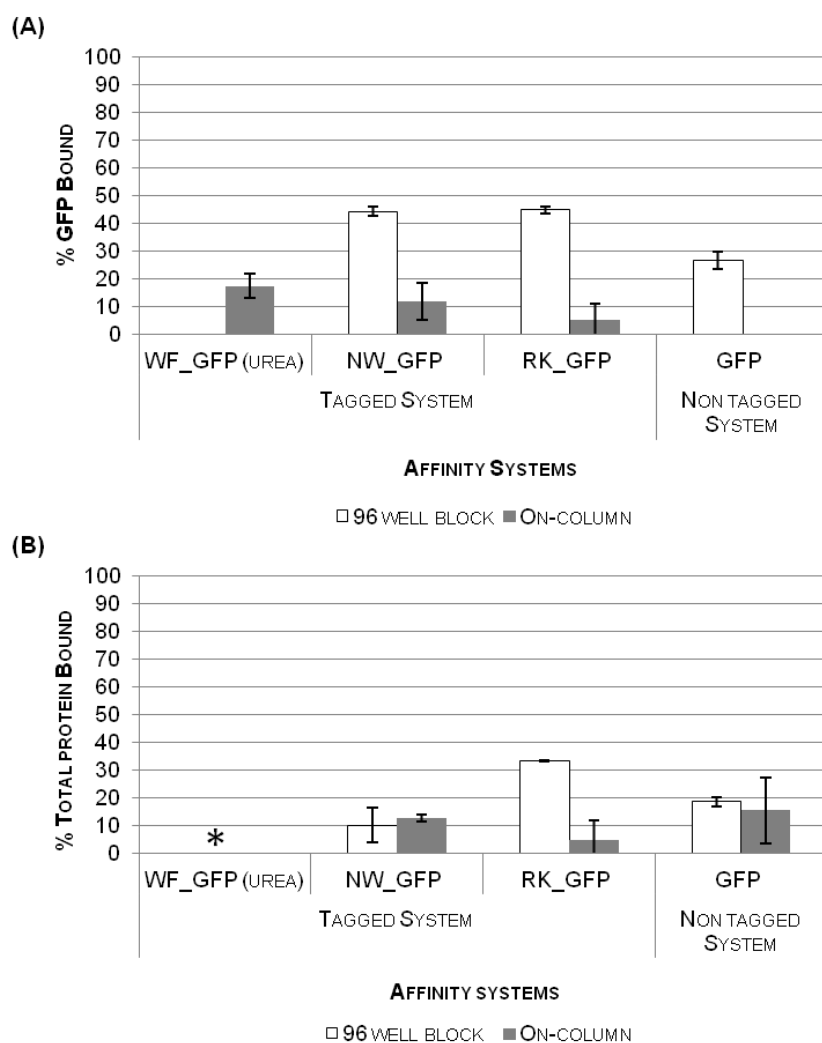


Figure 5.3- Comparison of the percentage of binding of (A) GFP protein and (B) total protein for the 96-well format and on-column for blank agarose. * Total protein bound and enrichment were not determined due to the high concentrations of chaotropic agents employed which interfered with the BCA assay. The screening assays were carried out in duplicates (n=2) % Bound = (protein bound x 100)/ (protein loaded), where the protein bound was given by protein bound = (Protein loaded)-(sum of the amount protein in flow-through and washes)

5.3 LEAD LIGANDS SELECTIVITY

The selectivity of the lead ligands for the NWNWNW and RKRKRK tagged GFP were evaluated by cross-screening ligand A3C4 against RKRKRK tagged GFP and GFP and ligand A7C1 against NWNWNW tagged GFP and GFP in on-column format. Ligand A4C7 revealed to bound ~100% to all GFP-tagged and non-tagged GFP proteins. The results obtained from these studies are shown in Figs. 5.4, 5.5. and 5.6.

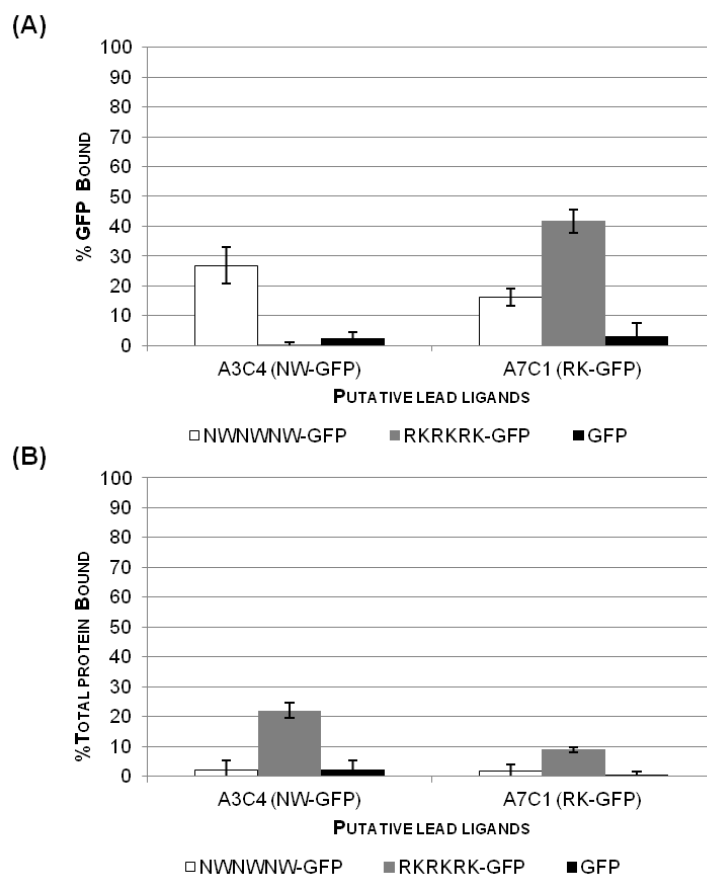


Figure 5.4 Selectivity of the ligands A3C4 and A7C1. The screening was performed on-column in duplicates (n=2) with the same strategy as in 96-well format. The amount loaded of GFP was 34.68 ± 0.71 , 15.78 ± 0.95 and 0.13 ± 0.00 μg for NWNWNW, RKRKRK tagged GFP and GFP, respectively. The respective amount of total protein was 0.61 ± 0.05 , 0.52 ± 0.04 and 0.95 ± 0.03 mg. % Bound = (protein bound x 100)/ (protein loaded), where the protein bound was given by protein bound = (Protein loaded)-(sum of the amount protein in flow-through and washes).

The results in Fig 5.4 (A) show that ligands A3C4 and A7C1 preferentially bound to their pairs indicating selectivity for the NWNWNW and RKRKRK tagged GFP, respectively. Regarding Fig. 5.4 (B), it is possible to observe a maximum 10% and 20% binding of total protein to ligands A7C1 and A3C4, respectively. These results were corroborated by SDS-PAGE analysis (Figs 5.5 and 5.6). In Fig. 5.5 (A) and (B), no selective binding of the RKRKRK tagged GFP and non-tagged GFP by ligand A3C4, and no evidence of protein eluted was observed.

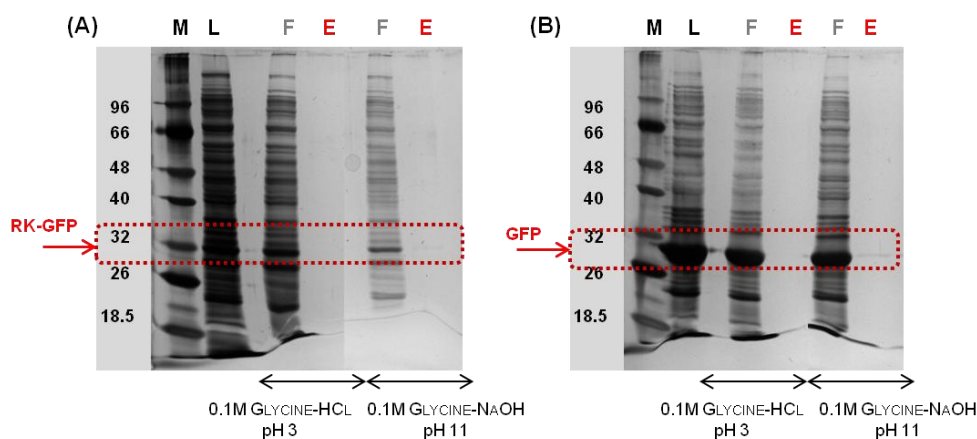


Figure 5.5 SDS-PAGE analysis of the binding A3C4 to (A) RKRKRK tagged GFP and (B) GFP. The Loading (L), flow-through (F) and elution fractions (E) obtained from on-column screening were loaded into SDS-PAGE gels (12.5% acrylamide), ran 1h at 120V and then silver stained. Two elution approaches 0.1M glycine-HCl pH 3 and 0.1M glycine-NaOH pH 11 were tested. (M=molecular weight marker).

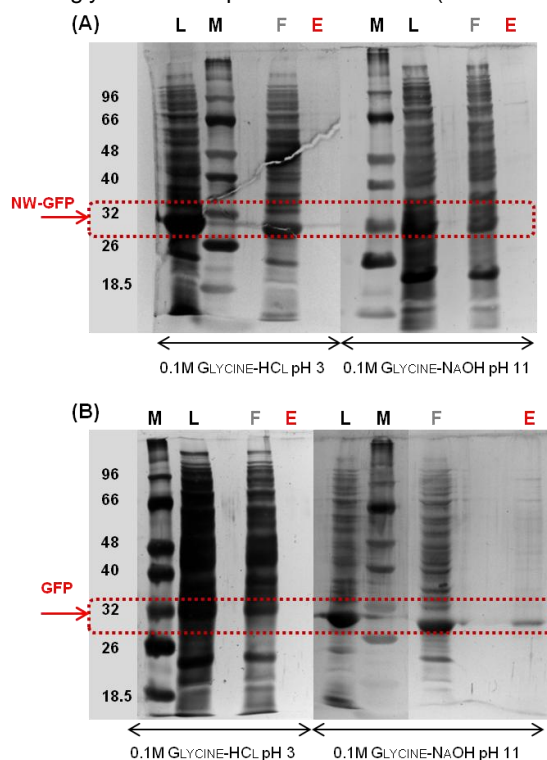


Figure 5.6. SDS-PAGE analysis of the binding A7C1 to (A) NWNWNW tagged GFP and (B) GFP. The Loading (L), flow-through (F) and elution fractions (E) obtained from on-column screening were loaded into SDS-PAGE gels (12.5% acrylamide), ran 1h at 120V and then silver stained. Two elution approaches 0.1M glycine-HCl pH 3 and 0.1M glycine-NaOH pH 11 were tested. (M=molecular weight marker).

A similar profile of low or inexistant selective binding of A7C1 was observed for NWNWNW tagged GFP and non-tagged GFP (Fig. 5.6 (A) and (B)). However, there was protein eluted under alkaline conditions that might be resultant from the non-specific binding of non-tagged GFP (<10%) to ligand A7C1. Overall, these results might infer about the selectivity of the lead ligands for their respective affinity tags.

5.4 DETERMINATION OF BINDING CONSTANTS

The adsorption step in affinity chromatography, if a unimolecular interaction occurs, involves the interaction between an immobilized ligand (L) and a target protein (P) to form a complex ligand-protein (LP). The following scheme (Fig. 5.7) and equation (5.1) describes the interaction, being the K_a , the association equilibrium constant and K_d the dissociation constant [34-35].

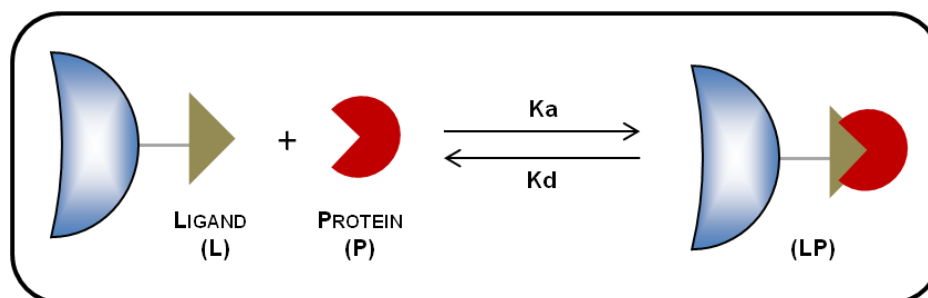


Figure 5.7. Adsorption and desorption scheme between an immobilized affinity ligand and a target protein.

$$K_a = \frac{[LP]}{[L] \times [P]} \quad \text{eq.5.1}$$

For affinity purification purposes, affinity constants should be within the range 10^3 M^{-1} - 10^9 M^{-1} [33, 249]. The association constant should not be low otherwise the separation process will be inefficient; nor too high because that will limit the recovery of the protein [33]. Examples of high affinity constants include natural affinity partners such as glutathione- glutathione-S-transferase (GST) protein [250], lectin (concanavalin A)- glycoprotein [251], staphylococcal protein A (spA) – IgG [77-78], avidin-biotin [113] with 10^5 , 10^5 , 10^8 and 10^{15} M^{-1} , respectively. When employing biomimetic ligands that mimic natural recognition, affinity constants are usually lower. Examples include triazine ligands 22/8 and 8/7 (artificial protein A and protein L, respectively), that bind to human IgG with a K_a of 10^5 and 10^4 M^{-1} , respectively [19-20]; Ligand 11/1 (glycoprotein-binding artificial receptor) which binds to glycoproteins purification with a K_a of 10^4 M^{-1} [22]. More recently, a new generation of biomimetic ligands based on the Ugi reaction have been developed for the purification of human IgG, displaying an affinity constant of 10^6 M^{-1} [38]. Another system with a high affinity constant, 10^7 M^{-1} , is the complex formed between metal ions and poly-histidine fusion proteins as determined by surface plasmon resonance (SPR) [181].

The detailed evaluation of the interaction between the ligand and the target protein can be determined by different analytical methods, separative and non-separative [252]. Separative methods involve liquid chromatography techniques such as quantitative affinity chromatography [252-253]. Non-separative methods rely on the evaluation of changes on the physicochemical properties of the target protein or the ligand after binding. These methods can be divided in different categories as spectroscopic (UV visible, fluorescence, NMR, circular dichroism),

calorimetric (isothermal titration calorimetry and differential scanning calorimetry) or surface plasma resonance (SPR).

Quantitative affinity chromatography presents several advantages: the matrix employed in the purification is also employed for the determination of the binding constants; it is a versatile technique that allows the characterization of a wide range of binding constants. It is possible to characterize from weak (10^3 M^{-1}) to strong interactions (10^9 M^{-1}) by partition equilibrium experiments, frontal or zonal affinity chromatography [249, 253]. In this particular case, the binding constants were determined by static partition equilibrium studies. These studies were performed by the incubation of a range of different concentrations of partitioning solute (GFP-tagged and non-tagged) with a fixed amount of immobilized ligand until the establishment of a chemical equilibrium [249, 253]. The correlation between the amount of protein adsorbed to the solid phase and the unbound protein after reaching the equilibrium is termed adsorption isotherm [34]. The isotherms are used to evaluate the binding capacity of the affinity adsorbent; Langmuir and Freundlich isotherms are typically employed in affinity chromatography [254]. However, the Langmuir isotherm is the most popular model and is often described by:

$$q = \frac{Q_{\max} \times K_a \times C}{1 + K_a \times C} \quad \text{eq. 5.2}$$

where q is the bound protein per mass of support (mg/g resin) and C corresponds to unbound protein in equilibrium (mg/ml). The Q_{\max} and K_a correspond to binding capacity and association equilibrium constant, respectively [34, 38, 254]. The assumptions of Langmuir isotherm include that all adsorption sites are identical and each one retains one target molecule; Moreover, all sites display uniform and independent energies of adsorption [254-255].

In order to evaluate the capacity of the affinity adsorbents and the association constants, partition equilibrium studies were performed for all the affinity systems on a 96-well format. The resultant parameters were fitted by a Langmuir isotherm and the respective constants are shown in Fig 5.8 and Table 5.1, respectively. For the systems with NWNWNW (Fig 5.8A) and RKRKRK tagged GFP (Fig. 5.8B), the isotherms displayed a linear profile, explained by the lower concentrations of protein used on these studies, which were not sufficient to reach the maximum adsorption capacity [256]. Regarding the studies with non-tagged GFP protein and A4C7 (Fig. 5.8C), the isotherm was linear for lower concentrations, stabilising after reaching a maximum binding capacity. The affinity constants obtained for the affinity systems tested were within the range for a suitable purification process, where the RKRKRK tagged GFP/ A7C1 and non-tagged GFP/A4C7 presented the same order of magnitude (10^5 M^{-1}) and the NWNWNW tagged GFP/A3C4 presented an affinity constant with a lower magnitude of 10^4 M^{-1} . The values of Q_{\max} presented a high error associated, but in general these are low when comparing if

other similar affinity systems. For example, the Ugi based affinity ligands for the purification of immunoglobulins presented Q_{max} values in the order of 25 and 16 mg IgGml⁻¹resin [38, 68].

These partition studies were carried out with unpurified protein from crude extracts and the higher concentration of GFP-tagged and non-tagged required was limited by the amount of protein produced in *E.coli* host. Also the determination of unbound protein was associated with a larger error. In order to overcome these issues, higher concentrations of protein were needed by using larger volumes of cell culture, concentrating the protein after the fractionation or by optimizing of the culture conditions to improve protein expression. On the other hand, a Langmuir isotherm was assumed as the best fitting model but other models might have been more appropriate.

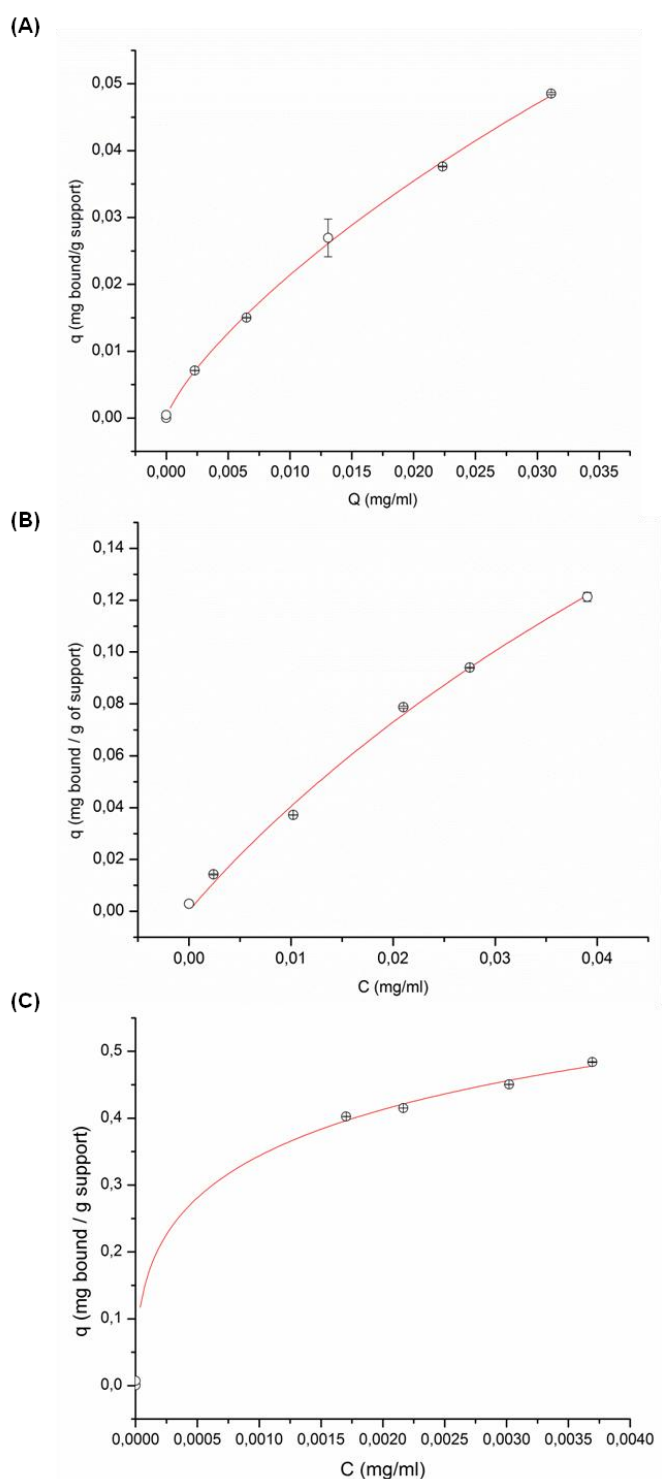


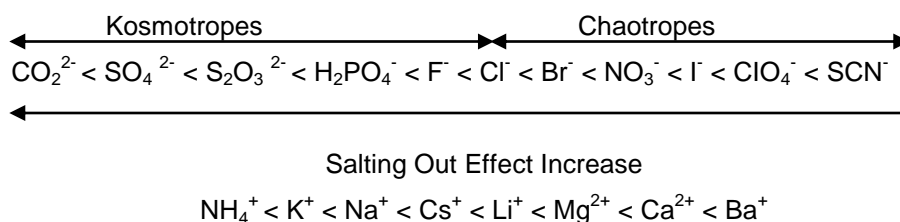
Figure 5.8- Binding isotherms and resulting static binding capacities for the affinity pairs (A) NWNWNW tagged GFP and A3C4, (B) RKRKRK tagged GFP and A7C1 and (C) GFP and A4C7. The Ugi-based adsorbents were incubated overnight at 4°C under orbital agitation with several dilutions of the correspondent protein. Bound and unbound protein was determined by measuring GFP fluorescence. Then the bound protein per mass of support, q , (mg/g resin) was determined and correlated with unbound protein in equilibrium, C , (mg/ml). Experimental data was plotted and fitted to the Langmuir model ($q = (Q_{MAX} \cdot K_A \cdot C) / (1 + K_A \cdot C)$). The fittings present correlation factors (r^2) of (A) 0.89, (B) 0.99 and (C) 0.99. The calculations were undertaken with OriginPro 8.0.

Table 5.1 Kinetic parameters obtained for each of the affinity pairs after fitting with Langmuir model.

Affinity System	K_a (M^{-1})	Q max (mg/g)	r^2
NWNWNW tagged GFP	7.22E+04	0.34 ± 0.58	0.99
RKRKRK tagged GFP	2.45E+05	0.47 ± 0.57	0.99
GFP	2.38E+05	1.02 ± 6.85	0.89

5.5 OPTIMIZATION OF THE ELUTION PROFILE FOR THE LEAD AFFINITY PAIRS

The role of the elution step in affinity purification is to disrupt the interactions formed between the complex ligand-target protein. In order to recover the target protein, a compromise between high yields, purity, stability and integrity of biological activity are taken into account when selecting elution buffers [35, 247, 257]. The elution process can be specific (i.e. competitive elution) or non-specific [34]. In the former, the elution is targeted to each particular case and mild conditions can be employed which contribute for the maintenance of the biological activity and integrity of the protein [34-35, 247]. However, when using the ligand as the competitive agent, the free ligand should be present in higher concentrations than the immobilized ligand [34], which is a disadvantage from an economic point of view [247]. The non-specific elution is related with a change in parameters as pH, ionic strength, polarity or temperature. A difference in pH can influence the overall charge of the ligand and the amino acids through protonation and deprotonation, and consequently can contribute for the disruption of the ionic interactions. However, some macromolecules can suffer from denaturation due to their instability under extreme pH conditions. Therefore, an additional concern about pH neutralization might be necessary [36, 247]. The use of different salt concentration and buffer salt composition can contribute for differences on the ion strength. High concentrations of salts can contribute for the disruption of ionic bonds but also can promote the formation of hydrophobic interactions [36, 247-248]. A class of salts that can be used are the chaotropic salts, which increase the solubility of non-polar molecules in water by disrupting or minimizing the hydrogen networking of the water molecules [258], disrupting hydrogen bonding and dissolving the hydrophobic binding regions into aqueous phase [36]. Both anions and cations of the chaotropic agents are arranged according with the Hofmeister series (Fig. 5.10), which is related with the ion's ability to alter the hydrogen bonding of water and is more pronounced for anions than for cations [36, 258-259].



The kosmotropes are a class of anions that are “water-structure makers” [259] as they possess the ability to perform hydrogen bonding with water molecules and contribute for the enhancement of hydrophobic interactions and salting-out effects. The chaotropic salts act by disordering water molecules, increasing the solubility of non-polar groups and reducing the hydrophobic interactions [259]. The urea, guanidine and detergents as sodium dodecyl sulphate (SDS) are also strong chaotropic agents [36, 258-259]. The greatest disadvantage of using chaotropic agents as eluents consists on the structural changes of the target protein, namely the loss of secondary and tertiary structure, leading to incorrect folding and denaturation, which require further removal after elution [34, 36, 247, 260]. The use of organic solvents also can affect the polarity of the affinity interaction, being the most common methanol, polyols as ethylene glycol (0-50%) and glycerol [247-248]. These solvents are polar reducing agents through the binding to hydrophobic sites of the target protein decreasing the hydrophobic interactions between the ligand and protein [261-262]. In general, ethylene glycol does not induce large conformational alterations in the protein, but high concentrations can cause ligand leaching [262]. The temperature and chelating substances are agents that also can improve the elution efficiency [34, 36]. The chelating agents such as EDTA are particular efficient in those affinity systems that require metals to enhance binding [36].

In order to develop a successful elution strategy, the eluents choice can be a combination of the parameters described above.

The optimization of the elution conditions was carried out after the binding assay, and the yield of the eluted fraction was monitored through GFP fluorescence. As already described, the GFP fluorophore is within a cylinder format with eleven beta strands around and a α -helix in both ends [244, 263], which confers protection to the GFP chromophore [206]. However the use of agents such as organic solvents and chaotropic agents can perturb the absorbance/excitation spectrum of native and recombinant wild-type *Aequorea* GFP [206]. The use of organic solvents, solvents with pH between 8 and 12, and solvents at high concentrations (> 60% (v/v)) can induce qualitative and quantitative spectral shift of GFP as previously noted [206]. The chaotropic agents when used in high concentrations as 8M urea and 4M guanidine ·HCl contribute for some loss of GFP fluorescence but when used it over longer periods and with higher temperatures a complete loss of fluorescence can be seen [206, 243].

The strategy for the optimization of elution conditions followed two different approaches in a 96-well format (Fig 5.9). As a first approach, the parameters analysed were the pH influence by using different buffers with pH between 3 and 10, different ionic strengths (0.15-1 M NaCl) and phosphate concentrations (0.01-1M) under mild conditions (PBS, 10mM sodium phosphate, 150mM NaCl pH 7.4). Afterwards, as second approach, the best pH that allows a better recovery of protein was used to evaluate the influence of different ionic strengths (0.15-1 M NaCl).and additives such as ethylene glycol (5-50% (v/v)). These last studies were conducted with two best pH conditions (pH 9 and 11) for non-tagged GFP, and for tagged systems just with the best pH condition (pH 11). Moreover, for the tagged systems, competitive elution also attempted, by using different concentrations of competitor such as tryptophan (5-15 mM) and arginine (10-750 mM) for the system NWWNWN and RKRKRK tagged GFP, respectively under mild conditions (PBS buffer) and at the best pH condition (pH 11).

The optimization of elution conditions was not addressed to the system WFWFWF tagged GFP because refolding on column should precede elution. The attempts on the refolding and elution were performed in an automated system and results are shown later in this Chapter (§ 5.7.1).

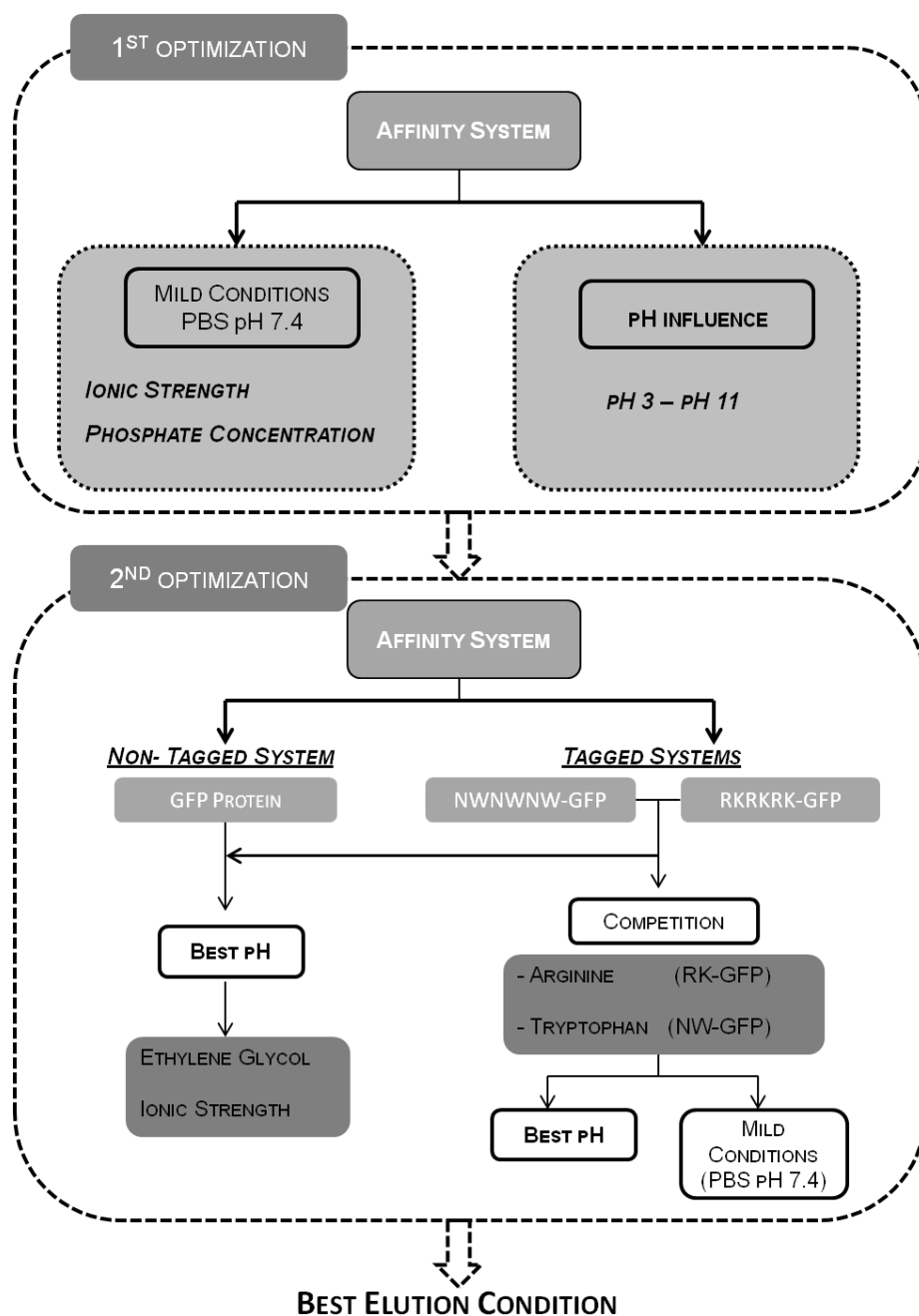


Figure 5.9- Strategy for the optimization of the elution conditions. Two different approaches were attempted in a 96-well format. The first attempt comprises the evaluation of best elution pH by using buffers with pH between 3 and 10, and the study of different ionic strengths (0.15-1 M NaCl) and phosphate concentration(0.01-1M) influence under mild conditions using PBS (10mM sodium phosphate, 150mM NaCl pH 7.4). The second approach involves the evaluation of GFP recovery with different ethylene glycol concentrations (5-50% (v/v)) and ionic strength (0.15-1 M NaCl) at the best pH condition, where a higher yield of GFP was eluted. In case of non-tagged GFP, these last studies were carried out with the two best pH elution conditions (pH 9 and 11) and for the tagged systems, only with one best pH condition (pH 11). In case of GFP tagged systems, the competitive elution was also considered by using arginine and tryptophan as the competitors for the RKRKRK-GFP and NWNWNW-GFP systems, respectively. The competitive elution was performed by using different concentrations of competitor (5-15 mM of tryptophan and 10-750 mM of arginine) at the best pH elution condition (pH 9) and also under mild conditions (PBS pH 7.4)

5.5.1 Affinity pair NWNWNW tagged GFP-A3C4

According to the chemical structure of the tag and ligand, it was expected that the basis of the affinity between the pair would be both hydrophobic interactions and hydrogen bonding (Fig. 5.10 and 5.11). The results obtained from the first approach of the elution optimization conditions with variation of phosphate concentration and ionic strength under mild conditions, indicated extremely low amounts of protein recovery (< 1%). When varying the pH, the same effect was also observed (< 1%), except for high alkaline conditions (pH > 11), at which the tag becomes negatively charged (Table 4.2). Overall, the system presented low reproducibility and recovery (< 3%).

The second attempt on the optimization of elution involved the use of elution buffer at pH 11 with different ionic strengths and concentrations of ethylene glycol. The competitive agent was tryptophan with a maximum concentration of 15 mM due to the low solubility in the buffers employed. Once again, the attempt to recover GFP tagged was not successful. The recovery of GFP tagged protein achieved a maximum of 3% in alkaline conditions with 50% ethylene glycol and 1% with all tryptophan concentrations at both conditions tested (PBS: 10 mM sodium phosphate, 150 mM pH 7.4 and 0.1M glycine-NaOH pH11). The influence of ethylene glycol in high concentrations seemed to recover a low amount of GFP-tagged but reproducibility was low. The competitive elution did not show any promising results, considering that only 1% of GFP tagged was eluted. In order to overcome this, more strongly agents should be tested to improve the efficiency of the elution. Alternatives can include chaotropic agents which disrupt most interactions (hydrogen-bonding, hydrophobic interactions and ionic bonds), and due to the harsh conditions employed, can present a higher impact on the elution of GFP tagged protein. Different competitors as an analogous compound of tryptophan should be used, with high solubility properties so that the easy dissolution on the mild conditions can be achieved. Considering all the conditions, the elution conditions chosen to proceed are 0.1 M Glycine-NaOH, pH 11 with 50%(v/v) of ethylene glycol.

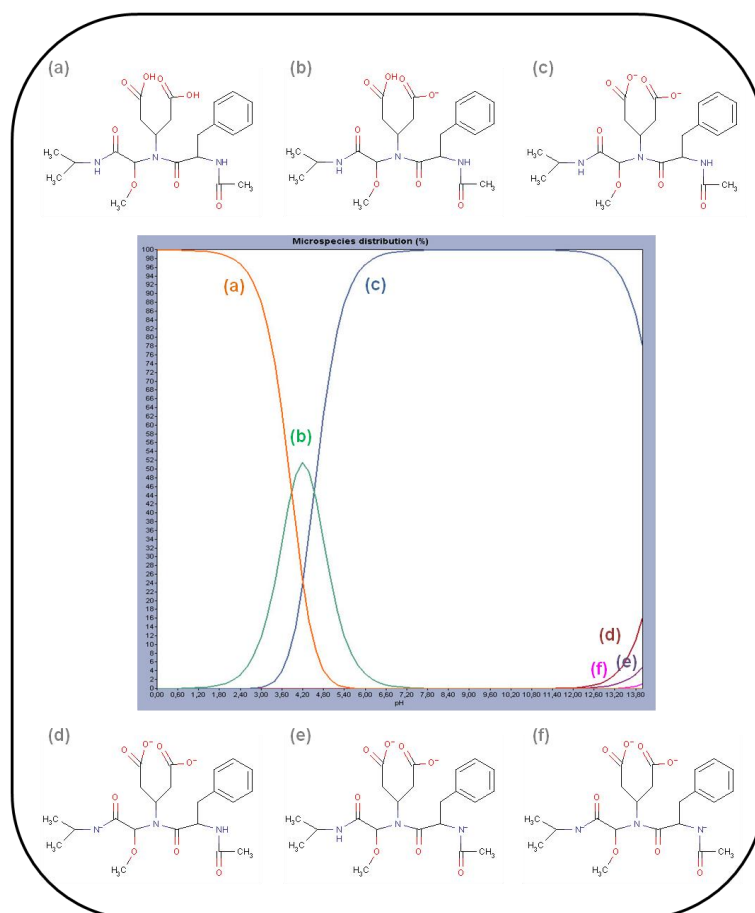


Figure 5.10 Microspecies distribution (%) for the ligand A3C4 over the pH values between 3 and 11. This distribution represents the various species that the ligand A3C4 can acquire over the pH values and was determined by the software Marvin Beans, ChemAxon.

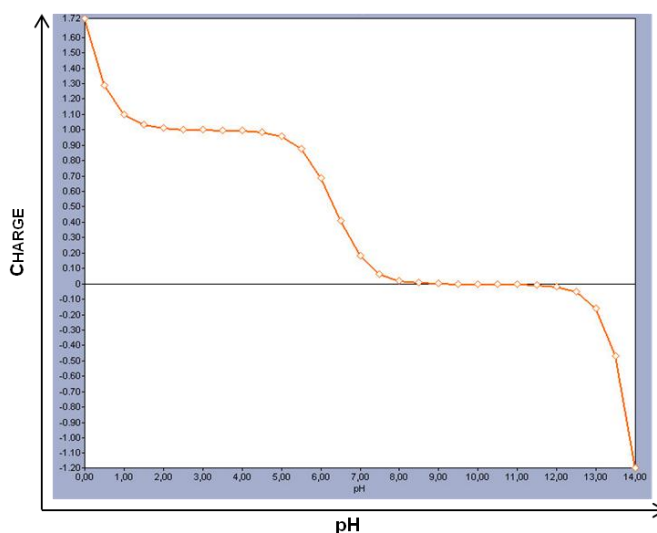


Figure 5.11 Influence of pH on the charge of the NWNWNW peptide without free C-terminal. This parameter was estimated by the software Marvin Beans, ChemAxon. This estimation took into account the C-terminal of the tag was linked to the recognition site of the enterokinase and GFP protein

5.5.2 Affinity pair RKRKRK tagged GFP - A7C1

The affinity system RKRKRK-GFP and A7C1 will probably interact mainly through electrostatic interactions and hydrogen bonding according to the structures of both affinity tag and receptor (Figs. 5.12 and 5.13). It was expected that an increase on ionic strength would disrupt the complex through the competition of the charged salts to the binding sites of the tag or the ligand. Variations on pH could also have a positive impact on the elution of the GFP tagged protein. The results obtained from the first elution tests are presented in Fig 5.14. Fig 5.14A shows the yield of fusion protein eluted for different phosphate and salt concentrations under mild conditions (PBS at pH 7.4). In terms of phosphate concentration, elution slightly decreases with the increase of salt concentration, being the best phosphate concentration 10 mM with a recovery of 20%. For the elution with varying concentrations of NaCl, the amount of fusion protein eluted increases from 150 to 250 mM NaCl, but there is not a significant difference from the concentration of 250 mM up to 1M of NaCl, corresponding to a recovery of 20%. High salt concentrations can break hydrogen bond networking and contribute for salting out effects of the protein, leading ultimately to protein precipitation and denaturation. Therefore, lower amounts of phosphate and NaCl in the elution buffers can be more beneficial, and a suitable buffer for mild elution is PBS with 10 mM of sodium phosphate and 250 mM NaCl at neutral pH. Regarding the effect of pH (Fig 5.14B), elution seemed to be more efficient at alkaline conditions. If the ionic interactions were predominant, the pH could influence the charge of the tag and the ligand by reducing or promoting the ionic forces. The ligand is negatively charged at pH values higher than 3 (Fig. 5.12), and the tag is highly positive charged until pH 8 where an accentuated decrease is observed until pH 13.35, where the charge is neutral (Fig. 5.13). Considering this, it was expected that from pH 8 onwards ionic interactions between the tag and the ligand would become weaker. The results obtained in Fig 5.14B corroborate these simulations, where pH 11 was the best condition for fusion protein elution.

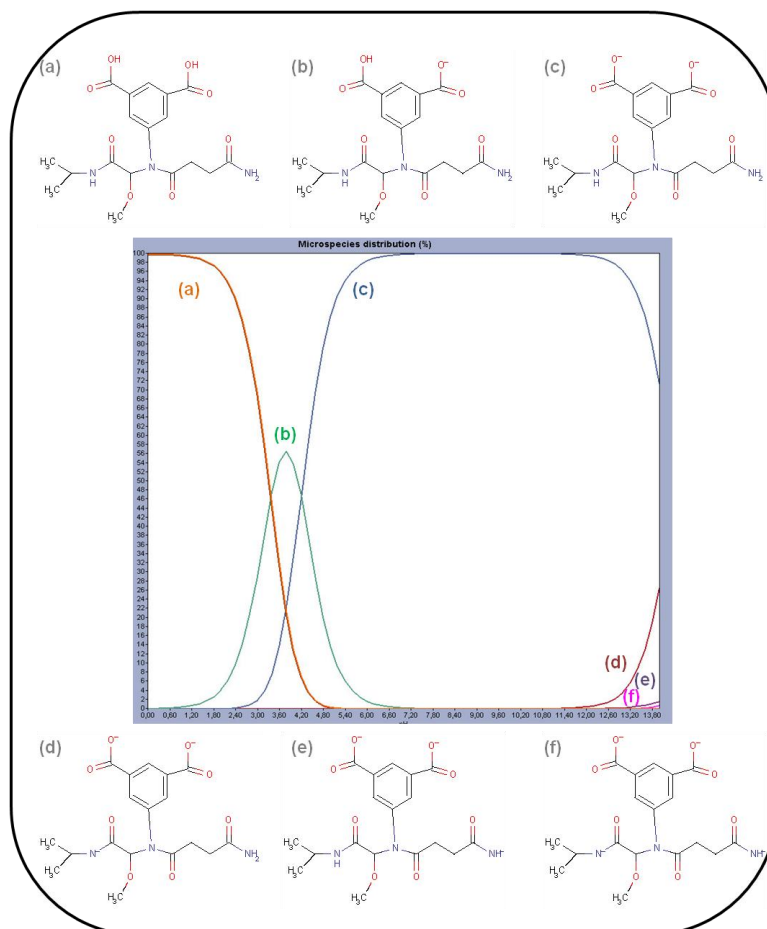


Figure 5.12 Microspecies distribution (%) for ligand A7C1 over pH values between 3 and 11 (determined by the software Marvin Beans, ChemAxon).

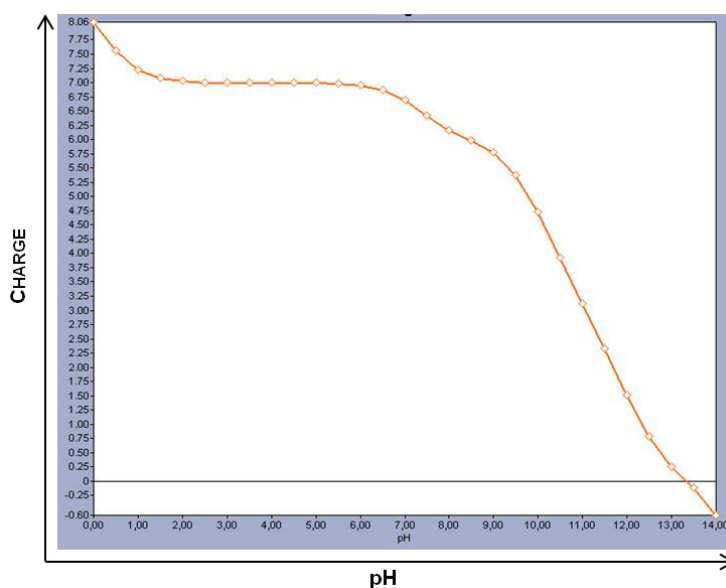
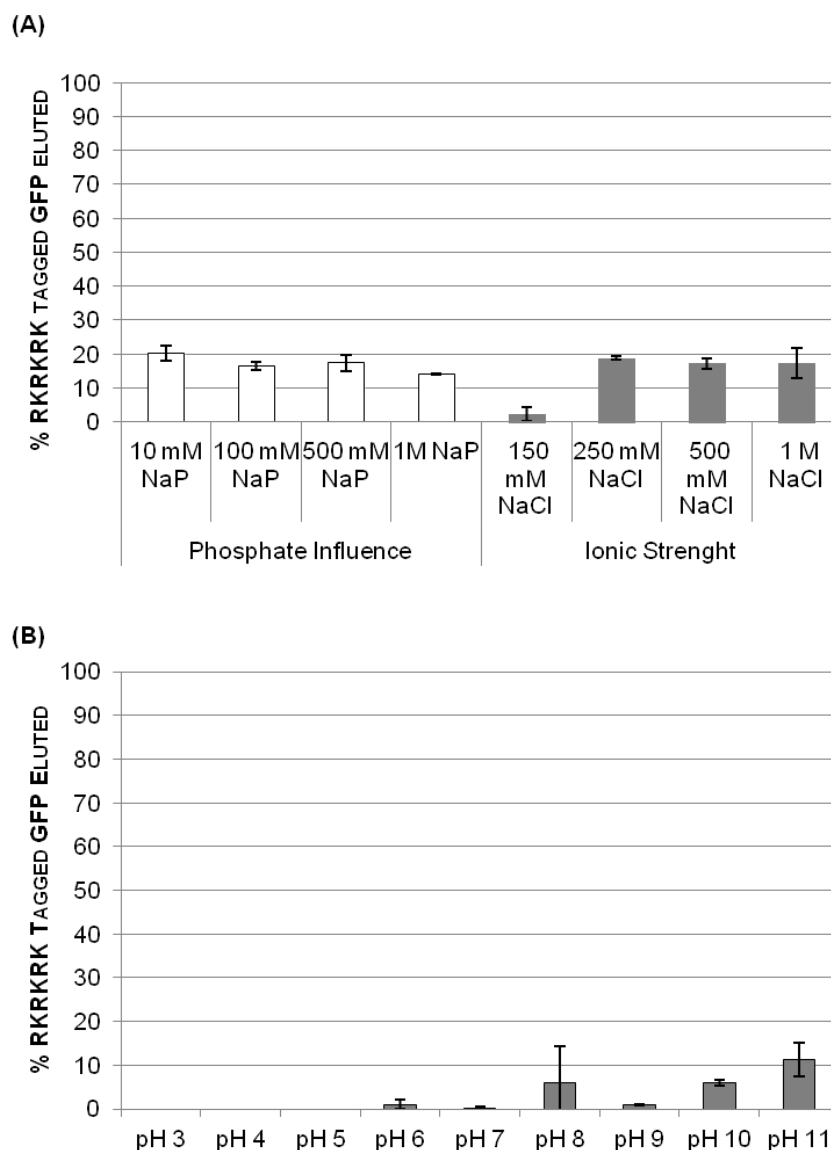


Figure 5.13 Influence of pH on the charge of RKRKRK peptide without free C-terminal estimated by the software Marvin Beans, ChemAxon. This estimation took into account the C-terminal of the tag being linked to the recognition site of the enterokinase and GFP.



5.14- Influence of the different elution buffers on the recovery of RKRKRK tagged GFP. As a first approach, the (A) influence of different phosphate concentrations (10mM-1M) and ionic strengths (150mM-1M NaCl) were studied under milder conditions with PBS (10 mM sodium phosphate , 150 mM NaCl, pH 7.4). The (B) influence of pH was also assessed by using different buffers with pH between the 3 and 10. The buffers used for pH 3-6, 7-8 and 9-10 were 0.1M citrate buffer, 0.1M sodium phosphate buffer and 0.1M glycine-HCl, respectively. The elution optimization was carried out in a 96-well format. The loading and washing steps were carried out as described in Fig. 5.21 with the additional elution step of addition of 5 c.v. to each well with the respective elution buffer after washing step. The total amount of RKRKRK-GFP bound was 5.7 µg, corresponding to a total of 60% of the amount loaded (9.5 µg). The error bars corresponds to duplicates. % Eluted = (protein eluted x 100) / protein bound.

The second approach on elution optimization was performed at pH11 in the presence of different concentrations of ethylene glycol and NaCl. Also competitive elution was performed by using arginine in PBS pH 7.4 and the buffer at pH 11. The maximum concentration of arginine used was 750mM due to solubilisation issues at higher concentrations (Fig 5.15). The

combination of the best pH and NaCl concentration improved the elution of the GFP tagged protein. The amount of GFP eluted appeared to be constant independently of the concentration of NaCl used (Fig.5.15 (A)), being the condition with the best GFP tagged recovery (30%) the elution buffer pH 11 with 150 mM NaCl. The effect of additives such as ethylene glycol was more accentuated with higher concentrations (Fig. 5.15 (A)).

The Fig 5.15 (B) shows the results about the competitive elution by using arginine as the competitor agent under mild conditions (PBS pH 7.4) and at pH 11. The arginine can compete with RKRKRK affinity tag for the ligand binding sites. Under milder conditions, the arginine concentrations that allow a better recovery of the protein are 500 and 750 mM, where under alkaline conditions, the results revealed similar amounts within concentrations between 100 and 750 mM. The results showed to be more promising at neutral conditions (PBS pH 7.4), and the arginine concentration selected was 500 mM with a protein recovery of 24%. Overall, the elution optimization studies contributed for an improvement of the elution being the most promising elution 0.1M Glycine-NaOH pH 11 with 150 mM of NaCl and the competitive elution condition as PBS pH 7.4 150 mM NaCl with 500 mM Arginine. The selected elution conditions can corroborate the selectivity between the affinity tag RKRKRK and ligand A7C1.

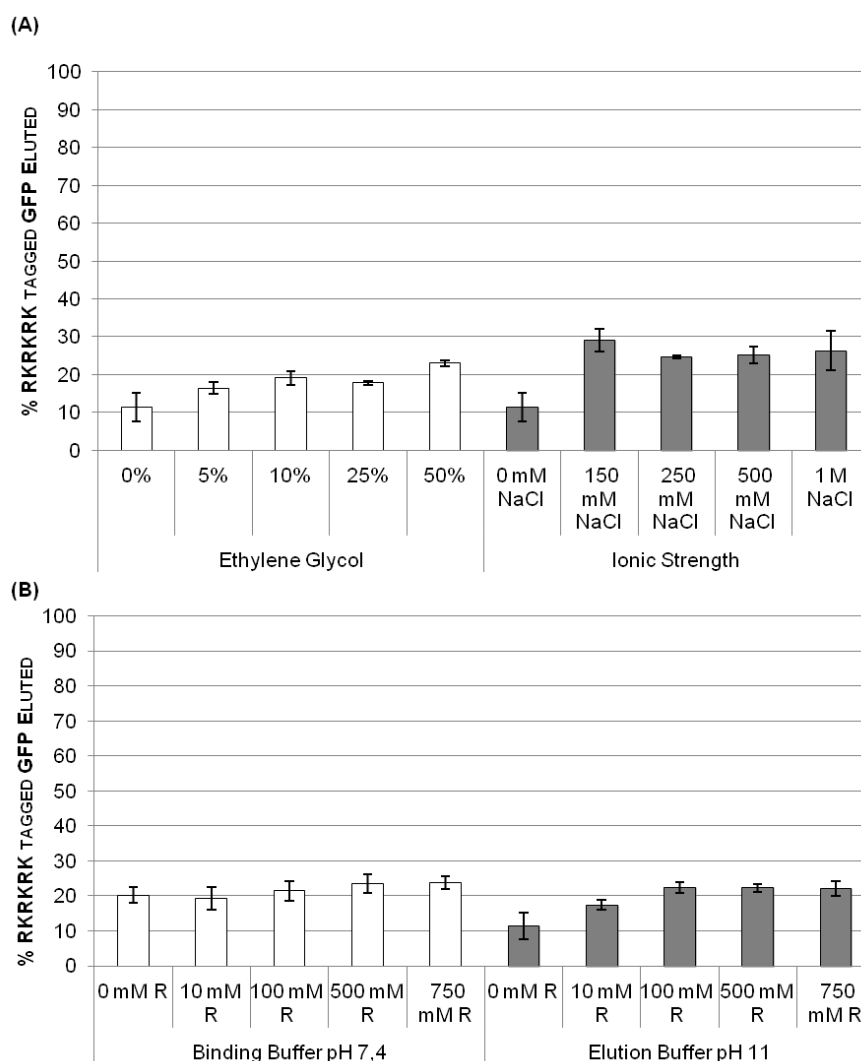


Figure 5.15- Second approach for the optimization of the elution conditions for the recovery of RKRKRK tagged GFP. The second approach for the elution optimization involves the evaluation of (A) different concentrations (v/v) of ethylene glycol (0-50%) and ionic strengths (0-1M NaCl) by using the best pH elution condition (0.1M Glycine-NaOH pH 11) determined on the first attempt of elution optimization. The competitive elution with arginine as competitor agent was evaluated by employing (B) different concentrations (0-750mM) under milder conditions (PBS, 150 mM NaCl, pH 7.4) and at pH 11 (0.1M Glycine-NaOH pH 11). The elution optimization was carried out in a 96-well format. The loading and washing steps were carried out as described in Fig. 5.21 with the additional elution step of addition of 5 c.v. to each well with the respective elution buffer after washing step. The total amount of RKRKRK-GFP bound was 20 µg, corresponding to a total of 60% of the amount loaded (32 µg). The error bars corresponds to duplicates. % Eluted = (protein eluted x 100) / protein bound.

5.5.3 Affinity pair GFP-A4C7

The main interactions involved in the affinity complex GFP and A4C8 were evaluated according to the structure of the lead ligand and GFP properties (Fig. 5.16 and 5.17). The optimization of the elution conditions might also give some indications about the interactions that have been established. Ligand A4C7 shows a highly hydrophobic structure with aromatic rings, which can

engage into hydrophobic interactions and as an acceptor of hydrogen bonding. No protein was eluted when using mild conditions as neutral phosphate buffers with different phosphate concentrations and ionic strengths. These parameters enhance hydrophobic interactions, and if these are the basis of the molecular recognition, the use of high concentrations of NaCl and sodium phosphate would contribute for a stronger interaction rather than disruption. When studying the influence of pH, it was observed that at pH 9, there was a recovery of 40% of bound GFP protein (Fig.5.15). At this pH value the overall charge of the protein is negative. The side chains of the amino acid present different pKa's of the ionisable groups. The pKa values can be shifted due to the influence of different factors as protein conformation and the polarizability of the solvent and protein [264]. Moreover, the binding of the ligand to the target protein can also affect the pKa values of the ionisable groups due to the transfer of the protons [265]. However if a shift on the pKa of the ionisable groups of the side chains of specific amino acids such as cysteine (pKa ~ 8), lysine (pKa ~ 10), tyrosine (pKa ~ 10), and arginine (pKa ~ 12) occurred, these groups could be deprotonated at pH 9. If these amino acids are interacting with the ligand via hydrogen bonding, at pH 9 these interactions should be reduced and the elution of the protein could be achieved. The protein can be interacting with ligand via hydrogen bonding considering that over the pH, the ligand presented a neutral charge eliminating the possibility of the establishment of the ionic interactions. However, in order to corroborate this hypothesis, it was necessary to understand if these amino acids are present in the surface of the GFP protein. The GFP sequence used on this work was based on the commercial sequence available in the plasmid pGFPuv (Clontech). The crystal structure used was the 1QYO from protein data bank (Fig.5.18). This structure is the one that presents a higher identity (more than 90%) with the GFP sequence used experimentally. By using the software Pymol, all the aminoacids whose pKa's of the side chains are close to 9 (lysine, tyrosine, arginine and cysteine) were selected to understand their position within the protein. All the aromatic residues were also selected in order to understand if they are available on surface to perform hydrophobic interactions.

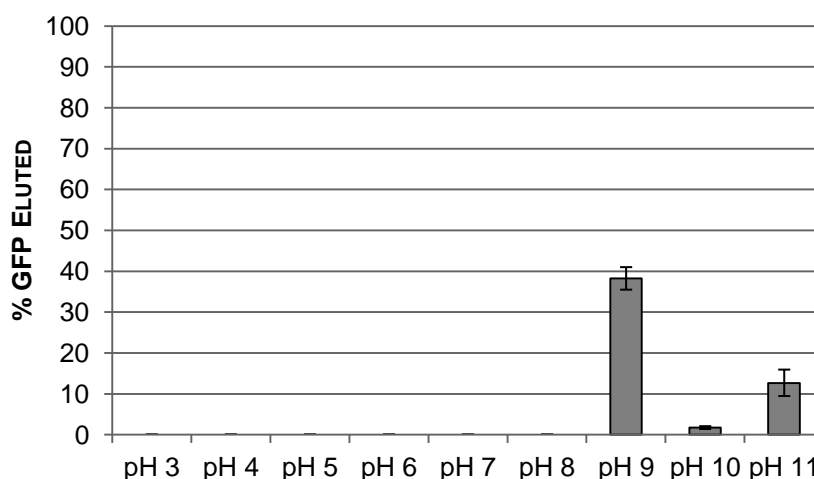


Figure 5.16- Influence of the different elution buffers on the recovery of GFP. As a first approach, the influence of pH was also assessed by using different buffers with pH between the 3 and 10 (B). The buffers used for pH 3-6, 7-8 and 9-10 were 0.1M citrate buffer, 0.1M sodium phosphate buffer and 0.1M glycine-HCl, respectively. The elution optimization was carried out in a 96-well format. The loading and washing steps were carried out as described in Fig. 5.27 with the additional elution step of addition of 5 c.v. to each well with the respective elution buffer after washing step, except for the condition 0.1M Glycine-NaOH pH 9, where 25 c.v. were added until GFP concentration reaches 0 mg/ml. The total amount of GFP bound was 111 µg, corresponding to a total of 99% of the amount loaded (112 µg). The error bars corresponds to duplicates. % Eluted = (protein eluted x 100) / protein bound.

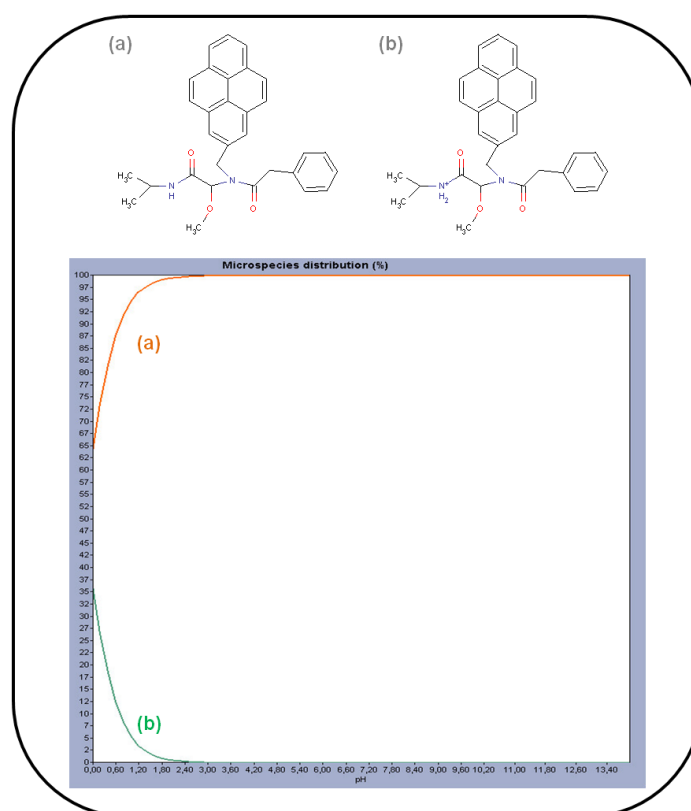


Figure 5.17- Microspecies distribution (%) for the ligand A4C7 over pH values between 3 and 11. This distribution represents the various species that the ligand A4C7 can acquire over the pH values and was determined by the software Marvin Beans, ChemAxon.

According to the Fig 5.18, a significant number of these residues are at the surface of the protein. Moreover, Fig 5.18 presents also a considerable number of aromatic residues (tyrosine, phenylalanine and tryptophan) mostly within the cylinder barrel (phenylalanine residues) and some present in higher amounts at the surface (tyrosine).

Overall, the best elution conditions found out from the first approach of the elution optimization were pHs 9 and 11. At pH 11, the total amount of GFP protein eluted is only 12%. Considering this, the second approach for the optimization of the elution buffers was performed and the results are shown in Fig 5.19.

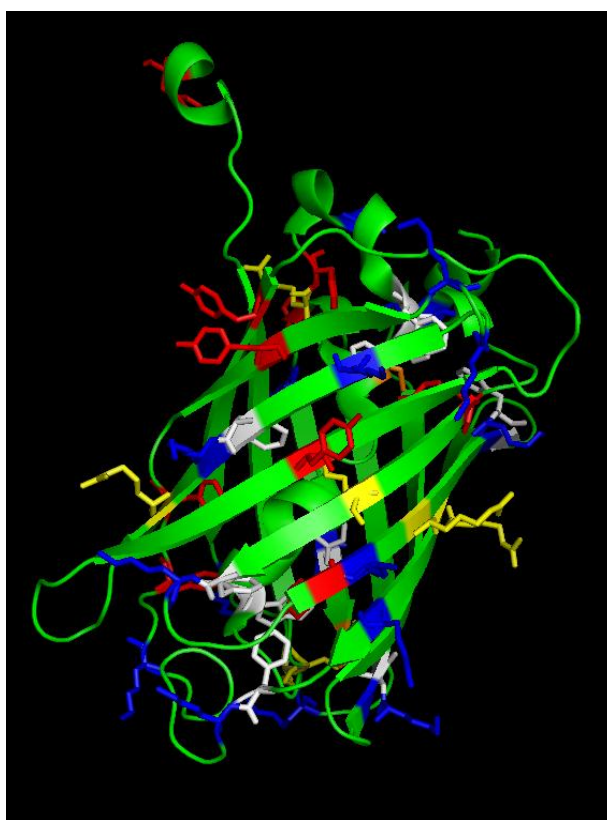


Figure 5.18- Crystal structure of GFP protein (1QYO). The surface of the GFP protein presents several aminoacids involved in different type of interactions such as lysines (blue), tyrosines (red), arginines (yellow), cysteines (orange) and aromatic residues (white) as tryptophan and phenylalanines.

The increase of ethylene glycol concentration increased the amount of GFP eluted, which corroborates the weight of hydrophobic interactions (Fig. 5.19(A)). As already discussed, ethylene glycol induces some perturbations in the fluorescence of GFP. This effect was observed as the percentage of GFP recovery corresponded to 150% when using pH 9 with 50% of ethylene glycol. In the case of pH 11 with 50% of ethylene glycol this corresponded to 99%. Even if the use of ethylene glycol inflated the results of the amount of GFP eluted, these two conditions seemed to be more promising. The combination of pH 9 and the use of 50% ethylene

contributed for the maximization of the disruption of more than one type of interactions. Other issue that was taken into account due to the higher percentages of GFP eluted, was ligand leaching. Ligand A4C7 presents a fluorescent amine with the same adsorption and emission wavelengths than GFP. Therefore, if ligand was leaching it could also be accounted in the quantification of the GFP protein. However, the SDS-PAGE analysis and the BCA assay confirmed the presence of protein during the elution.

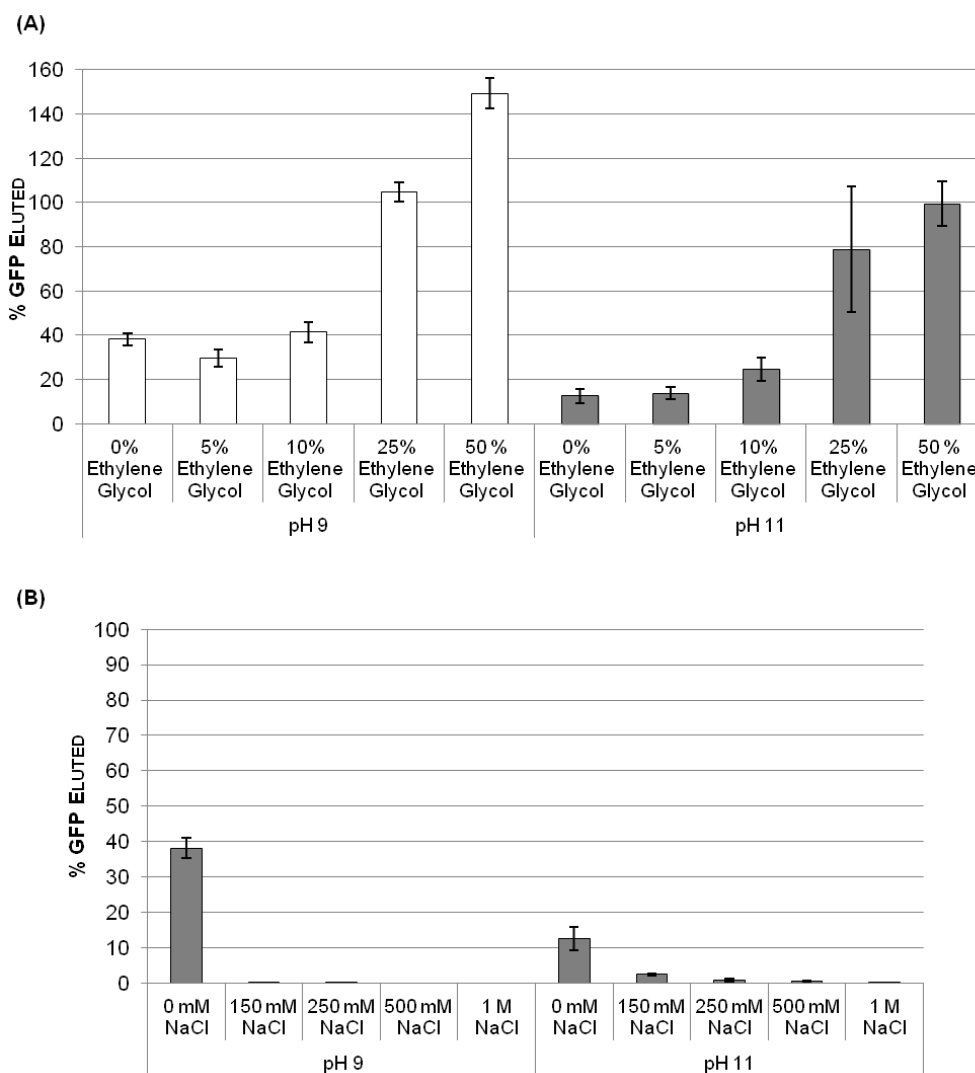


Figure 5.19 Second approach for the optimization of the elution conditions for the recovery of GFP
 The second approach for the elution optimization involves the evaluation of (A) different concentrations (v/v) of ethylene glycol (0-50%) and ionic strengths (0-1M NaCl) by using the two best pH elution conditions (0.1M Glycine-NaOH pH 9 and 11) determined on the first attempt of elution optimization. Also the influence of (B) of different ionic strength (0-1M NaCl) was evaluated in the two best pH elution conditions (0.1M Glycine-NaOH pH 9 and 11). The elution optimization was carried out in a 96-well format. The loading and washing steps were carried out as described in Fig. 5.27 with the additional elution step of addition of 5 c.v. to each well with the respective elution buffer after washing step, except for the conditions 0.1M Glycine-NaOH pH 9 with 5, 10, 25 and 50% ethylene glycol (v/v), where 18 c.v. were added until GFP concentration reaches 0 mg/ml. The total amount of GFP bound was 149 µg, corresponding to a total of 99% of the amount loaded (150 µg). The error bars corresponds to duplicates. % Eluted = (protein eluted x 100) / protein bound.

The results shown in Fig 5.19 revealed that different ionic strengths did not contribute for the disruption of interactions between the affinity complex. If the predominant interactions were involved were hydrophobic and hydrogen bonding, the results obtained can corroborate this hypothesis. Overall the best elution conditions selected to proceed were 0.1mM Glycine-NaOH pH 9 with and without 50% of ethylene glycol. These conditions might present selectivity on GFP recovery due to total amount of total protein also eluted that is below 13% as determined by BCA colorimetric analysis.

5.6 SCALE UP OF THE AFFINITY PAIRS

The scale-up of the affinity systems involved the use of higher amounts of the affinity adsorbents (~1g) in an automated preparative liquid chromatography system (AKTA™ avant 25). Therefore, it was possible to perform cycles of affinity purification at controlled flow rate, with 1ml/min as a first approach. The choice of this flow rate was the best approximation of the gravitational flow rate (0.75 ± 0.09 ml/min) obtained on-column screening, presented as the most suitable flow rate according to affinity based protein purification manuals [266] and was also the best flow compared to others (0.3 and 0.5 ml/min) to prevent resin damage. Two separate cycles of binding, elution and regeneration were performed for the affinity pairs RKRKRK tagged GFP/A7C1, NWNWNW tagged GFP/A3C4 and non-tagged GFP/A4C7 using *E.coli* protein extracts. The system GFP/A4C7 was also employed in the purification of a GFP fusion protein, produced in a different host (*Lactococcus lactis*). The affinity system WFWFWF-GFP/A4C8 was subjected to a cycle with an additional step of on column refolding after binding under denaturing conditions and prior to elution. Samples obtained in each step of the purification process were analysed by GFP fluorescence and total protein measurements as well as SDS-PAGE analysis for purity evaluation.

5.6.1 Affinity pair WFWFWF tagged GFP - A4C8 under denaturing conditions

The affinity pair WFWFWF tagged GFP/A4C8 presented different properties from the other systems, as the extremely hydrophobic tag induced the formation of inclusion bodies. In this case, as ligand A4C8 binds to the tagged protein in the presence of the chaotropic agent (urea), it was interesting to explore the potential of this system for on-column refolding of the protein prior to elution. The protocol for the refolding on column was based on other cases presented in the literature [49, 138, 190], and involved simultaneously a decreasing gradient of urea and an increasing gradient of the ionic strength up to 1M NaCl to avoid non-specific interactions [49]. During the refolding process (Fig. 5.20), one important aspect is the flexibility of the affinity tag to allow the proper protein structure folding while the interaction between the tag and the

immobilized ligand is maintained. Therefore, the refolding yield and rate are dependent on each target protein [165].

If the refolding is successfully achieved, the protein will acquire the correct folding and the fluorescence of chromophore can be observed. As the optimization studies (§ 5.6) were not performed for this affinity system due to impossibility to have an accurate quantification of GFP fluorescence and total protein, it was expected that, by removing urea, the fluorescence of WFWFWF tagged GFP could be slowly restored. As a first approach, the refolding on-column was conducted with 50 c.v. of PBS with 1M NaCl to dilute the 8M urea presented in the binding buffer prior to elution step. However, after the refolding process, the WFWFWF tagged GFP did not restore the proper folding due to inexistence of fluorescence obtained on the elution and regeneration samples. A longer refolding time with a gradient of PBS 8M urea: PBS 1M NaCl from 100:0 to 0:100 in 250 c.v. was used to improve the protein folding kinetic (Fig.5.20) and all the stages were monitored for GFP fluorescence. However, no GFP fluorescence was again observed. The SDS-PAGE gel of the samples collected confirmed that the affinity ligand bound to the GFP-tagged, when comparing the loading and flow-through lanes (Fig. 5.21). However, it is unlikely that GFP refolding occurred as despite the bands of GFP in the elution and regeneration lanes, no fluorescence was detected by spectrofluorimetry. During the first steps of refolding there was a small amount of tagged GFP being washed out. Still, as there was evidence of WFWFWF tagged GFP on the elution and regeneration lanes, there was still protein bound to the affinity matrix during refolding. This might be related with the fact that high salt concentrations can improve hydrophobic interactions between the ligand and the tag and therefore influence the renaturation process [165]. The elution step was conducted by a drastic change in the pH conditions. If a rational selection was performed, maybe the use of ethylene glycol that disrupts the hydrophobic interactions, could improve the recovery of GFP-tagged. The used strategy was just performed once as a first approach and therefore some efforts should be done in future so that a total refolding of the GFP could be achieved.

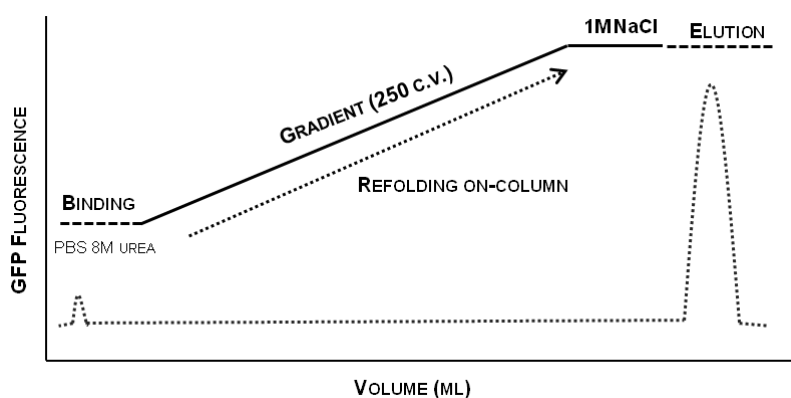


Figure 5.20 Strategy used for the refolding of the WFWFWF tagged GFP on A4C8 functionalized matrix performed on the AKTA™ avant 25 with the expected chromatogram. The refolding strategy was based on the dilution of the binding buffer PBS 8M urea with 250 c.v. of PBS with 1M NaCl, prior to the elution step with 0.1M glycine-NaOH pH 11.

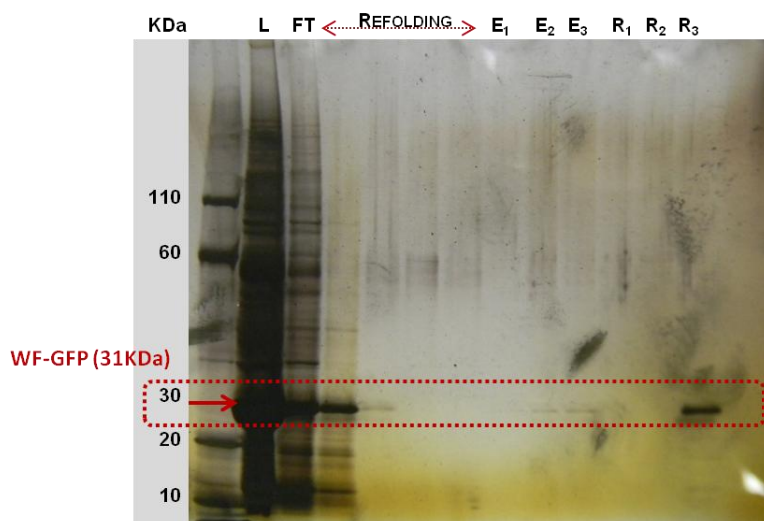


Figure 5.21- SDS-PAGE analysis of the samples from the purification of WF-FWF tagged GFP by ligand A4C8 under denaturing conditions. A4C8 adsorbent was packed (1g) on column with a flow of $0.3 \text{ ml}\cdot\text{min}^{-1}$ and prepared for analysis using an AKTA[™] avant 25. Afterwards the solid support was subjected to the following steps with a flow rate of $1 \text{ ml}\cdot\text{min}^{-1}$: regeneration (0.1M NaOH, 30% isopropanol (v/v) alternated with H₂O, 10 c.v. each in a total of 30 c.v.), equilibration (10 mM sodium phosphate, 150 mM NaCl, pH 7.4 15 c.v.) and loading (1 c.v. of 1ml of WF-FWF tagged GFP with 12 μg loaded/g of resin. In this particular case, after the loading, a refolding step was performed by using a gradient of 1M NaCl in binding buffer in order to dilute urea (250 c.v.). The elution was proceeded with 30 c.v. of the 0.1M Glycine-NaOH pH 11. The regeneration was done as already described with a total of 20 c.v. (0.1M NaOH, 30% isopropanol (v/v) (10 c.v.) alternated with H₂O (5 c.v.). All the fractions were collected from the loading in 1 ml 96-well block in all the steps except for the refolding step, where the samples were collected in 50ml centrifuge tubes. The GFP fluorescence was then evaluated. The ligand density corresponded to $\sim 20 \mu\text{mol}$ per gram of moist gel. The gel was 12.5% acrylamide, ran 1h at 120V and was silver stained.

5.6.2 Affinity pair NWNWNW tagged GFP-A3C4

The affinity pair NWNWNW tagged GFP and ligand A3C4 showed a poor performance in an automated system according with the results shown in Figs. 5.22, 5.23 and 5.24. The adsorption step revealed less than 10% of GFP-tagged binding, much lower from those obtained with 96-well format (Table 5.2). This was also corroborated by the SDS-PAGE analysis, where there was no evidence of GFP-tagged enrichment in the flowthrough (Fig.5.23). This can indicate that higher residence times and incubation in batch with orbital agitation were promoting the interaction between adsorbent and protein (Table 5.2)

Regarding the elution step, the conditions used were not effective on the recovery of the low amounts of NWNWNW-GFP bound to the affinity matrix. The conditions employed in the regeneration step, were also not strict enough to remove the protein bounded.

Overall, the affinity system NWNWNW-GFP/A3C4 was considered not suitable for affinity purification purposes, under the conditions employed in this study. The affinity constant of the pair was acceptable for purification processes; the estimated Q_{max} value was extremely low.

An effort on the binding and elution conditions for this affinity system can be an alternative to increase its performance. Otherwise, a better evaluation of the promising lead ligands for NWNWNW-GFP should be performed in order to find out other potential affinity ligand.

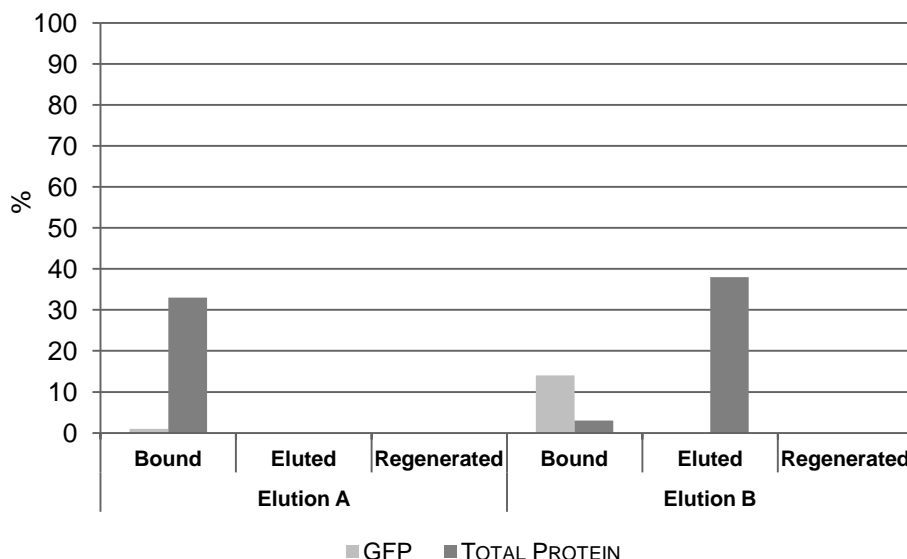


Figure 5.22- Percentages of NWNWNW tagged GFP bound, eluted and regenerated from affinity adsorbent containing A3C4. Two different elution conditions were tested (A) 0.1M Glycine-NaOH pH 11 and (B) 0.1M Glycine-NaOH pH 11 50% ethylene glycol. A3C4 adsorbent was packed (1g) on column with a flow of $0.3 \text{ ml} \cdot \text{min}^{-1}$ and prepared for analysis using an AKTA[™] avant 25. Afterwards the solid support was subjected to the following steps with a flow rate of $1 \text{ ml} \cdot \text{min}^{-1}$: regeneration (0.1M NaOH, 30% isopropanol (v/v) alternated with H₂O, 10 c.v. each in a total of 30 c.v.), equilibration (10 mM sodium phosphate, 150 mM NaCl, pH 7.4 15 c.v.) and loading (1 c.v. of 1ml of NWNWNW tagged GFP of (A) 20 μg and 5.5 mg and (B) 58 μg and 25 mg of GFP-tagged and total protein, respectively). Then the column was washed with binding buffer (20 c.v.), and then a step of elution with the respective buffer (10 c.v.) was performed and at the end other regeneration step was done as already described with a total of 20 c.v. (0.1M NaOH, 30% isopropanol (v/v) (10 c.v.) alternated with H₂O (5 c.v.)). All the fractions were collected from the loading in 1 ml 96-well block and were quantified by GFP fluorescence and Bradford assay. The ligand density corresponds to $\sim 20 \mu\text{mol}$ per gram of moist gel. The percentage of protein bound was determined according to $\% \text{ Bound} = (\text{protein bound} \times 100) / (\text{protein loaded})$, where the protein bound was given by $\text{protein bound} = (\text{Protein loaded}) - (\text{sum of the amount protein in flow-through and washes})$.

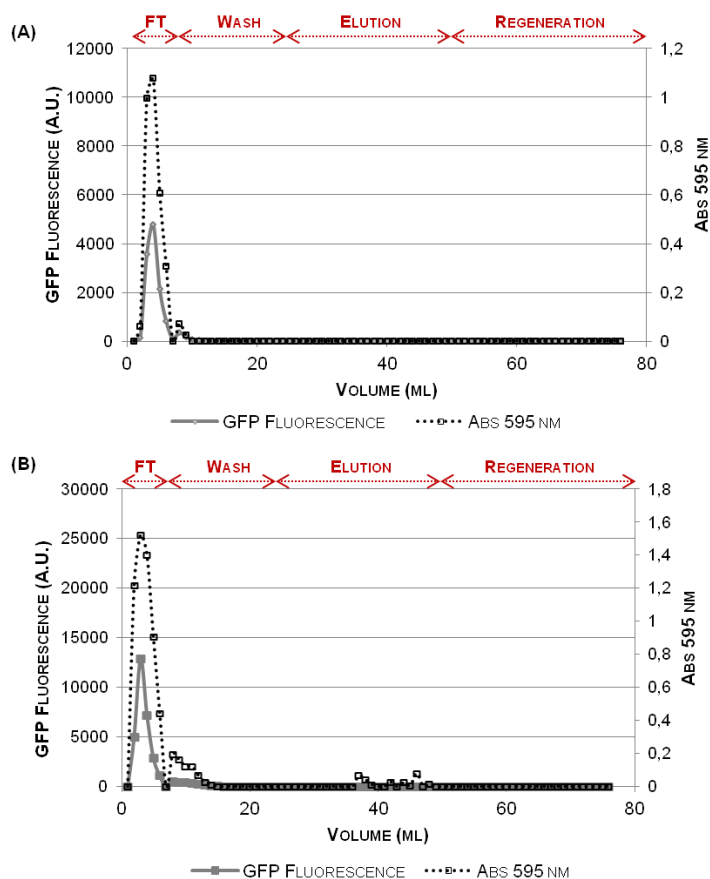


Figure 5.23- Chromatograms obtained for the purification of NWNWNW tagged GFP with affinity ligand A3C4 when using the elution conditions (A) 0.1M Glycine-NaOH pH 11 and (B) 0.1M Glycine-NaOH pH 11 50% ethylene glycol. The chromatograms were obtained from the purification of NWNWNW tagged GFP in Äkta avant 25 by using the conditions employed in Fig. 5.21, with the monitoring of NWNWNW tagged GFP through GFP fluorescence and the total protein by absorbance at 595 nm according to Bradford assay. FT- Flow-through

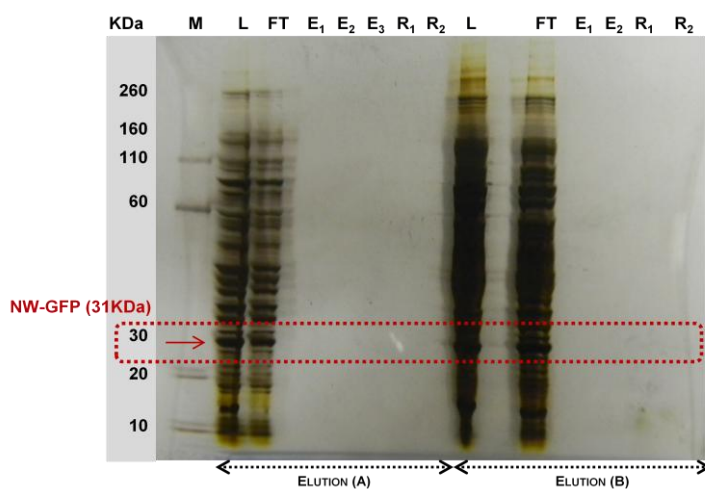


Figure 5.24 SDS-PAGE analysis of the samples from the purification of NWNWNW tagged GFP by the ligand A3C4 when using (A) 0.1M Glycine-NaOH pH 11 and (B) 0.1M Glycine-NaOH pH 11 50% ethylene glycol. The gel present 12.5% acrylamide ran 1h at 120V and silver stained. The purity of the eluted fractions was not estimated as there is no evidence of any amount of eluted protein.

Table 5.2. Comparison of the results obtained for the affinity pair NWNWNW tagged GFP/A3C4 in all chromatographic steps for the screening formats used: 96-well block, on-column and on a preparative liquid chromatographic system ÄKTA avant 25 system. The binding step was carried at physiological conditions (PBS pH 7.4 150 mM NaCl) and the elution conditions used where the two best conditions obtained by elution tests. The error bars correspond to duplicates, except for the binding on 96-well block that corresponds to an average of 16 samples obtained from the binding assays on the elution optimization studies. The individual runs performed on ÄKTA AVANT 25 were from different crude extracts containing different amounts of tagged GFP and total protein. Therefore the amounts on loading and binding corresponds to the conditions (A) 0.1M glycine-NaOH pH 11 and (B) 0.1M glycine-NaOH pH 11 50% ethylene glycol. * The binding of total protein was 0%, in the cases where GFP tagged was eluted, due to limitations on detection using colorimetric assay, BCA. ** The run was just conducted once.

CHROMATOGRAPHIC STEPS		96-WELL BLOCK	ON COLUMN	ÄKTA AVANT 25
RESIN AMOUNT (g)		0.25	0.25	1
Loading	NWNWNW tagged GFP (µG)	66.4 ± 7.2	45.25 ± 0.45	(A) 20 ** (B) 58 **
	Total Protein (mg)	1.9 ± 0.1	0.87 ± 0.01	(A) 5.55 ** (B) 25.00 **
Binding	NWNWNW tagged GFP (µG bound / g resin)	156.1 ± 12.5 (60%)	48.44 ± 11.01 (27%)	(A) 0.1 (1%) ** (B) 8 (14%) **
	Total Protein (mg bound / g resin)	0.6 ± 0.2 (38%)	0.07 ± 0.11 (2%)	(A) 2 (33%) ** (B) 0.8 (3%) **
Elution Condition (A) 0.1M Glycine-NaOH pH 11	NWNWNW tagged GFP (mg eluted / g bound)	31 ± 37 (1%)	-	0 (0%) **
	Total Protein (mg eluted / g bound)	0 (0%)*	-	0 (0%) **
Regeneration (Elution Condition a)	NWNWNW tagged GFP (mg regenerated / g bound)	-	-	0 (0%) **
	Total Protein (mg regenerated / g bound)	-	-	0 (0%) **
Elution Condition (B) 0.1M Glycine-NaOH pH 11 50% Ethylene glycol	NWNWNW tagged GFP (mg eluted / g bound)	31.9 ± 33 (3%)	-	0 (0%) **
	Total Protein (mg eluted / g bound)	354.1 ± 12.6 (35%)	-	365.6 (38%) **
Regeneration (Elution Condition B)	NWNWNW tagged GFP (mg regenerated/ g bound)	-	-	0 (0%) **
	Total Protein (mg regenerated/ g bound)	-	-	0 (0%) **

6.6.3 Affinity pair RKRKRK tagged GFP-A7C1

The results obtained for the affinity pair RKRKRK-GFP/A7C1 when employing a flow rate of 1ml/min are shown in Figs. 5.25, 5.26 and 5.27. Under the tested conditions, the adsorbent bound approximately 50% of loaded GFP (22.5 µg bound/g resin). The chromatograms obtained (Fig. 5.26) revealed a large peak in the flowthrough fraction for the total protein. The peak correspondent to RKRKRK-GFP is lower when compared with the maximum of fluorescence of the y-axis that is correlated with the fluorescence of the sample loaded. The fact that some amount of RKRKRK-GFP is being washed out can be related with limited binding capacity of the adsorbent at the flow rate and packing conditions tested. In terms of protein recovery, the two best elution buffers selected in section § 5.6 were tested. The best elution condition obtained in the scale-up studies was when using arginine as a competitor with 12% of GFP-tagged bound recovered with an estimated purity of 45%. The purity of RKRKRK-GFP eluted with condition B, as determined by densitometry analysis of SDS-PAGE gels in Fig 5.27, indicated a slightly higher value of 50%. It was also attempted to elute the protein with the PBS buffer with 500 mM lysine or 500 mM lysine: 500 mM arginine, but no protein was verified in the elution.

Regarding the regeneration step, it was not effective for the desorption of the remaining protein bound to the affinity matrix.

As a proof of concept, the affinity pair RKRKRK-GFP/A7C1 worked in an automated preparative liquid chromatography, but the results obtained on GFP binding and elution showed lower efficiency than 96-well block and on-column format. Moreover, the parameters used including flow rate were selected as a first approach and require further optimization to promote the efficiency on adsorption and desorption of GFP-tagged. The results obtained in the automated liquid chromatography differ from those obtained previously (Table 5.3).

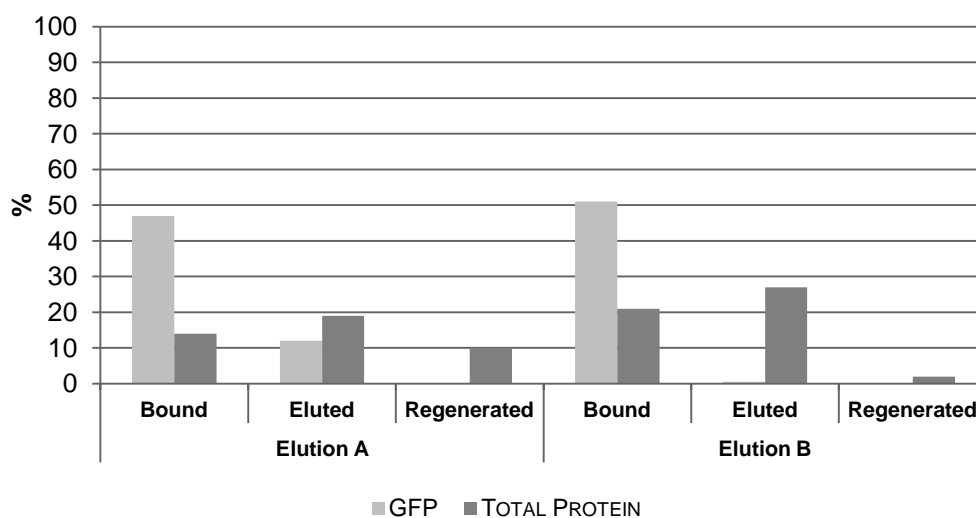


Figure 5.25- Percentages of RKRKRK tagged GFP bound, eluted and regenerated from affinity adsorbent containing A7C1. Two different elution conditions were tested (A) 10mM Naphosphate 150 mM NaCl pH 7.4 with 500mM Arginine and (B) 0.1M Naphosphate 150 mM NaCl pH 7.4. A7C1 adsorbent was packed (1g) on column ($0.3 \text{ ml}\cdot\text{min}^{-1}$) and prepared for analysis using AKTA[™] avant 25 system. Afterwards the solid support was subjected to the following steps at a flow rate of $1 \text{ ml}\cdot\text{min}^{-1}$: regeneration (0.1M NaOH, 30% isopropanol (v/v) alternated with H₂O, 10 c.v each in a total of 30 c.v.), equilibration (10 mM sodium phosphate, 150 mM NaCl, pH 7.4, 15 c.v) and loading (1 c.v of 1ml of RKRKRK tagged GFP: 44 μg of GFP-tagged and 12 mg of total protein). The column was washed with binding buffer (20 c.v.), and then a step of elution with the respective buffer (10 c.v.) was performed and at the end other regeneration step was done as already described with a total of 20 c.v. (0.1M NaOH, 30% isopropanol (v/v) (10 c.v.) alternated with H₂O (5 c.v.)). All the fractions were collected from the loading in 1 ml 96-well block and were quantified by GFP fluorescence and Bradford assay. The ligand density was $\sim 20 \mu\text{mol}$ per gram of moist gel. The percentage of protein bound was determined according to $\% \text{ Bound} = (\text{protein bound} \times 100) / (\text{protein loaded})$, where the protein bound was given by $\text{protein bound} = (\text{Protein loaded}) - (\text{sum of the amount protein in flow-through and washes})$.

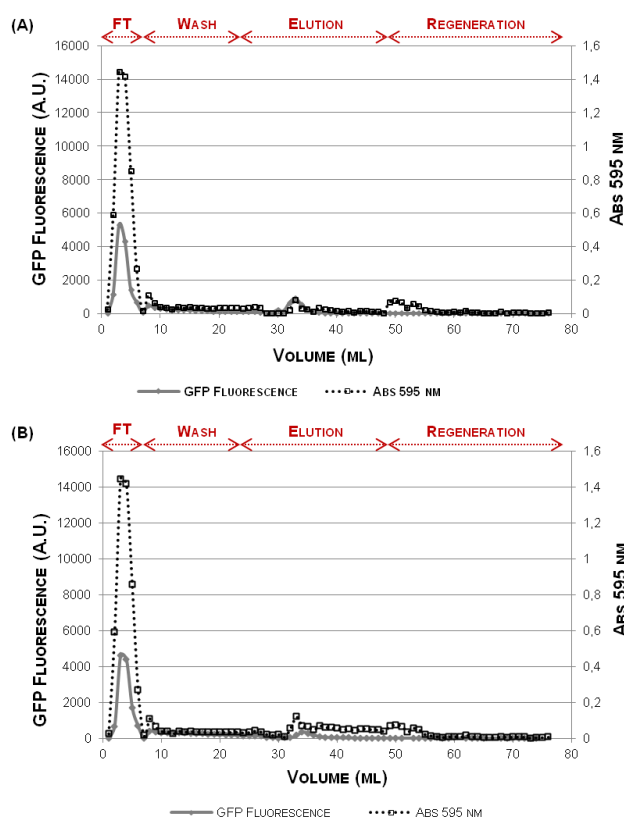


Figure 5.26 – Chromatograms obtained for the purification of RKRKRK tagged GFP with affinity ligand A7C1 when using the elution conditions (A) PBS pH 7.4 150 mM NaCl with 500mM Arginine and (B) 0.1M PBS pH 7.4 150 mM NaCl. The chromatograms were obtained from the purification of RKRKRK tagged GFP in Äkta avant 25 by using the conditions employed in Fig. 5.21, with the monitorization of NWNWNW tagged GFP through GFP fluorescence and the total protein by absorbance at 595 nm according to Bradford assay. FT- Flow-through

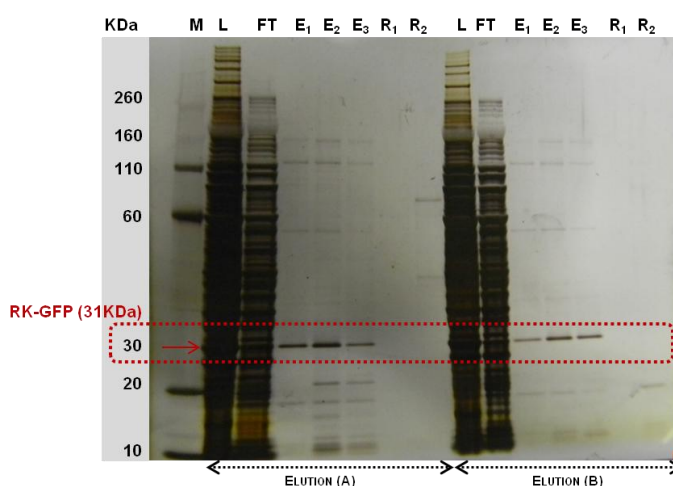


Figure 5.27 SDS-PAGE analysis of the samples from the purification of RKRKRK tagged GFP by the ligand A7C1 when using A) PBS pH 7.4 150 mM NaCl with 500mM Arginine and (B) 0.1M PBS pH 7.4 150 mM NaCl. The gel presents 12.5% acrylamide, ran 1h at 120V and silver stained. The purity of the eluted fractions was estimated with the software Image J revealing a purity of 45% for elution condition (A) and 50 % for elution condition (B).

Table 5.3 Comparison of the results obtained for the affinity pair RKRKRK tagged GFP/A7C1 in all chromatographic steps for the screening formats used: 96-well block, on-column and on a preparative liquid chromatographic system ÄKTA avant 25 system. The binding step was carried at physiological conditions (PBS pH 7.4 150 mM NaCl) and the elution conditions used where the two best conditions obtained by elution tests. The error bars correspond to duplicates, except for the binding on 96-well block that corresponds to an average of 16 samples obtained from the binding assays on the elution optimization studies. * The binding of total protein was 0%, in the cases where GFP tagged was eluted, due to limitations on detection using colorimetric assay, BCA. ** The run was just conducted once.

CHROMATOGRAPHIC STEPS		96-WELL BLOCK	ON COLUMN	ÄKTA AVANT 25
Resin amount (g)		0.25	0.25	1
Loading	RKRKRK tagged GFP (µG)	37.00 ± 1.11	33.00 ± 0.49	46
	Total Protein (mg)	1.55 ± 0.25	0.98 ± 0.01	12
Binding	RKRKRK tagged GFP (µG bound / g resin)	77.45 ± 4.07 (52%)	55.00 ± 5.20 (42%)	22.5 ± 0.71 (49%)
	Total Protein (mg bound / g resin)	0.49 ± 0.06 (46%)	0.35 ± 0.03 (9%)	2.10 ± 0.56 (18%)
Elution Condition (A) PBS with 500 mM Arginine	RKRKRK tagged GFP (mg eluted / g bound)	235 ± 25.75 (30%)	-	111 (12%) **
	Total Protein (mg eluted / g bound)	15.72 ± 5.46 (2%)	-	214.28 (19%) **
Regeneration (Elution Condition a)	RKRKRK tagged GFP (mg regenerated / g bound)	-	-	0 (0%) **
	Total Protein (mg regenerated / g bound)	-	-	80.95 (10%) **
Elution Condition (B) 0.1M Glycine-NaOH pH 11 150 mM NaCl	RKRKRK tagged GFP (mg eluted / g bound)	290.67 ± 30.63 (24%)	-	4 (0.5%) **
	Total Protein (mg eluted / g bound)	0 (0%) *	-	228.57 (27%) **
Regeneration (Elution Condition B)	RKRKRK tagged GFP (mg regenerated/ g bound)	-	-	0 (0%) **
	Total Protein (mg regenerated/ g bound)	-	-	22.85 (9%) **

5.6.4 Affinity pair non-tagged GFP- ligand A4C7

For the affinity pair GFP/A4C7, the results obtained with the automated system are in agreement with those obtained in the 96-well format and on-column (Table 5.4) (Figs. 5.28, 5.29 and 5.30). In both formats, the binding percentage was ~100%. The immobilized ligand seemed to bind strongly to GFP independently of the format used. Moreover, it was possible to recover almost the total amount of GFP bound, although results for elution B (pH 9 with 50% of ethylene glycol) might have some error due to the presence of ethylene glycol. Consequently the mass balances are affected due to the increase of fluorescence. In order to confirm that the GFP fluorescence was being influenced just by the 50% of ethylene glycol and not by ligand leaching, a column run was performed in the ÄKTA avant 25 with the ligand and 50 c.v of the 0.1M glycine-NaOH 50% ethylene glycol and then 1ml fractions were collected in a 96-well block. As the ligand and GFP absorbs and emits fluorescence at equal wavelengths, the fractions were quantified as GFP fluorescence. The results obtained revealed no fluorescence in all fractions, indicating that the ligand was not being leached from the column during the elution step.

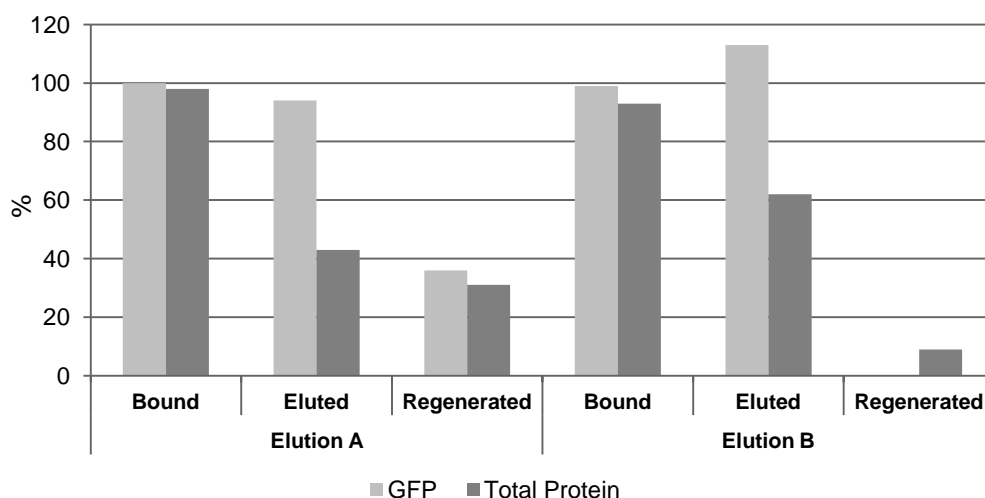


Figure 5.28- Percentages of GFP bound, eluted and regenerated from affinity adsorbent containing A4C7. Two different elution conditions were tested (A) 0.1M Glycine-NaOH pH 9 and (B) 0.1M Glycine-NaOH pH 9 50% ethylene glycol. A3C4 adsorbent was packed (1g) on column with a flow of $0.3 \text{ ml}\cdot\text{min}^{-1}$ and prepared for analysis using an AKTA[™] avant 25. Afterwards the solid support was subjected to the following steps with a flow rate of $1 \text{ ml}\cdot\text{min}^{-1}$: regeneration (0.1M NaOH, 30% isopropanol (v/v) alternated with H_2O , 10 c.v. each in a total of 30 c.v.), equilibration (10 mM sodium phosphate, 150 mM NaCl, pH 7.4 15 c.v.) and loading (1 c.v. of 1ml of GFP with 104 μg of GFP-tagged and 9 mg of total protein for both conditions). Then the column was washed with binding buffer (20 c.v.), and then a step of elution with the respective buffer (10 c.v.) was performed and at the end other regeneration step was done as already described with a total of 20 c.v. (0.1M NaOH, 30% isopropanol (v/v) (10 c.v.) alternated with H_2O (5 c.v.). All the fractions were collected from the loading in 1 ml 96-well block and were quantified by GFP fluorescence and Bradford assay. The ligand density corresponds to $\sim 20 \mu\text{mol}$ per gram of moist gel. The percentage of protein bound was determined according to $\% \text{ Bound} = (\text{protein bound} \times 100) / (\text{protein loaded})$, where the protein bound was given by $\text{protein bound} = (\text{Protein loaded}) - (\text{sum of the amount protein in flow-through and washes})$.

The flow rate used in the automated system seemed to be effective on binding, however the elution conditions used seemed to display different peaks on the GFP elution according with the chromatograms presented in Fig. 5.29.

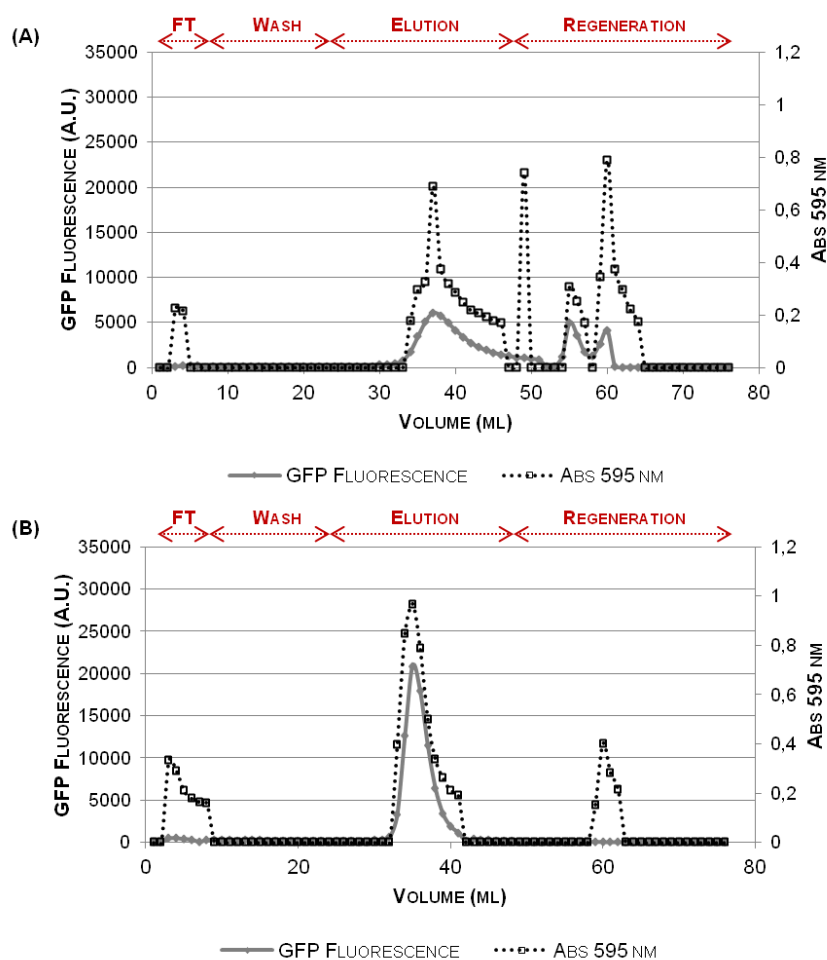


Figure 5.29 -Chromatograms obtained for the purification of GFP with affinity ligand A4C7 when using the elution conditions (A) 0.1M Glycine-NaOH pH 9 and (B) 0.1M Glycine-NaOH pH 9 50% ethylene glycol. The chromatograms were obtained from the purification of non-tagged GFP in Äkta avant 25 by using the conditions employed in Fig. 5.27, with the monitorization of GFP through GFP fluorescence and the total protein by absorbance at 595 nm according to Bradford assay. FT- Flow-through

By analysing both chromatograms, the elution on Fig. 5.29(A) shows a broad elution peak. On the subsequent regeneration step, there was a high amount of protein regenerated; indicating that elution A is not so effective to completely dissociate the affinity complex. The chromatogram on Fig. 5.28(B) shows a better profile on the elution, revealing a more accentuated peak, inferring about elution efficiency. On the regeneration step, there was no evidence of GFP only total protein presented on crude extract.

These results were also corroborated by SDS-PAGE analysis (Fig. 5.30), and the purity of GFP eluted was quantified by the software Image J. The purity of GFP eluted corresponded to 94% on elution (A) and 66% to elution (B), although more protein was recovered in the latter.

In this way, a compromise should be done between the amount and the higher yields, and therefore the elution condition A is preferably.

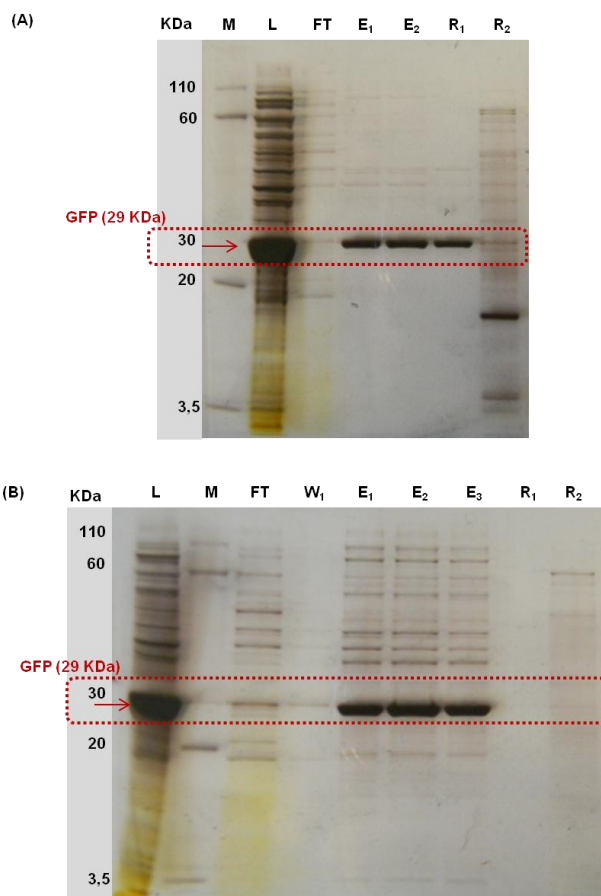


Figure 5.30- SDS-PAGE analysis obtained from the chromatographic assays from the purification of GFP and the ligand A4C7 when using (A) 0.1M Glycine-NaOH pH 9 and (B) 0.1M Glycine-NaOH pH 9 50% ethylene glycol. The gels present 12.5% acrylamide on its composition and were run 1h at 120V and then stained with silver. The purity of the eluted fractions were estimated with the software Image J revealing a purity of 94% for condition (A) and 67 % for condition (B).

Overall, these results indicate that the affinity ligand A4C7 although non specific for binding to GFP, allowed a significant amount of protein recovered with high yields of purity.

Table 5.4. Comparison of the results obtained for the affinity pair non-tagged GFP/ ligand A4C7 in all chromatographic steps for the screening formats used: 96-well block, on-column and on a preparative liquid chromatographic system ÄKTA avant 25 system. The binding step was carried at physiological conditions (PBS pH 7.4 150 mM NaCl) and the elution conditions used where the two best conditions obtained by elution tests. The error bars correspond to duplicates, except for the binding on 96-well block that corresponds to an average of 16 samples obtained from the binding assays on the elution optimization studies. ¹ The elution was higher than 100% due to the interference of 50% of ethylene glycol* The binding of total protein was 0%, in the cases where GFP tagged was eluted, due to limitations on detection using colorimetric assay, BCA. ** The run was just conducted once.

CHROMATOGRAPHIC STEPS		96-WELL BLOCK	ON COLUMN	ÄKTA AVANT 25
Resin amount (g)		0.25	0.25	1
Loading	GFP (µG)	148.30 ± 10.47	135.82 ± 3.60	104.00 ± 9.89
	Total Protein (mg)	2.81 ± 0.33	1.08 ± 0.01	9
Binding	GFP (µG bound / g resin)	590.72 ± 2.70 (99%)	512.00 ± 6.02 (94%)	103.50 ± 9.19 (99%)
	Total Protein (mg bound / g resin)	11.00 ± 0.07 (98%)	4.14 ± 0.21 (95%)	8.53 ± 0.33 (98%)
Elution Condition (A) 0.1M Glycine-NaOH pH 9	GFP (mg eluted / g bound)	382.24 ± 27.52 (40%)	-	888.00 (94%) **
	Total Protein (mg eluted / g bound)	16.62 ± 0.04 (2%)	-	445.22 (43%) **
Regeneration (Elution Condition a)	GFP (mg regenerated / g bound)	-	-	338.16 (36%) **
	Total Protein (mg regenerated / g bound)	-	-	316.34 (31%) **
Elution Condition (B) 0.1M Glycine-NaOH pH 9 50% Ethylene glycol	GFP (mg eluted / g bound)	1493.55 ± 68.00 (150%) ¹	-	1217.39 (113 %) ^{1**}
	Total Protein (mg eluted / g bound)	129.49 ± 9.66 (13%) *	-	651.43 (62%) **
Regeneration (Elution Condition B)	GFP (mg regenerated/ g bound)	-	-	0 (0%) **
	Total Protein (mg regenerated/ g bound)	-	-	1054.48 (9%) **

5.6.5 Purification of GFP fusion proteins with A4C7 ligand

In order to show the suitability of ligand A4C7 to purify GFP fusion proteins, GFP was fused to a *B. megaterium* spore cortex lytic enzyme SleL and then produced in *Lactococcus lactis* [267] with a total molecular weight of 78 KDa. This GFP fusion protein also presented a histidine tag after the GFP sequence. Besides using a real GFP fusion protein, this one was produced in a different host from *E.coli*. The results obtained in Fig.5.31 were promising; revealing highly percentages of binding ~90%, indicating that the affinity ligand still recognized GFP, even when fused to other protein. However, in terms of elution, condition (A), where only a displacement of pH was used, was not effective to recover the bound protein. The GFP-fusion protein is eluted only during the regeneration step. These results were corroborated by the chromatogram on Fig.5.32 and SDS-PAGE analysis on Fig. 5.33. The elution condition B seemed more promising with ~50% of recovery. This indicates that not only a displacement of pH was effective for desorption of the complex but the presence of an agent such as ethylene glycol disrupted hydrophobic interactions.

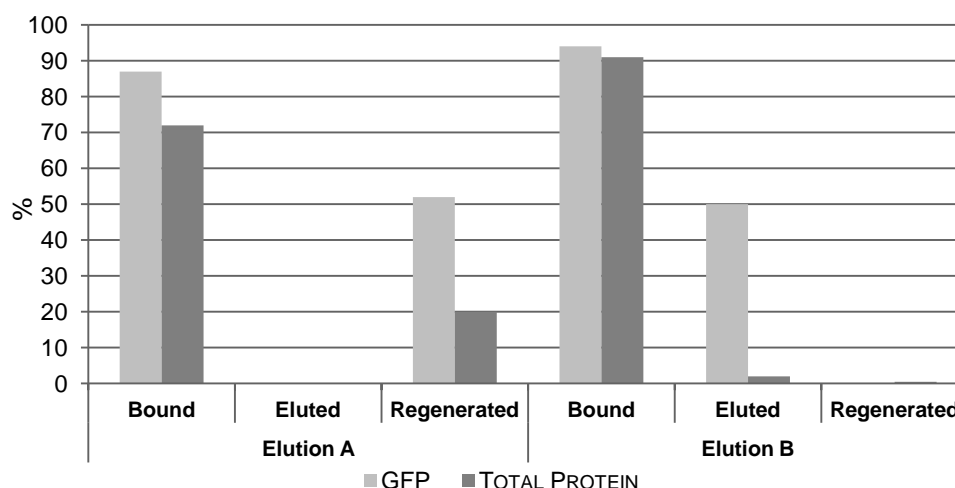


Figure 5.31- Percentages of GFP fusion protein bound, eluted and regenerated from affinity adsorbent containing A4C7. Two different conditions were tested (A) 0.1M Glycine-NaOH pH 9 and (B) 0.1M Glycine-NaOH pH 9 50% ethylene glycol. A4C7 adsorbent was packed (1g) on column with a flow of 0.3 ml.min⁻¹ and prepared for analysis using an AKTA[™] avant 25. Afterwards the solid support was subjected to the following steps with a flow rate of 1 ml.min⁻¹: regeneration (0.1M NaOH, 30% isopropanol (v/v) alternated with H₂O, 10 c.v. each in a total of 30 c.v.), equilibration (10 mM sodium phosphate, 150 mM NaCl, pH 7.4 15 c.v.) and loading (1 c.v. of 1ml of GFP with 13 µg of GFP and 9 mg of total protein in both situations. Then the column was washed with binding buffer (20 c.v.), and then a step of elution with the respective buffer (10 c.v.) was performed and at the end other regeneration step was done as already described with a total of 20 c.v. (0.1M NaOH, 30% isopropanol (v/v) (10 c.v.) alternated with H₂O (5 c.v.). All the fractions were collected from the loading in 1 ml 96-well block and were quantified by GFP fluorescence and Bradford assay. The ligand density corresponds to ~ 20 µmol per gram of moist gel. The percentage of protein bound was determined according to % Bound = (protein bound x 100)/ (protein loaded), where the protein bound was given by protein bound = (Protein loaded)-(sum of the amount protein in flow-through and washes).

The results were in accordance with the chromatogram from Fig. 5.33, that shows a peak of the GFP fusion protein on the elution. The respective results obtained on SDS-PAGE analysis for the eluted fractions were not as conclusive as just only a faint band appeared on the gel. Western blot analysis was performed with anti-GFP antibodies in order to detect the GFP-fusion proteins. The western blot was revealed by chemiluminescence reactions and the results are shown in Fig. 5.34 (B).

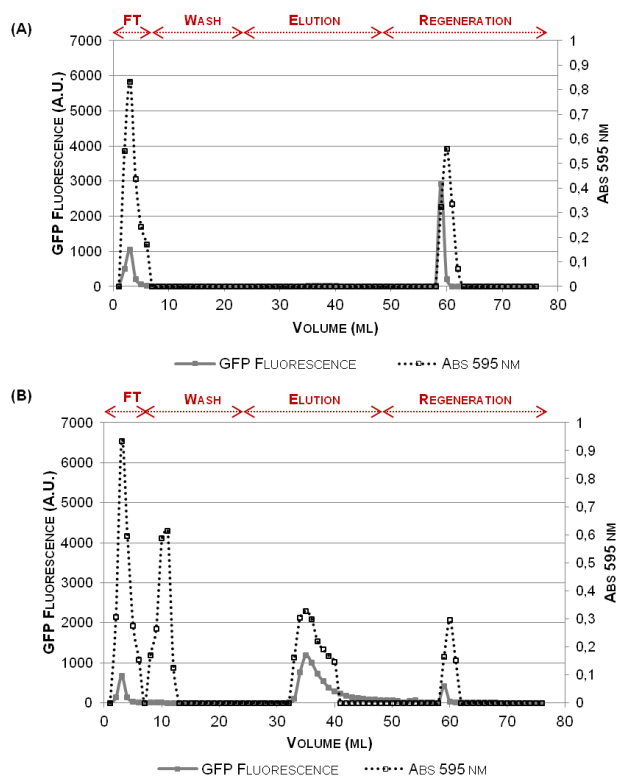


Figure 5.32 -Chromatograms obtained for the purification of GFP fusion protein with affinity ligand A4C7 when using the elution conditions (A) 0.1M Glycine-NaOH pH 9 and (B) 0.1M Glycine-NaOH pH 9 50% ethylene glycol. The chromatograms were obtained from the purification of RKRKRK tagged GFP in Äkta avant 25 by using the conditions employed in Fig. 5.27, with the monitorization of GFP fusion protein through GFP fluorescence and the total protein by absorbance at 595 nm according to Bradford assay. FT- Flow-through

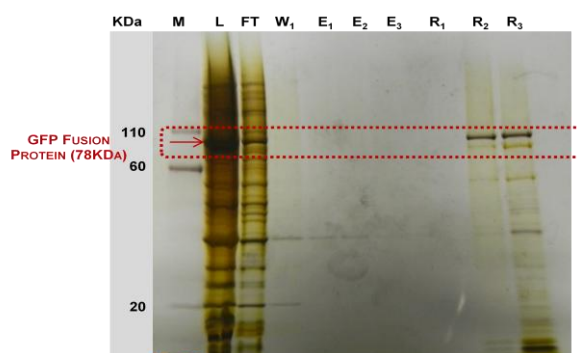


Figure 5.33-SDS-PAGE analysis of the samples from the purification of GFP fusion protein by the ligand A4C7 when using 0.1M Glycine-NaOH pH 9. The gel presents 12.5% acrylamide, ran 1h at 120V and silver stained. The purity of the eluted fractions was not assessed as there is no evidence of the elution of the GFP fusion protein.

The detection method of the western blot used presents the greatest sensitivity of any available detection method according to the supplier instructions. Due to this, it was possible to visualize the correspondent bands of the GFP-fusion protein on the elution and regeneration steps. The bands on the elution were not so evident on the SDS-PAGE analysis and therefore the purity evaluation was not assessed.

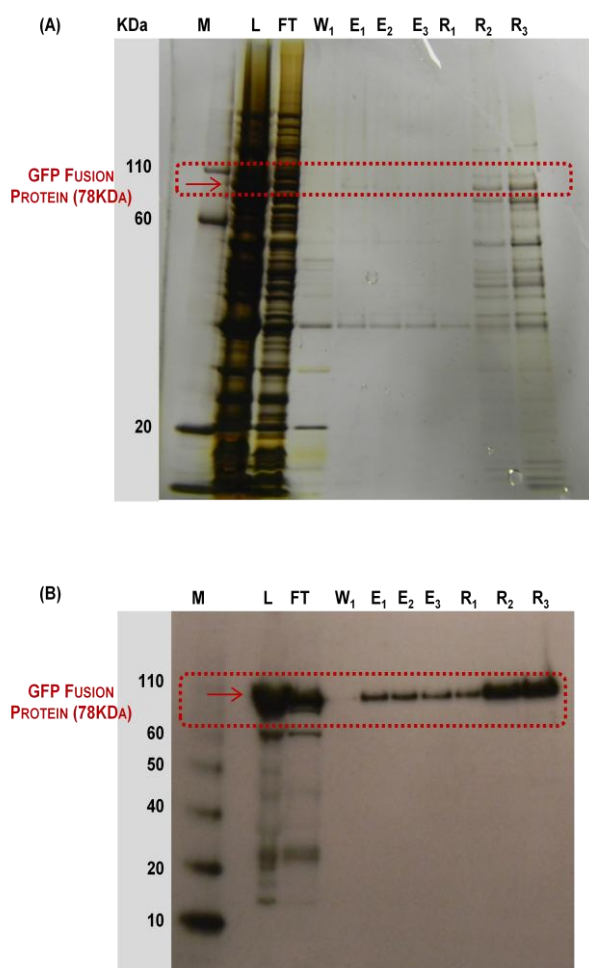


Figure 5.34 SDS-PAGE analysis obtained of the samples from the purification of GFP fusion protein by the ligand A4C7 when using 0.1M Glycine-NaOH pH 9 50% ethylene glycol, where (A) corresponds to SDS-PAGE analysis and (B) western blot. The gels present 12.5% acrylamide ran 1h at 120V. After, one gel was silver stained and the other was used for further western blot analysis. The western blot was performed by using anti-gfp antibodies that only recognized gfp.

These preliminary results indicate that the affinity ligand A4C7 might be effective on the purification of GFP fusion proteins, inferring about the selectivity for the GFP. In this particular case, the elution condition that indicates to be more effective was the buffer B, which includes high pH and 50% ethylene glycol. Moreover, this elution condition can also indicate the establishment of strong hydrophobic interactions between the pair. Overall, this affinity ligand seems to be a promising strategy for a novel purification method for GFP-fusion proteins.

5.7 CONCLUDING REMARKS

The optimization of chromatographic conditions on a 96-well block was effective on narrow to search of the most promising conditions for the binding and elution steps. This platform seems to be reliable, cost- and time-effectiveness. The results obtained from the 96-well block were comparable with those obtained by an automated liquid chromatography system the main difference being flow-rate and resin packing. This influences the performance of the adsorption and desorption processes between the affinity pair.

As a proof of concept, the development of an affinity pair “tag-receptor” was successfully achieved with the pair RKRKRK tagged GFP and ligand A7C1. The pair presented an affinity constant on order of 10^5 M^{-1} but the maximum binding capacity low. The binding and elution conditions used on an automated system still need further optimizations, mainly the flow-rate and residence time, in order to improve the adsorption and desorption steps. This affinity pair can be similar to an ion-exchange chromatography but with increased selectivity. The recovery of the RKRKRK tagged GFP was not possible with a pH displacement but when using arginine as a competitor.

Regarding the affinity pair NWNWNW tagged GFP and ligand A3C4, this pair revealed poor ability on the binding and elution of the target protein. Despite the tag the neutral character of the tag can facilitate the purification of challenging proteins, some studies should be performed in the future by using other putative lead ligands selected on Chapter 4.

The affinity ligand A4C8 bound preferentially to WFWFWFW tagged GFP in presence of urea., which is promising for proteins are expressed as inclusion bodies. Although the renaturation and the proper folding of the protein were not achieved, further studies are highly required to promote refolding. These studies should be performed on an automated liquid chromatography system and the use of other additives (e.g. arginine) should be considered for a matrix-assisted protein refolding [138].

An affinity ligand for the purification of GFP fusion proteins was successfully accomplished. Ligand A4C7 seemed highly specific for GFP and its selectivity was maintained when using a proper GFP fusion protein produced in other hosts.

Overall, the development of an affinity pair for the purification of recombinant proteins was accomplished but some efforts are still required for the adjustment of binding and elution conditions to improve the kinetics of the affinity pair. Other important parameter when applied this system in a commercial view is the matrix reuse. In other words, the immobilized ligand should also hold different cycles of cleaning-in-place, so that can be further used for the purification of tagged proteins without compromising its activity.

The proof of concept was made with two GFP proteins, one non-tagged and a fusion protein, but the ligand must be challenged in the future with other GFP fusion proteins.

CHAPTER 6**CONCLUDING REMARKS**

The work presented in this thesis aimed at the development of new affinity pairs “tag-receptor” for the purification of recombinant proteins, which combined the best of small peptide tags with the best of synthetic affinity ligands.

Five promising hexapeptides with different hydrophobic / hydrophilic characters were previously selected as potential affinity tags, and the respective complementary binding partners based on the one-pot multicomponent Ugi reaction were *de novo* designed *in silico* [207]. This work served as inspiration for the synthesis of a combinatorial library comprising 64 Ugi-based affinity ligands in solid phase (aldehyde-activated agarose), and diversity was introduced through the variation of the amine and carboxylic acid compounds, whereas the isocyanide functionality was kept constant.

In order to improve the selection of the best affinity ligands to target the tag peptides, the latter were genetically fused to Green Fluorescent Protein (GFP) and produced in *Escherichia coli* (*E. coli*). As a successful proof of concept, the DNA fragment that codifies for the RWRWRW tagged GFP was cloned on the expression vector pET21c for further expression on *E. coli* cells (BL21 (DE3)). A particularity of this DNA fragment is that GFP is within two multicloning sites and this vector could be extended for the production of other proteins. Moreover, a nucleotide sequence between the tag and the protein was designed to codify for the enterokinase recognition site, useful for future tag removal. The production of GFP fused to the affinity tags NWNWNW, RKRKRK and NNNNNN was straightforward leading to a production of mostly or totally soluble protein, with a percentage of GFP-tagged in the crude extract of 9, 4 and 5% (w/w), respectively for 1L of culture medium. Regarding the production of GFP tagged with RWRWRW and WFWFWF, the percentage of soluble tagged protein was as low as 2% (w/v) in 1L of culture. In addition, the RWRWRW peptide presented antimicrobial properties with a minimum inhibitory concentration of 2.5mM for *E. coli* K12. The production of this peptide fused to GFP was found to be a burden for bacterial cells, and the fusion with GFP worked as a limited mask for peptide toxicity. For these reasons, the RWRWRW peptide was not considered as suitable affinity tag to be used in bioseparation. The WFWFWF tagged GFP was produced

as inclusion bodies, and different culture conditions such as temperature and inductor levels were studied to improve the amount of soluble protein. Therefore, the screening with this affinity tag was performed in both denaturant and non-denaturant conditions.

The bacterial extracts containing GFP-tagged proteins were further screened for binding to a solid phase second generation combinatorial library based on the Ugi reaction comprising of 64 affinity ligands. Screening was performed on a 96-well batch mode by micro-scale affinity chromatography as it allows testing a wide range of adsorbents under different binding/elution conditions [226]. As a first approach, the selection of the putative lead ligands was possible for all affinity tags, except for NNNNNN tagged GFP, which was not used for further studies. The putative lead ligands were then re-screened for binding to GFP-tagged proteins and recovered samples were analysed by SDS-PAGE, where selective binding was estimated on the basis of the loading and flow-through samples. The overall analysis led to the selection of three lead affinity pairs tag-receptor: WFWFWF tagged GFP – ligand A4C8 (in denaturant conditions), NWNWNW tagged GFP – ligand A3C4 and RKRKRK tagged GFP – ligand A7C1. Ligand A4C8 bound 11 ± 0.5 μg of WFWFWF tagged GFP per gram of resin (90%), ligand A3C4 bound 156 ± 12 μg of NWNWNW tagged GFP per gram of resin (58%) and ligand A7C1 bound to 77 ± 4 μg of RKRKRK tagged GFP per gram of resin (50%). As a negative control, the second generation library was also screened against GFP extracts from *E. coli*. The affinity ligand A4C7 was found to bind ~100% GFP (590 ± 3 μg of GFP per gram of resin (99%)), all tagged-GFP proteins and most host proteins. Despite the apparent lack of selectivity the affinity pair A4C7-GFP was taken into account for further optimizations. Partition equilibrium experiments fitted to Langmuir isotherm with good correlation factors were performed to estimate affinity constants (K_a) and binding capacity (Q_{max}) for the lead affinity systems, although limited by the amount of unpurified protein produced by the bacterial cells. The K_a values obtained for the affinity pairs NWNWNW tagged GFP-A3C4 and RKRKRK tagged GFP-A7C1 were in order of $7.22 \times 10^4 \text{ M}^{-1}$ and $2.45 \times 10^5 \text{ M}^{-1}$, respectively. The Q_{max} values obtained were apparently low with high errors associated.

The four affinity systems (i) WFWFWF tagged GFP-A4C8 (denaturant conditions), (ii) NWNWNW tagged GFP-A3C4, (iii) RKRKRK tagged GFP-A7C1 and (iv) GFP-A4C7 were selected for further studies in terms of selectivity, optimization of elution conditions, binding kinetics and further use on an automated chromatographic system. The affinity systems (i) to (iii) were mainly developed to be used as an affinity pair “tag-receptor”, where (i) might be employed as matrix-assisted refolding strategy. The affinity system (iv) was interesting to explore for the purification of GFP fusion proteins.

The affinity pair WFWFWF tagged GFP- A4C8 might be an alternative matrix-assisted refolding strategy considering that the A4C8 ligand was found to selectively bind to WFWFWF tagged GFP in presence of 8M urea. Hydrophobic interaction is expected to prevail on the molecular recognition between WFWFWF tagged GFP and A4C8. The preliminary attempts performed for proper on-column refolding of protein prior to elution were not successful. Still, the optimization of this system could represent an alternative for refolding strategies, specially for those based on structural ligands such as ion-exchange resins and metal chelate ligands with the respective binding partners Z_{basic} [48-49] and His-tag [134]. Ion-exchange resins are cost-effective, however charge-charge interactions can lead to protein aggregation [8]. The His-tag is compatible with denaturant conditions, but metal leaching can occur and the imidazole side chain groups of histidine interfere with protein folding [74].

The affinity pair NWNWNW tagged GFP-A3C4 showed very low binding capacity (less than 10% of loaded protein) and protein recovery (elution yields of 3%) when employing the automated liquid chromatography system, which points toward system optimization. Another alternative could be to further optimize other potential lead candidates.

The affinity pair RKRKRK tagged GFP-A7C1 revealed to be promising for the purification of fusion proteins, as it showed a reproducible and consistent binding profile, although the amount of GFP-tagged per gram of resin has slightly decreased when applying an automated system (Table 5.3 § 5.7.3). However, this can be improved by an adjustment of the flow rate and residence time. Regarding the optimization of elution conditions carried out on a 96-well format, the most promising were the competitive elution with 10mM PBS, 500 mM arginine and 0.1M Glycine-NaOH pH 11, 150 mM NaCl, with a protein recovery of 30 and 24%, respectively. Both conditions when applied to automated liquid chromatography system led to a purity of 50% and protein recovery of 30%. The affinity pair RKRKRK- A7C1 represents increased selectivity when compared with a common IEX based process, as the elution process is more complex, indicating mixed modes of interaction. Also, the novel RKRKRK tag is recognised by a robust and inexpensive synthetic ligand. An overview and comparison of the affinity pair performance with currently established platforms is shown in Table 6.1.

Table 6.1 Comparison of the RKRKRK affinity tag with similar tags commercially available

Affinity tag	Size	Resin	Ka (M ⁻¹)	Method	Dynamic binding capacity (mg/ml)
Strep-tag II	8 aa	StrepTactin™	1.4x10 ⁴	ITC ¹ [101]	-6 mg/ml (GE Healthcare) -9 mg/ml (QIAGEN)
FLAG tag	8 aa	-ANTI-FLAG M1 -ANTI-FLAG M2 -ANTI-FLAG antibody magnetic beads	2.4x10 ⁶	SPR ² [268]	-0.6 mg/ml (ANTI-FLAG M1, M2) (SIGMA-ALDRICH) -1.3 mg/ml (magnetic beads) (Clontech)
His-tag	6 aa	-Ni-IDA Sepharose -HisPur Ni-NTA -HisPur Cobalt -His60 Ni (Ni-IDA) -Ni-NTA	10 ⁷	SPR ² [181]	- 40mg/ml (GE Healthcare) - 20 mg/ml (Piercenet) -20 mg/ml (Piercenet) -60 mg/ml (Clontech) -5-10 mg/ml (Invitrogen)
RKRKRK	6 aa	A7C1 (Ugi based ligand)	2.5x10 ⁵	Static partition equilibrium	-0.5 mg/g ³

¹ ITC - isothermal titration calorimetry; ² SPR - surface Plasmon resonance; ³ Static binding capacity.

A novel strategy for the purification of GFP fusion proteins was also developed based on Ugi-based ligand A4C7. Although the affinity constant Ka (2.38x10⁵ M⁻¹) was suitable for affinity based chromatography processes, the binding capacity is still low. Despite the low selectivity in binding, the fine tuning of elution conditions allowed the selective GFP recovery (92-100%) with a purity of 94% and 66% for the conditions 0.1M glycine-NaOH pH 9 and 0.1M glycine-NaOH pH 9 with 50% ethylene glycol, respectively. The affinity ligand A4C7 was also employed on the purification of GFP fused to a *B. megaterium* spore cortex lytic enzyme SleL produced in *Lactococcus lactis* [267] with a total molecular weight of 78 KDa. Adsorbents containing ligand A4C7 bound a high binding percentage (90%) of GFP, with a 50% recovery. This work sets the ground for a deeper exploitation of ligand A4C7 for purification of GFP fusion proteins. To date, there is no selective synthetic affinity ligand towards GFP and the purification of GFP fusion

proteins is often cumbersome (Table 6.2.) with single or multiple chromatographic steps needed. Immunoaffinity chromatography which requires the use of biological ligands such as monoclonal antibodies is the most popular strategy. The novel affinity adsorbent based on A4C7 can yield similar results in terms of recovery and purity and can overcome the limitations of the described methods.

Table 6.2 Purification methods used on the purification of green fluorescent protein (GFP)

GFP source	Purification Method	Yield (% Recovery)	Purity (%)	Reference
<i>E.coli</i> cell lysate	Organic extraction	60	90	[269]
<i>E.coli</i> cell lysate	Chromatofocusing	60	80	[270]
<i>E.coli</i> cell lysate	Two IMAC steps ¹	80-66	n.a.	[271]
<i>E.coli</i> cell lysate	ATPE ² + IMAC ¹	32-81	n.a.	[272]
<i>E.coli</i> cell lysate	Three phase partitioning ³	78	n.a.	[273]
<i>E.coli</i> cell lysate	(NH ₄) ₂ SO ₄ Precipitation + HIC	74-80	97	[274]
Culture tobacco cells lysate	Multiple step ⁴	95	80	[275]
<i>E.coli</i> cell lysate	Immunoaffinity	90	97	[276]
<i>E.coli</i> cell lysate	Novel Affinity chromatography based on A4C7 affinity ligand	92	94	-

¹The immobilized metal affinity chromatography (IMAC) was performed by using histidine residues available on the surface of GFP without the use of a histidine tail; **n.a.**: data not available; ²aqueous two-phase extraction (ATPE) ³ the method involves a combination of ammonium sulphate and tert-butanol precipitation methods; ⁴ the multiple step chromatography purification method involved the methods ammonium sulphate precipitation, hydrophobic interaction chromatography (HIC) and ion-exchange chromatography (IEX)

The novel affinity pair RKRKRK tagged GFP-A7C1 seems to combine increased selectivity with the cost-effectiveness of structural ligands, representing an interesting alternative to explore fusion protein production, purification and immobilization. The synthetic ligand A4C7 shows interesting selectivity on the recovery of bound GFP protein, and therefore can serve as a capturing agent for GFP alone or GFP fusion proteins, with a wide range of applications in the bioseparation and diagnostic fields.

BIBLIOGRAPHY

1. Martínez, J.L., et al., *Current Opinion in Biotechnology*, 2012. **23**: p. 965-971.
2. Carter, P.J., *Experimental Cell Research*, 2011. **317**: p. 1261-1269.
3. Carta, G. and A. Jungbauer, *Downstream Processing of Biotechnology Products*, in *Protein Chromatography*. 2010, Wiley-VCH Verlag GmbH & Co. KGaA. p. 1-55.
4. Cramer, S.M. and M.A. Holstein, *Current Opinion in Chemical Engineering*, 2011. **1**: p. 27-37.
5. Walsh, G., *The drug manufacturing process*, in *Biopharmaceuticals Biochemistry and Biotechnology* L. John Wiley & Sons, Editor. 2003. p. 93 - 185.
6. Lowe, C.R., *Affinity Chromatography and Related Techniques: Perspectives and Trends*, in *Advances in Molecular and Cell Biology*, B.D. E. Edward Bittar and B. Lelf, Editors. 1996, Elsevier. p. 513-522.
7. Arnau, J., et al., *Protein expression and purification*, 2006. **48**: p. 1 - 13.
8. Hedhammar, M., T. Gräslund, and S. Hober, *Chemical Engineering & Technology*, 2005. **28**: p. 1315 - 1325.
9. Nolan, T., N. Singh, and C.R. McCurdy, *Ligand Macromolecule Interactions: Theoretical Principles of Molecular Recognition*. 2010. p. 13-29.
10. Waugh, D.S., *Trends in Biotechnology*, 2005. **23**: p. 316 - 320.
11. Young, C.L., Z.T. Britton, and A.S. Robinson, *Biotechnology Journal* 2012. **7**: p. 620-634.
12. Lowe, C.R., *Current Opinion in Chemical Biology*, 2001. **5**: p. 248 - 256.
13. Roque, A.C.A. and C.R. Lowe, *Biotechnology Advances*, 2006. **24**: p. 17 - 26.
14. Roque, A.C.A. and C.R. Lowe, *Affinity Chromatography: History, Perspectives, Limitations and Prospects*, in *Affinity Chromatography: Methods and Protocols*, M. Zachariou, Editor. 2007, Human Press. p. 1-23.
15. Block, H., et al., *Immobilized-Metal Affinity Chromatography (IMAC) : A Review*, in *Methods in Enzymology: Guide to Protein Purification*, R.R. Burgess and M.P. Deutscher, Editors. 2009. p. 439-473.
16. Charlton, A. and M. Zachariou, *Immobilized Metal Ion Affinity Chromatography of Native Proteins*, in *Affinity Chromatography : Methods and Protocols*. 2007. p. 25-36.
17. Gaberc-Porekar, V. and V. Menart, *Journal of Biochemical and Biophysical Methods*, 2001. **49**: p. 335-360.
18. Li, R., et al., *Nature Biotechnology*, 1998. **16**: p. 190-195.
19. Roque, A.C.A., M.A. Taipa, and C.R. Lowe, *Journal of Chromatography A*, 2005. **1064**: p. 157 - 167.
20. Teng, S.F., et al., *Journal of Chromatography B*, 2000. **740**: p. 1 - 15.

21. Palanisamy, U.D., D.J. Winzor, and C.R. Lowe, *Journal of Chromatography B: Biomedical Sciences and Applications*, 2000. **746**: p. 265-281.
22. Gupta, G. and C.R. Lowe, *Journal of Molecular Recognition*, 2004. **17**: p. 218-235.
23. Renou, E.N.S., et al., *Journal of Molecular Recognition*, 2004. **17**: p. 248-261.
24. Morrill, P.R., et al., *Journal of Chromatography B*, 2002. **774**: p. 1-15.
25. Filippusson, H., L.S. Erendsson, and C.R. Lowe, *Journal of Molecular Recognition*, 2000. **13**: p. 370 - 381.
26. Roque, A.C.A. and C.R. Lowe, *Biotechnology Advances*, 2006. **24**: p. 17 - 26.
27. Evers, P., *The future of the biologicals market: Market overview, innovations and company profiles 2010*.
28. Gottschalk, U., K. Brorson, and A.A. Shukla, *Nature Biotechnology*, 2012. **30**: p. 489-492.
29. Clonis, Y.D., *Journal of Chromatography A*, 2006. **1101**: p. 1 - 24.
30. Lowe, C.R., A.R. Lowe, and G. Gupta, *Journal of biochemical and biophysical methods*, 2001. **49**: p. 561 - 574.
31. Milne, J., *Scale-Up of Protein Purification: Downstream Processing Issues*, in *Protein Chromatography*, D. Walls and S.T. Loughran, Editors. 2011, Humana Press. p. 73-85.
32. Walter, J. and U. Gottschalk, *Concepts for Disposables in Biopharmaceutical Manufacture*, in *Current Trends in Monoclonal Antibody Development and Manufacturing*, S.J. Shire, et al., Editors. 2010, Springer New York. p. 87-99.
33. Janson, J.-C., *Trends in Biotechnology*, 1984. **2**: p. 31-38.
34. Lowe, C.R. and P.D.G. Dean, *Affinity Chromatography*. 1974: Jonh Wiley & Sons.
35. Urh, M., D. Simpson, and K. Zhao, *Chapter 26 Affinity Chromatography: General Methods*, in *Methods in Enzymology*, R.B. Richard and P.D. Murray, Editors. 2009, Academic Press. p. 417-438.
36. Hermanson, G.T., A.K. Mallia, and P.K. Smith, *Immobilized affinity ligand techniques*. 1992: Academic Press.
37. Rowe, L., G. El Khoury, and C.R. Lowe, *Affinity Chromatography: Historical and Prospective Overview*, in *Biopharmaceutical production technology 1*, G. Subramanian, Editor. 2012, Wiley-VCH: Weinheim.
38. Haigh, J.M., et al., *Journal of Chromatography B*, 2009. **877**: p. 1440-1452.
39. Clonis, Y.D., et al., *Journal of Chromatography A*, 2000. **891**: p. 33 - 44.
40. Lowe, C.R., *Current opinion in chemical biology*, 2001. **5**: p. 248 - 256.
41. Björck, L. and G. Kronvall, *The Journal of Immunology*, 1984. **133**: p. 969-74.
42. Duhamel, R.C., et al., *Journal of Immunological Methods*, 1979. **31**: p. 211-217.
43. Lindmark, R., K. Thorén-Tolling, and J. Sjöquist, *Journal of Immunological Methods*, 1983. **62**: p. 1-13.
44. Füglistaller, P., *Journal of Immunological Methods*, 1989. **124**: p. 171-177.

45. Nilson, B.H.K., et al., *Journal of Immunological Methods*, 1993. **164**: p. 33-40.
46. Vretblad, P., *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 1976. **434**: p. 169-176.
47. Smith, G.P. and V.A. Petrenko, *Chemical Reviews*, 1997. **97**: p. 391-410.
48. Gräslund, T., et al., *Protein Engineering*, 2000. **13**: p. 703-709.
49. Hedhammar, M., et al., *Biotechnology Journal*, 2006. **1**: p. 187-196.
50. Hedhammar, M., et al., *Protein Engineering Design and Selection*, 2004. **17**: p. 779-786.
51. Cummins, P.M. and B.F. O'Connor, *Hydrophobic Interaction Chromatography*, in *Protein Chromatography*, D. Walls and S.T. Loughran, Editors. 2011, Humana Press. p. 431-437.
52. Porath, J., *Biotechnology Progress*, 1987. **3**: p. 14-21.
53. Liu, X.-C. and W.H. Scouten, *Journal of Chromatography A*, 1994. **687**: p. 61-69.
54. Liu, X.-C. and W.H. Scouten, *Boronate Affinity Chromatography*, in *Affinity Chromatography*, P. Bailon, et al., Editors. 2000, Humana Press. p. 119-128.
55. Bouriotis, V., I.J. Galpin, and P.D.G. Dean, *Journal of Chromatography A*, 1981. **210**: p. 267-278.
56. Liu, X.-C., *Chinese Journal of Chromatography*, 2006. **24**: p. 73-80.
57. Borlido, L., et al., *Journal of Chromatography B*, 2012. **903**: p. 163-170.
58. Zhao, G., X.-Y. Dong, and Y. Sun, *Journal of Biotechnology*, 2009. **144**: p. 3-11.
59. Brenac Brochier, V., et al., *Journal of Chromatography A*, 2008. **1177**: p. 226-233.
60. Mant, C.T. and R.S. Hodges, *Journal of Separation Science*, 2008. **31**: p. 2754-2773.
61. Yon, R.J., *Analytical Biochemistry*, 1981. **113**: p. 219-228.
62. Burton, S.C., N.W. Haggarty, and D.R.K. Harding, *Biotechnology and Bioengineering*, 1997. **56**: p. 45-55.
63. Entzeroth, M., *Current Opinion in Pharmacology*, 2003. **3**: p. 522 - 529.
64. Kodadek, T., *Chemical Communications*, 2011. **47**: p. 9757-9763.
65. Roque, A.C.A., M.Â. Taipa, and C.R. Lowe, *Journal of Chromatography A*, 2005. **1064**: p. 157-167.
66. Wang, J., et al., *Separation and Purification Technology*, 2006. **50**: p. 141-146.
67. Wu, F. and J. Yu, *Biochemical and Biophysical Research Communications*, 2007. **355**: p. 673-678.
68. Qian, J., et al., *Journal of Chromatography B*, 2012. **898**: p. 15-23.
69. Khoury, G., L.A. Rowe, and C.R. Lowe, *Biomimetic Affinity Ligands for Immunoglobulins Based on the Multicomponent Ugi Reaction*, in *Chemical Genomics and Proteomics : Reviews and Protocols*, E.D. Zanders, Editor. 2012, Humana Press. p. 57-74.
70. Malhotra, A., *Chapter 16 Tagging for Protein Expression*, in *Methods in Enzymology*, R.R. Burgess and M.P. Deutscher, Editors. 2009, Academic Press. p. 239-258.

BIBLIOGRAPHY

71. Nilsson, J., et al., Protein expression and purification, 1997. **11**: p. 1 - 16.
72. Hu, S.-M., A.H.-J. Wang, and T.-F. Wang, *Expression Tags for protein production*, in *Encyclopedia of life sciences*. 2001.
73. Walls, D. and S.T. Loughran, *Tagging recombinant proteins to enhance solubility and aid purification*, in *Protein Chromatography: Methods and protocols* D. Walls and S.T. Loughran, Editors. 2011, Humana Press.
74. Hearn, M.T.W. and D. Acosta, Journal of Molecular Recognition, 2001. **14**: p. 323 - 369.
75. Terpe, K., Applied Microbiology and Biotechnology, 2003. **60** p. 523 - 533.
76. Waugh, D.S., Protein Expression and Purification, 2011. **80**: p. 283-293.
77. Nilsson, B. and L. Abrahmsén, Methods in Enzymology, 1990. **185**: p. 144 - 161.
78. Nilsson, B., L. Abrahmsén, and M. M Uhlén, The EMBO journal, 1985. **4**: p. 1075-1080.
79. Nilsson, B., et al., Protein Engineering, 1987. **1**: p. 107-113.
80. Akerström, B., E. Nielsen, and L. Björck, Journal of Biological Chemistry, 1987. **262**: p. 13388-13391.
81. Nygren, P.-Å., et al., Journal of Molecular Recognition, 1988. **1**: p. 69-74.
82. Einhauer, A. and A. Jungbauer, Journal of Biochemical and Biophysical Methods, 2001. **49**: p. 455-465.
83. Allen, S.J., et al., Bioresource Technology, 2003. **88**: p. 143-152.
84. Hopp, T.P., et al., Nature Biotechnology, 1988. **6**: p. 1204-1210.
85. Prickett, K.S., D.C. Amberg, and T.P. Hopp, Biotechniques, 1989. **7**: p. 580-9.
86. Evan, G.I., et al., Molecular and Cellular Biology, 1985. **5**: p. 3610 - 3616.
87. Kipriyanov, S.M., et al., Journal of Immunological Methods, 1996. **196**: p. 51-62.
88. Jarvik, J.W. and C.A. Telmer, Annual Review of Genetics, 1998. **32**: p. 601-618.
89. Studier, F.W. and B.A. Moffatt, Journal of Molecular Biology, 1986. **189**: p. 113-130.
90. Wilson, I.A., et al., Cell, 1984. **37**: p. 767-778.
91. Field, J., et al., Molecular and Cellular Biology, 1988. **8**: p. 2159-2165.
92. Foreman, P.K. and R.W. Davis, Gene, 1994. **144**: p. 63-68.
93. Burgess, R.R. and N.E. Thompson, Current Opinion in Biotechnology, 2002. **13**: p. 304-308.
94. Duellman, S.J., N.E. Thompson, and R.R. Burgess, Protein Expression and Purification, 2004. **35**: p. 147-155.
95. Thompson, A.J., et al., Analytical Biochemistry, 2003. **75**: p. 3232-3243.
96. Edwards, A.M., et al., Proceedings of the National Academy of Sciences, 1990. **87**: p. 2122 - 2126.
97. Smith, D.B. and K.S. Johnson, Gene, 1988. **67**: p. 31-40.
98. Stofko-Hahn, R.E., D.W. Carr, and J.D. Scott, FEBS Letters, 1992. **302**: p. 274-278.
99. Kim, J.S. and R.T. Raines, Protein Science, 1993. **2**: p. 348 - 356.

100. Schmidt, T.G.M. and A. Skerra, *Protein Engineering Design and Selection*, 1993. **6**: p. 109-122.
101. Schmidt, T.G.M., et al., *Journal of Molecular Biology*, 1996. **255**: p. 753-766.
102. Lamla, T. and V.A. Erdmann, *Protein Expression and Purification*, 2004. **33**: p. 39-47.
103. Keefe, A.D., et al., *Protein Expression and Purification*, 2001. **23**: p. 440-446.
104. Schmidt, T.G.M. and A. Skerra, *Nature Protocols*, 2007. **2**: p. 1528-1535.
105. Gaj, T., S.C. Meyer, and I. Ghosh, *Protein Expression and Purification*, 2007. **56**: p. 54-61.
106. Donaldson, D.D., et al., *The Journal of Immunology*, 1998. **161**: p. 2317-2324.
107. Fuchs, S.M. and R.T. Raines, *Protein Science*, 2005. **14**: p. 1538 - 1544.
108. Thompson, N.E., T.M. Arthur, and R.R. Burgess, *Analytical Biochemistry*, 2003. **323**: p. 171-179.
109. di-Guan, C., et al., *Gene*, 1988. **67**: p. 21 - 30.
110. Tomme, P., et al., *Journal of Chromatography B: Biomedical Sciences and Applications*, 1998. **715**: p. 283-296.
111. Chong, S., et al., *Gene*, 1997. **192**: p. 271-281.
112. Luoqing, C., C. Ford, and Z. Nikolov, *Gene*, 1991. **99**: p. 121-126.
113. Bayer, E.A. and M. Wilchek, *Journal of Chromatography A*, 1990. **510**: p. 3-11.
114. Korndörfer, I.P. and A. Skerra, *Protein Science*, 2002. **11**: p. 883-893.
115. Lamla, T. and V.A. Erdmann, *Journal of Molecular Biology*, 2003. **329**: p. 381-388.
116. Vaillancourt, P., et al., [22] *Affinity purification of recombinant proteins fused to calmodulin or to calmodulin-binding peptides*, in *Methods in Enzymology*, J. Thorner, S.D. Emr, and J.N. Abelson, Editors. 2000, Academic Press. p. 340-362.
117. Melkko, S. and D. Neri, *Calmodulin as an Affinity Purification Tag*, in *Methods in Molecular Biology - E. coli gene expression protocols* P. Vaillancourt, Editor. 2003, Academic Press. p. 69-77.
118. Karpeisky, M.Y., et al., *FEBS Lett*, 1994. **339**: p. 209-212.
119. Boyer, T.D., *Hepatology*, 1989. **9**: p. 486-496.
120. Smith, D.B. and K.S. Johnson, *Gene*, 1988. **67**: p. 31 - 40.
121. Smith, D.B., [17] *Generating fusions to glutathione S-transferase for protein studies*, in *Methods in Enzymology*, J. Thorner, S.D. Emr, and J.N. Abelson, Editors. 2000, Academic Press. p. 254-270.
122. Frangioni, J.V. and B.G. Neel, *Analytical Biochemistry*, 1993. **210**: p. 179-187.
123. Kaplan, W., et al., *Protein Science*, 1997. **6**: p. 399-406.
124. Kellermann, O.K. and T. Ferenci, [75] *Maltose-binding protein from Escherichia coli*, in *Methods in Enzymology - Carbohydrate Metabolism - Part E*, A.W. Willis, Editor. 1982, Academic Press. p. 459-463.
125. Nikaido, H., *FEBS Letters*, 1994. **346**: p. 55-58.

BIBLIOGRAPHY

126. di Guana, C., et al., *Gene*, 1988. **67**: p. 21-30.
127. Miller, D.M., et al., *Journal of Biological Chemistry*, 1983. **258**: p. 13665-72.
128. Kapust, R.B. and D.S. Waugh, *Protein Science*, 1999. **8**: p. 1668-1674.
129. Kurek, D.V., S.A. Lopatin, and V.P. Varlamov, *Applied Biochemistry and Microbiology*, 2009. **45**: p. 1-8.
130. Cheung, R., J. Wong, and T. Ng, *Applied Microbiology and Biotechnology*, 2012. **96**: p. 1411-1420.
131. Porath, J., *Protein Expression and Purification*, 1992. **3**: p. 263-281.
132. Porath, J., et al., *Nature*, 1975. **258**: p. 598-599.
133. Porath, J. and B. Olin, *Biochemistry*, 1983. **22**: p. 1621-1630.
134. Hochuli, E., et al., *Nature Biotechnology*, 1988. **6**: p. 1321-1325.
135. Janknecht, R., et al., *Proceedings of the National Academy of Sciences*, 1991. **88**: p. 8972-8976.
136. Westra, D.F., et al., *Journal of Chromatography B: Biomedical Sciences and Applications*, 2001. **760**: p. 129-136.
137. Hochuli, E., *Journal of Chromatography A*, 1988. **444**: p. 293-302.
138. Dashivets, T., et al., *ChemBioChem*, 2009. **10**: p. 869-876.
139. Kato, K., H. Sato, and H. Iwata, *Langmuir*, 2005. **21**: p. 7071-7075.
140. Wegner, G.J., et al., *Analytical Chemistry*, 2003. **75**: p. 4740-4746.
141. Wilson, D.S. and S. Nock, *Current Opinion in Chemical Biology*, 2002. **6**: p. 81-85.
142. Chaga, G., et al., *Journal of Chromatography A*, 1999. **864**: p. 247-256.
143. Mooney, J.T., D. Fredericks, and M.T.W. Hearn, *Journal of Chromatography A*, 2011. **1218**: p. 92-99.
144. Sassenfeld, H.M. and S.J. Brewer, *Nature Biotechnology*, 1984. **2**: p. 76-81.
145. Stempfer, G., B. Holl-Neugebauer, and R. Rudolph, *Nature Biotechnology*, 1996. **14**: p. 329-334.
146. Kweon, D.-H., et al., *Biotechnology Progress*, 2004. **20**: p. 277-283.
147. Dalboge, H., et al., *Bio/Technology*, 1987. **5**: p. 161 - 164.
148. Stubenrauch, K., et al., *Journal of Chromatography B: Biomedical Sciences and Applications*, 2000. **737**: p. 77-84.
149. Zhang, C., et al., *Biotechnology Progress*, 2001. **17**: p. 161-167.
150. Becker, K., J.V. Alstine, and L. Büllow, *Journal of Chromatography A*, 2008. **1202**: p. 40 - 46.
151. Fexby, S. and L. Büllow, *Trends in Biotechnology*, 2004. **22**: p. 511-516.
152. Ohana, R.F., et al., *Protein Expression and Purification*, 2009. **68**: p. 110-120.
153. Peng, C.-C., et al., *Journal of Biotechnology*, 2004. **111**: p. 51-57.
154. Bhatla, S.C., V. Kaushik, and M.K. Yadav, *Biotechnology Advances*, 2010. **28**: p. 293-300.

155. McLean, M., et al., *Transgenic Research*, 2012. **21**: p. 1291-1301.
156. Floss, D.M., et al., *Trends in Biotechnology*, 2010. **28**: p. 37-45.
157. Hassouneh, W., S.R. MacEwan, and A. Chilkoti, *Chapter nine - Fusions of Elastin-Like Polypeptides to Pharmaceutical Proteins*, in *Methods in Enzymology*, K.D. Wittrup and L.V. Gregory, Editors. 2012, Academic Press. p. 215-237.
158. Meyer, D.E. and A. Chilkoti, *Nature Biotechnology*, 1999. **17**: p. 1112-1115.
159. Banki, M.R., L. Feng, and D.W. Wood, *Nature Methods*, 2005. **2**: p. 659-662.
160. Wu, W.-Y., et al., *Nature Protocols*, 2006. **1**: p. 2257-2262.
161. Fong, B.A., W.-Y. Wu, and D.W. Wood, *Trends in Biotechnology*, 2010. **28**: p. 272-279.
162. Li, Y., *Biotechnology Letters*, 2011. **33**: p. 869-881.
163. Fong, B.A., W.-Y. Wu, and D.W. Wood, *Trends Biotechnol*, 2010. **28**: p. 272-279.
164. Demain, A.L. and P. Vaishnav, *Biotechnology Advances*, 2009. **27**: p. 297-306.
165. Lange, C. and R. Rudolph, *Production of Recombinant Proteins for Therapy, Diagnostics, and Industrial Research by in Vitro Folding*, in *Protein Folding Handbook*. 2008, Wiley-VCH Verlag GmbH. p. 1245-1280.
166. García-Fruitós, E., et al., *Trends in Biotechnology*, 2012. **30**: p. 65-70.
167. Katti, S.K., D.M. LeMaster, and H. Eklund, *Journal of Molecular Biology*, 1990. **212**: p. 167-184.
168. Li, S.-J. and M. Hochstrasser, *Nature*, 1999. **398**: p. 246-251.
169. Marblestone, J.G., et al., *Protein Science*, 2006. **15**: p. 182-189.
170. Gusarov, I. and E. Nudler, *Cell*, 2001. **107**: p. 437-449.
171. Liu, K. and M.M. Hanna, *Journal of Molecular Biology*, 1995. **247**: p. 547-558.
172. LaVallie, E.R., et al., *Nature Biotechnology*, 1993. **11**: p. 187-193.
173. LaVallie, E.R., et al., *[21] Thioredoxin as a fusion partner for production of soluble recombinant proteins in Escherichia coli*, in *Methods in Enzymology - Applications of Chimeric Genes and Hybrid Proteins: Gene expression and protein purification*, J. Thorner, S.D. Emr, and J.N. Abelson, Editors. 2000, Academic Press. p. 322-340.
174. Malakhov, M., et al., *Journal of Structural and Functional Genomics*, 2004. **5**: p. 75-86.
175. Harrison, R.G., *inNovations*, 2000. **11**: p. 4-7.
176. Günzl, A. and B. Schimanski, *Tandem Affinity Purification of Proteins*, in *Current Protocols in Protein Science*. 2001, John Wiley & Sons, Inc. p. 19.19.1-19.19.16.
177. Puig, O., et al., *Methods*, 2001. **24**: p. 218-229.
178. Xu, X., et al., *Protein Expression and Purification*, 2010. **72**: p. 149-156.
179. Ahmed, S., A. Daulat, and S. Angers, *Tandem Affinity Purification and Identification of Heterotrimeric G Protein-Associated Proteins*, in *Signal Transduction Protocols*, L.M. Luttrell and S.S.G. Ferguson, Editors. 2011, Humana Press. p. 357-370.
180. Ma, H., et al., *Molecular & Cellular Proteomics*, 2012. **11**: p. 501-511.
181. Müller, K.M., et al., *Analytical Biochemistry*, 1998. **259**: p. 54-61.

182. Choi, S., Il, et al., *Biotechnology and Bioengineering* 2001. **75**: p. 718-724.
183. Dougherty, W.G., et al., *The EMBO Journal*, 1988. **7**: p. 1281-7.
184. Jenny, R.J., K.G. Mann, and R.L. Lundblad, *Protein Expression and Purification*, 2003. **31**: p. 1-11.
185. Chang, J.-Y., *European Journal of Biochemistry*, 1985. **151**: p. 217-224.
186. Santana, S.D.F., A.S. Pina, and A.C.A. Roque, *Journal of Biotechnology*, 2012. **161**: p. 378-382.
187. Novagen, *pET System Manual*. 2003.
188. Barbas, C.F., et al., *Cold Spring Harbor Protocols*, 2007. **2007**: p. pdb.ip47.
189. Sambrook, J. and D.W. Russel, *Gel Electrophoresis of DNA and pulsed-field agarose gel electrophoresis in Molecular Cloning: A laboratory manual* C.S.H.L. Press, Editor. 2001: Cold Spring Harbor, New York. p. 5.1 - 5.86.
190. Burgess, R.R., *Chapter 17 Refolding Solubilized Inclusion Body Proteins*, in *Methods in Enzymology*, R.B. Richard and P.D. Murray, Editors. 2009, Academic Press. p. 259-282.
191. Terpe, K., *Applied Microbiology and Biotechnology*, 2006. **72**: p. 211-222.
192. Zerbs, S., A.M. Frank, and F.R. Collart, *Chapter 12 Bacterial Systems for Production of Heterologous Proteins*, in *Methods in Enzymology*, R.B. Richard and P.D. Murray, Editors. 2009, Academic Press. p. 149-168.
193. Brondyk, W.H., *Chapter 11 Selecting an Appropriate Method for Expressing a Recombinant Protein*, in *Methods in Enzymology*, R.B. Richard and P.D. Murray, Editors. 2009, Academic Press. p. 131-147.
194. Burgess-Brown, N.A., et al., *Protein Expression and Purification*, 2008. **59**: p. 94-102.
195. Chen, D. and D.E. Texada, *Gene Therapy and Molecular Biology*, 2006. **10**: p. 1-12.
196. Larrick, J.W. and D.W. Thomas, *Current Opinion in Biotechnology*, 2001. **12**: p. 411-418.
197. Xu, J., et al., *Biotechnology Advances*, 2012. **30**: p. 1171-1184.
198. Houdebine, L.-M., *Comparative Immunology, Microbiology and Infectious Diseases*, 2009. **32**: p. 107-121.
199. Graf, A., et al., *FEMS Yeast Research*, 2009. **9**: p. 335-348.
200. Dyck, M.K., et al., *Trends in Biotechnology*, 2003. **21**: p. 394-399.
201. Wurm, F.M., *Nature Biotechnology*, 2004. **11**: p. 1393 - 1398.
202. Porro, D., et al., *Molecular Biotechnology*, 2005. **31**: p. 245-259.
203. Georgiou, G., *AIChE Journal*, 1988. **34**: p. 1233-1248.
204. Zimmer, M., *Chemical Reviews*, 2002. **102**: p. 759-782.
205. Valdivia, R.H., B.P. Cormack, and S. Falkow, *The Uses of Green Fluorescent Protein in Prokaryotes*, in *Green Fluorescent Protein*. 2005, John Wiley & Sons, Inc. p. 163-178.

206. Ward, W.W., *Biochemical and Physical Properties of Green Fluorescent Protein*, in *Green Fluorescent Protein*. 2005, John Wiley & Sons, Inc. p. 39-65.
207. Pina, A.S., *Synthesis and screening of rationally designed combinatorial libraries for proteomics applications* Master Thesis, 2008, Departamento de Quimica, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Monte da Caparica
208. Pina, A.S., C.R. Lowe, and A.C.A. Roque, *Separation Science and Technology*, 2010. **45**: p. 2187-2193.
209. Samuelson, J., *Bacterial Systems*, in *Production of Membrane Proteins*. 2011, Wiley-VCH Verlag GmbH & Co. KGaA. p. 11-35.
210. Sørensen, H.P. and K.K. Mortensen, *Journal of Biotechnology*, 2005. **115**: p. 113-128.
211. Zhang, C. and X. Xing, *Chinese Journal of Chemical Engineering*, 2010. **18**: p. 863-869.
212. Liu, Z., et al., *Antimicrobial Agents and Chemotherapy*, 2007. **51**: p. 597-603.
213. Strøm, M.B., et al., *Journal of Medicinal Chemistry*, 2003. **46**: p. 1567-1570.
214. Mohan, K.V.K., S.S. Rao, and C.D. Atreya, *Transfusion*, 2010. **50**: p. 166-173.
215. Jenssen, H., P. Hamill, and R.E.W. Hancock, *Clinical Microbiology Reviews*, 2006. **19**: p. 491-511.
216. Li, Y., *Biotechnology and Applied Biochemistry*, 2009. **54**: p. 1-9.
217. Zorko, M. and R. Jerala, *Production of Recombinant Antimicrobial Peptides in Bacteria in The Antimicrobial Peptides*. 2010. p. 61-76.
218. Ventura, S. and A. Villaverde, *Trends in Biotechnology*, 2006. **24**: p. 179-185.
219. Wiegand, I., K. Hilpert, and R.E.W. Hancock, *Nature Protocols*, 2008. **3**: p. 163-175.
220. Hou, S., et al., *Applied and Environmental Microbiology*, 2010. **76**: p. 1967-1974.
221. Skosyrev, V.S., et al., *Protein Expression and Purification*, 2003. **27**: p. 55-62.
222. Singh, S.M. and A.K. Panda, *Journal of Bioscience and Bioengineering*, 2005. **99**: p. 303-310.
223. Jonasson, P., et al., *Biotechnology and Applied Biochemistry*, 2002. **35**: p. 91-105.
224. Missiakas, D. and S. Raina, *Journal of Bacteriology*, 1997. **179**: p. 2465-71.
225. Coffman, J.L., J.F. Kramarczyk, and B.D. Kelley, *Biotechnology and Bioengineering*, 2008. **100**: p. 605-618.
226. Chhatre, S. and N.J. Titchener-Hooker, *Journal of Chemical Technology & Biotechnology*, 2009. **84**: p. 927-940.
227. Łacki, K.M., *Biotechnology Journal*, 2012n/a-n/a.
228. Rege, K. and M. Heng, *Nature Protocols*, 2010. **5**: p. 408-417.
229. Hopp, J., et al., *Biotechnology Progress*, 2009. **25**: p. 1427-1432.
230. Rege, K., et al., *Biotechnology and Bioengineering*, 2006. **93**: p. 618-630.
231. Hanora, A., et al., *Journal of Chromatography A*, 2005. **1087**: p. 38-44.
232. Bergander, T., et al., *Biotechnology Progress*, 2008. **24**: p. 632-639.

233. Porzelle, A. and W.-D. Fessner, *Angewandte Chemie International Edition*, 2005. **44**: p. 4724-4728.
234. Dömling, A. and I. Ugi, *Angewandte Chemie International Edition*, 2000. **39**: p. 3168-3210.
235. Marcaccini, S. and T. Torroba, *Nature Protocols*, 2007. **2**: p. 632-639.
236. Dömling, A., *Current Opinion in Chemical Biology*, 2000. **4**: p. 318-323.
237. Chéron, N., et al., *The Journal of Organic Chemistry*, 2012. **77**: p. 1361-1366.
238. Bienaymé, H., et al., *Chemistry - A European Journal*, 2000. **6**: p. 3321-3329.
239. Böhm, H.-J. and G. Schneider, eds. *Protein-Ligand Interactions: From Molecular Recognition to Drug Design*. ed. R. Mannhold, H. Kubinyi, and G. Folkers. Vol. 19. 2003, Wiley.
240. Böhm, H.J., *Prediction of Non-bonded Interactions in Drug Design*, in *Protein-Ligand Interactions*. 2005, Wiley-VCH Verlag GmbH & Co. KGaA. p. 3-20.
241. Szmacinski, H., K. Ray, and J.R. Lakowicz, *Analytical Biochemistry*, 2009. **385**: p. 358-364.
242. Leo, A., C. Hansch, and D. Elkins, *Chemical Reviews*, 1971. **71**: p. 525-616.
243. Reid, B.G. and G.C. Flynn, *Biochemistry*, 1997. **36**: p. 6786-6791.
244. Yang, F., L.G. Moss, and G.N. Phillips, *Nature Biotechnology*, 1996. **14**: p. 1246-1251.
245. Levitt, M. and M.F. Perutz, *Journal of Molecular Biology*, 1988. **201**: p. 751-754.
246. Roque, A.C.A., C.S.O. Silva, and M.A. Taipa, *Journal of Chromatography A*, 2007. **1160**: p.
247. Firer, M.A., *Journal of Biochemical and Biophysical Methods*, 2001. **49**: p. 433 - 442.
248. Gallant, S.R., V. Koppaka, and N. Zecherle, *Dye Ligand Chromatography*. 2007. p. 61-70.
249. Winzor, D.J., *Journal of Biochemical and Biophysical Methods*, 2001. **49**: p. 99-121.
250. Schramm, V.L., et al., *Journal of Biological Chemistry*, 1984. **259**: p. 714-722.
251. Bøgg-Hansen, T.C. and K. Takeo, *ELECTROPHORESIS*, 1980. **1**: p. 67-71.
252. Vuignier, K., et al., *Analytical and Bioanalytical Chemistry*, 2010. **398**: p. 53-66.
253. Winzor, D.J., *Journal of Chromatography A*, 2004. **1037**: p. 351-367.
254. Livingston, A.G. and H.A. Chase, *Journal of Chromatography A*, 1989. **481**: p. 159-174.
255. Limousin, G., et al., *Applied Geochemistry*, 2007. **22**: p. 249-275.
256. Arnold, F.H., H.W. Blanch, and C.R. Wilke, *The Chemical Engineering Journal*, 1985. **30**: p. B9-B23.
257. Denizli, A. and E. Pişkin, *Journal of Biochemical and Biophysical Methods*, 2001. **49**: p. 391-416.
258. Sawyer, W.H. and J. Puckridge, *Journal of Biological Chemistry*, 1973. **248**: p. 8429-8433.
259. Zhang, Y. and P.S. Cremer, *Current Opinion in Chemical Biology*, 2006. **10**: p. 658-663.

260. Subramanian, A., *Molecular Biotechnology*, 2002. **20**: p. 41-47.
261. Arakawa, T. and S.N. Timasheff, *Biochemistry*, 1985. **24**: p. 6756-6762.
262. Lowe, C.R. and K. Mosbach, *European Journal of Biochemistry*, 1975. **52**: p. 99-105.
263. Ormö, M., et al., *Science*, 1996. **273**: p. 1392-1395.
264. Bashford, D. and M. Karplus, *Biochemistry*, 1990. **29**: p. 10219-10225.
265. Bas, D.C., D.M. Rogers, and J.H. Jensen, *Proteins: Structure, Function, and Bioinformatics*, 2008. **73**: p. 765-783.
266. Healthcare, G., *Purifying Challenging Proteins* G.H.B.-S. AB, Editor. 2007.
267. Geertsma, E. and B. Poolman, *Production of Membrane Proteins in Escherichia coli and Lactococcus lactis*, in *Heterologous Expression of Membrane Proteins*, I. Mus-Veteau, Editor. 2010, Humana Press. p. 17-38.
268. Einhauer, A. and A. Jungbauer, *Journal of Chromatography A*, 2001. **921**: p. 25-30.
269. Yakhnin, A.V., et al., *Protein Expression and Purification*, 1998. **14**: p. 382-386.
270. Narahari, C.R., et al., *Biotechnology Progress*, 2001. **17**: p. 150-160.
271. Li, Y., et al., *Journal of Chromatography A*, 2001. **909**: p. 183-190.
272. Li, Y. and R.R. Beitle, *Biotechnology Progress*, 2002. **18**: p. 1054-1059.
273. Jain, S., R. Singh, and M.N. Gupta, *Journal of Chromatography A*, 2004. **1035**: p. 83-86.
274. McRae, S.R., C.L. Brown, and G.R. Bushell, *Protein Expression and Purification*, 2005. **41**: p. 121-127.
275. Peckham, G.D., R.C. Bugos, and W.W. Su, *Protein Expression and Purification*, 2006. **49**: p. 183-189.
276. Zhuang, R., et al., *Protein Expression and Purification*, 2008. **59**: p. 138-143.